Fault Detection Utilizing Convolution Neural Network on Timeseries Synchrophasor Data From Phasor Measurement Units

Mohammad Alqudah, *Student Member, IEEE*, Martin Pavlovski, *Student Member, IEEE*, Tatjana Dokic, *Member, IEEE*, Mladen Kezunovic, *Life Fellow, IEEE*, Yi Hu, *Fellow, IEEE*, and Zoran Obradovic, *Senior Member, IEEE*

***Abstract*—An end-to-end supervised learning method is proposed for fault detection in the electric grid using Big Data from multiple Phasor Measurement Units (PMUs). The approach consists of preprocessing steps aimed at reducing data noise and dimensionality, followed by utilization of six classification models considered for detecting faults. Three of the models were variants of Convolutional Neural Network (CNN) architectures that consider a single type of measurement (voltage, current or frequency) at all PMUs or all types together also at all PMUs. CNN based models were compared to traditional methods of Logistic Regression (LR), Multi-layer Perceptron (MLP) and Support Vector Machine (SVM). Evaluation was conducted on two-year data measured by PMUs at 37 locations in a large electric grid. The response variable for classification were extracted from the grid-wide outage event log. Experiments show that CNN-based models outperformed traditional methods on one year out-of-sample outage detection over the entire grid.**

***Index Terms*—Big data applications, event detection, machine learning, phasor measurement units, power system faults, dimensionality reduction, smart grids, time series analysis, neural networks, convolutional neural networks.**

I. Introduction

PHASOR measurement Units (PMUs) provide measured estimates of voltage and current phasors, as well as frequency and rate of change of frequency (ROCOF) to the transmission grid operators and planners [1], providing a high-resolution real-time network view for situational awareness, as well as historical data for post-mortem analysis of various types of system disturbances. With the deployment of more and more PMUs in the nation's power grid, the increase in the amount of available data has reached a challenging level for utilities. To fully exploit the value of such large datasets, new techniques are being developed that can provide more automated and proactive practices. Many utilities are starting to rely on the Big Data technologies to store, process, and analyze the large PMU datasets. Despite increased deployment, currently, the PMUs are still sparsely located covering typically less than 5% of the system electrical buses. In addition, the placement density in various geographical regions may be uneven.

A. Problem Statement and Study Objectives

Several research questions related to power system fault detection are addressed in this study. First question is how to detect and characterize the power system faults based on a reduced set of PMUs in the Western interconnection in the USA (representing a small proportion of the actual network PMU's), where the faults may be causing a system wide manifestation (ex. frequency events) while actually being localized (ex. line faults). Given the small number of PMUs and the occurrence of local faults that could be anywhere in the system, the chances of the PMU placement coinciding with the location of the fault being detected is rather small, so the detection and classification of the faults based on the measurements taken at a distance from the fault occurrence location is an additional challenge that the proposed techniques need to handle. Second question is how to design automated detection systems that doesn't rely on extensive manual study of data and feature engineering. Due to real-time operation needs at the control centers, processing of large number of PMU measurement streams within a required time interval creates an additional constraint.

B. Related Work

Event detection and classification is a classical machine learning problem with multiple effective and well-studied methods proposed such as decision trees, support vector machine (SVM) and Bayesian models [1]. However, such methods might not perform well on sensor data due to high dimensionality, autocorrelation, and other factors. An overview of various data mining techniques and their use in power system analysis based on PMU measurements can be found in [2]. The existing methods vary mostly based on a classification/clustering algorithm, and the

Manuscript received 18 January 2021; revised 26 May 2021 and 12 September 2021; accepted 27 November 2021. Date of publication 14 December 2021; date of current version 19 August 2022. This work was supported by the Department of Energy under Award DE-OE0000913. Paper no. TPWRS-00088-2021. *(Corresponding author: Martin Pavlovski.)*

Mohammad Alqudah, Martin Pavlovski, and Zoran Obradovic are with the Temple University, Philadelphia, PA 19122 USA (e-mail: mohammad.alqudah@temple.edu; martin.pavlovski@temple.edu; zoran.obradovic@temple.edu).

Tatjana Dokic and Mladen Kezunovic are with the Texas A&M University, College Station, TX 77843-3128 USA (e-mail: tanjans88@gmail.com; kezunov@ece.tamu.edu).

Yi Hu is with the Quanta Technology, Raleigh, NC 27607 USA (e-mail: yhu@quanta-technology.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2021.3135336.

Digital Object Identifier 10.1109/TPWRS.2021.3135336

feature extraction method used to generate inputs to the classification algorithm. A PMU-based fault detection using Wavelet analysis was proposed in [4]. There is also a method based on the fast variant of the Discrete S-Transform feature extraction using Extreme Learning Machine (ELM) classifier [5]. Due to the high volume of PMU data, multiple studies were based on dimensionality reduction using Principal Component Analysis [6], [7]. Several studies used Minimum Volume Enclosing Ellipsoid to extract features needed for the classification and clustering algorithms [8], [9]. There is also a study using an unsupervised learning technique based on Agglomerative Hierarchical Clustering [8]. An explainable pattern-recognition method based on domain-specific Shapelets was proposed, where both K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) were successfully used for classification [3].

Various event detection techniques were suggested for classifying temporal data from a timeseries perspective. Some reported results include feature-based techniques [9], shapelet subsequence matching methods [10], timeseries modeling [11], techniques with adapted distance matrices and kernels methods such as SVM with Dynamic Time Warping. Such techniques are representationally powerful but their applicability to big PMU measurements might be limited as they can suffer from slow execution times, curse of dimensionality and the need of time-consuming handcrafted features.

Several recent studies utilized Convolutional Neural Networks in grid fault detection. In a recent study [12], Pseudo continuous quadrature wavelet transform was used to generate featured scalograms, then the scalograms (treated as two dimensional images) were used to train CNN network for fault classification. ROCOF signal and the radial active shift (RAS) signal were utilized as indicators for frequency disturbance faults [13]. Images were constructed from the RAS and ROCOF signals and used to train a convolutional network for fault classification. Similar, the data were embedded as images and used as an input to convolutional networks [14]. Another approach of using CNNs is introduced in [15], where 2-dimentional filters were used on matrices representing the different signal types used in the study. Convolutional networks were trained to detect faults in a wide area measurement system. Also, a convolutional based model that relies on extracting features from the signals before using them in the convolutional network was introduced in [16].

C. Purpose and Novelty

The purpose of this study is to show how to utilize PMUs measurements along with the supplied event log to design an event detection system, which can be leveraged in an online mode for automated power systems fault detection. Our study shows how to overcome common problems with PMU measurements such as sparsity of the PMU measurements, data quality issues and the need for tedious and costly manual processing and feature engineering.

The novelty of the proposed approach is in designing end-to-end power systems fault detection with automated data reduction, de-noising techniques, and automated feature learning from measurements of a reduced set of PMUs.

TABLE I
COMPARISON OF THE PROPOSED TO ALTERNATIVE EVENT DETECTION MODELS. (√) REPRESENTS IF THE REFERENCED MODEL(S) HAS ONE OR MORE OF THE COMPARISON CATEGORIES

	Data Driven	Scalability	Data Representation
[3]			✓
[7, 10]	✓		✓
[8]			✓
[4, 5, 12, 13, 14, 15, 16]		✓	
Proposed	✓	✓	✓

D. Contribution and Advantages

In this study, six classification models are considered for fault detection. Three of the considered classifiers were variants of Convolutional Neural Network (CNN) architectures that consider a single type of input variable (voltage, current or frequency measurements) at all PMUs or all types together also at all PMUs. These models, named Single Channel CNN (SC-CNN), Parallel Channel Embedding CNN (PCE-CNN), and Simultaneous Channel Embedding CNN (SCE-CNN) were compared to traditional Logistic Regression (LR), Multi-layer Perceptron (MLP) and Support Vector Machine (SVM) using measurements from all PMUs. We propose a different approach to using CNN based models where automated de-nosing techniques are applied to the measurements and then parallel and concatenation-based convolutional models are utilized for fault detection, which is different from image based convolutional models and feature based convolutional models discussed earlier.

In the conducted experiments, CNN-based models outperformed traditional methods on one year out-of-sample outage detection over the entire grid, providing evidence that the convolution operations leveraged in CNN models seem to capture the sub-signal patterns relevant to fault detection missed by traditional approaches. The best AUC (0.83) was obtained from the multi-channel model SCE-CNN. The obtained results provide evidence that localized fault-related line outage can be detected with good and reliable performance using CNN based models.

The proposed method is a data driven approach, where manual feature engineering is avoided. Furthermore, this model is scalable. Also, it avoids using costly representations of data (such as images), which is usually associated with CNNs. Table I shows a comparison of alternative models using the aforementioned advantages.

The paper is organized as follows: after the Introduction, we explain the fault detection modeling in Section II, discuss data management problems in Section III, provide experimental results in Section IV, and summarize conclusions in Section V.

II. FAULT DETECTION MODELING

Two Convolutional Neural Network (CNN) architectures for fault detection from PMU-generated data streams are proposed, one being a generalization of the other.

Problem definition: Suppose that multiple PMU devices $P = \{p_1, p_2, \ldots, p_{|P|}\}$ in a system are ***measuring several***

variables (such as voltage magnitude, current magnitude, frequency, etc.) during a time period $[t-\Delta_1,\ t+\Delta_2]$. In this context, *each variable* measured by the PMU is called a channel. Given a multi-channel signal:

$$\mathbf{s}\ (t-\Delta_1,\ t{+}\Delta_2) = \left[\mathbf{s}^{(1)}\left(t-\Delta_1,\ t{+}\Delta_2\right), \ldots,\ \mathbf{s}^{(C)}\left(t{-}\Delta_1,\ t{+}\Delta_2\right)\right], \quad (1)$$

each $\mathbf{s}^{(c)}(t-\Delta_1,\ t+\Delta_2) \in \mathbb{R}^l$ is the signal summarizing the c -th variable (channel) across all PMUs, for $c = 1,\ldots,C$. For simplicity of notation, we will use $\mathbf{s}$ and $\mathbf{s}^{(c)}$ to represent $\mathbf{s}(t-\Delta_1,\ t+\Delta_2)$ and $\mathbf{s}^{(c)}(t-\Delta_1,\ t+\Delta_2)$, respectively.

Given $\mathbf{s}$, or one of its channels $\mathbf{s}^{(c)}$, the fault detection objective is to predict $y \in \{01\}$ indicating whether or not a fault occurred during $[t-\Delta_1,\ t+\Delta_2]$. The fault of interest in this study is line outage, but the methodology is applicable to a larger range of faults.

From a machine learning perspective, the aforementioned fault detection problem can be formulated as a binary classification problem, which has been thoroughly studied in the machine learning community. Given a collection of measurements, one may, in principle, apply some of the readily available classification models to differentiate faults from signals indicating normal operation. The traditional classifiers would treat the measurements taken at a certain timestep across the signals as independent from the measurements taken at other timesteps. However, the heterogeneity of the different types of PMU-generated measurements would be difficult to account for in traditional classification frameworks that typically assume homogeneous inputs. Thus, we propose adopting convolutional neural network-based approaches as they are inherently designed to learn from correlated heterogeneous measurements.

A. Single-Channel Convolutional Neural Networks (SC-CNN)

The objective of SC-CNN is to model a single-channel (i.e., one-dimensional) signal $\mathbf{s}^{(c)}$. The signal originating from a single channel suggests that it can only be of a certain type, i.e., it can contain measurements from a single variable (such as voltage magnitude, voltage angle, current magnitude, current angle, or frequency). For example, $\mathbf{s}^{(c)}$ can carry information only about the voltage magnitude across all PMUs in the system, but not about any other variable that was measured by the same PMUs.

Convolution layers: When analyzing only signals from one PMU channel, the patterns that characterize abnormal behavior (such as a fault-related line outage) are typically manifested in their sub-signals. To capture such patterns, the original signal $\mathbf{s}^{(c)}$ is transformed such that the sub-signal patterns are emphasized upon its transformation. For this purpose, $\mathbf{s}^{(c)}$ is passed through a so-called convolution layer. Namely, $\mathbf{s}^{(c)}$ is initially convolved with a 1-D convolution window (also referred to as kernel) $\mathbf{k}_j^{[1]}$ of size k_1 :

$$v_{i,j}^{[1]} = \sum_{w=1}^{k_1} s_{i+w}^{(c)} * k_{q,j}^{[1]}, \qquad \forall i\ = 1,\ldots,l-k_1+1\,. \quad (2)$$

Note that, in the above equation, the bias terms are omitted for simplicity.

This operation is performed for m_1 different kernels $\mathbf{k}_1^{[1]},\ldots,\mathbf{k}_{m_1}^{[1]}$, thus defining m_1 output filters, i.e.,

$$\mathbf{v}_j^{[1]} = \left[v_{1,j}^{[1]},\ \ldots,\ v_{l-k_1+1,j}^{[1]}\right]\ . \quad (3)$$

for each $j\ =\ 1,\ldots,m_1$.

Essentially, the convolved outputs $\mathbf{v}_j^{[1]}$ can be thought of as a summarization of the original signal, in which certain characteristic 'shapes' from $\mathbf{s}^{(c)}$ should be emphasized. Nevertheless, some problems may require capturing higher-order characteristics, beyond the shapes captured by the initial convolution layer. We consider fault detection to be one of these challenging problems, thus an additional convolution layer is introduced. More precisely, the outputs from the first convolutional layer are further convolved m_2 times:

$$v_{q,r}^{[2]} = \sum_{w=1}^{k_2}\sum_{j=1}^{m_1} v_{q+w,j}^{[1]} * k_{w,j,r}^{[2]}\ , \qquad \forall q\ =\ 1,\ldots,l-k_1-k_2+2\,. \quad (4)$$

where $\mathbf{K}_q^{[2]} = [k_{w,j,r}^{[2]}]_{w=1,j=1}^{k_2,m_1}$ is of size $k_2 \times m_1$, for each $r\ =\ 1,\ldots,m_2$. Similarly, as used in the first convolution layer, the bias terms are omitted from the notation for simplicity.

Output layer: The resulting hidden vector reshaped and mapped to a single neuron whose value is subsequently passed to a sigmoid activation function to obtain the fault occurrence $(y\ =\ 1)$ probability:

$$P\ (y|\mathbf{h}) =\ sigmoid\ (\mathbf{hw}) = \frac{1}{1+e^{-\mathbf{hw}}}\ , \quad (5)$$

where $\mathbf{w} \in \mathbb{R}^d$.

Learning: Provided a collection of signals $S = \{\mathbf{s}_1,\ldots,\mathbf{s}_N\}$, the parameters of the whole single-channel CNN architecture are determined such that the binary cross-entropy loss is minimized, that is

$$\sum_{n\ =\ 1}^{N} -y_n \log P\left(y_n|\mathbf{h}_n\right) - \left(1-y_n\right)\ \log\left(1-P\left(y_n|\mathbf{h}_n\right)\right) \quad (6)$$

B. Multi-channel Convolutional Neural Networks (MC-CNN)

Although SC-CNN transforms a single channel $\mathbf{s}^{(c)}$ of the input signals to capture sub-signal patterns that characterize abnormal behavior, such patterns might exist across multiple channels. Moreover, patterns relating to faults may not be evident from looking into a single channel, but rather across multiple channels of the same PMU. We also propose two concatenation-based CNN variants for fault detection across multiple channels.

Parallel Channel Embedding based Convolutional Neural Network (PCE-CNN): Each $\mathbf{s}^{(c)}$ is passed through a separate 1-D convolution layer to obtain its corresponding convolutional embeddings. The convolutional embeddings from all C channels are then concatenated in an extended vector. Subsequently,

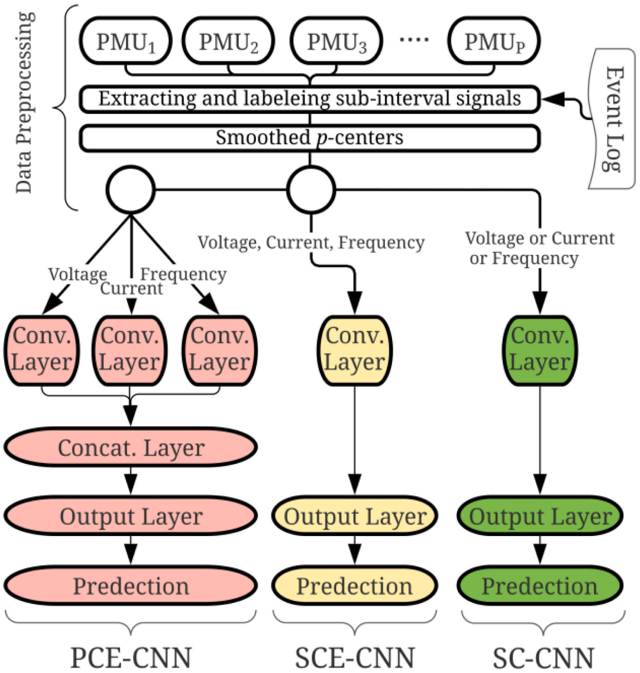


Fig. 1. Diagram representing the introduced model variants.

the convolved are reshaped and mapped to a single output to determine the probability of a fault occurrence.

Simultaneous Channel Embedding based Convolutional Neural Network (SCE-CNN): Instead of learning C parallel sets of convolutional embeddings, this variant concatenates the input signals from all channels $\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(C)}$ and passes their concatenation in a 2-D convolution layer. Subsequently, the convolved vectors are reshaped and mapped to an output layer which determines the fault occurrence probability conditioned on the original multi-channel measurement.

Fig. 1 represents an overview of the proposed models.

III. Data Management

A. Data Description

The dataset used in this study was collected from a set (P) of 37 PMUs from the Western interconnection in the United States. The dataset is completely anonymized to remove all critical energy infrastructure information including the network topology, the location of PMUs or events described in the event log (E). The PMU inputs are multiple voltage and current waveforms, and PMU outputs are phasor signals with multiple variables characterizing separate properties of the input waveforms such as magnitude, angle and frequency. Quite often the output phasors for three phases of voltages and currents are transformed into a single positive sequence voltage or current phasor. For each PMU $p \in P$, measurements were collected for 2 years (2016, 2017) at 30 to 60 frames per second (FPS). For each timestep t (in UTC), voltage, current and system frequency are reported. For voltage and current, both 3-phase and positive sequence values are reported, and both magnitude and angle are reported for the individual measurements.

Along with the measurements, an event log (E) is provided. In this event log, a certain number of events that happened in the grid are listed. For each $e \in E$, the start and end time stamps to a minute resolution is provided in some cases in addition to a description of the event type. The event log doesn't contain any information regarding the geographical location of the fault or the proximity of a fault to a certain PMU.

Several characteristics of this problem are in line with the defined fundamental characteristics of Big Data [17], such as Volume (the used dataset size is in the Terabytes), Velocity (Phasor data is reported at 30 or 60 FPS), Variety (several vintage points are measured by each PMU) and Validity (PMU data has several data quality issues such as missing data).

B. Data Preprocessing

The provided PMU measurements in its original format is very large and noisy, and not suitable for model training. To mitigate this, a series of data management steps were performed before proceeding with the model training and evaluation.

1) Extracting Labeled Sub-Interval Signals: The first data preprocessing step is data windowing, where the PMU measurements is split into sub-signals (windows) of 1 minute, using the UTC timestamps. Since the measurements are sampled at different sampling rates, the 1-minute sub-interval signals can vary in size, for example 1 minute at 30 FPS is equivalent to 1800 data points and 1 minute at 60 FPS is equivalent to 3600 data points. To unify the data, all sub-interval signals are down sampled (using averaging) to 30 FPS. Next step was to label the sub-signals. Labeling in this context refers to intersecting the sub-signal timesteps with event log (E) provided with the dataset. This step produces a binary label, which indicates if an event occurred at this sub-interval signal or not. The labeled sub-signals created in this process become hugely unbalanced due to inherited characteristics of PMU data where the majority of the data is not labeled as events (normal operations). This result rendered model training problematic. Furthermore, normal operation data is fairly stable (including some normal variation due to load adaptations and other electric grid operational properties). Given aforementioned characteristics of the data, not all data was necessary for model training. The temporal data around the event is considered the most informative, hence a time window around the event is kept and the remaining parts of the data are discarded. If an event is marked at timestamp t, a time period of $[t - \Delta_1,\ t + \Delta_2]$ is kept, where $\Delta_1,\ \Delta_2$ refers to two separate time differences before and after the event. By examining the data and consulting with the subject matter experts, values of 3 mins and 6 seconds, and 54 seconds were given for $\Delta_1,\ \Delta_2$ respectively allowing the total of time around the event to equal 4 minutes. This process of extracting the sub-signals is done individually for each $p \in P$. The resulting dataset contains different types of events. To keep the data homogeneous, line outage faults was the type of event considered in this study.

2) Computing Smoothed $\boldsymbol{p}$ *-Centers:* After performing the steps discussed earlier, each sub-signal is represented as a tensor of $1800 \times \#\, signals \times size\, of\, P.$ This representation introduces a computational challenge in model training and model inference, since each tensor representing 1 minute of PMU measurements (sub-signals from all PMUs) can have a large number of data points, for example, for 50 PMUs and only 3 signals the number of data points can reach 270000, which is challenging for model training and inference. Another issue inherited from the PMU data is that the fault's signature characteristics are not necessarily evident on all PMUs, which can be due to geographical factors (i.e., how far the faults are from the PMU). Also, this limitation can be related to a topological characteristic of the grid (for example, if the PMU is behind a power transformer, the effect of a fault can be diminished for such a PMU). This high dimensionality of noisy data can be an impediment for model training. One technique to mitigate this problem is to find a simplified representation of the data that maintains the information contained in such tensors but reduces the data size and noise.

The first step taken to reduce the data is down sampling. The technique used in down sampling should be chosen to preserve the information of the data. One characteristic of fault-related line outages is that a sudden disturbance is seen in the PMU measurements, the realization reached by manually examining a certain number of faults across different time periods. By correlating the event log with the collected measurements, sudden disturbances are usually seen when a fault-related line outage happens. In this case, to down sample the data and preserve such characteristic, down sampling is performed as following: each 1800 data point sub-signal is divided into a list of non-overlapping windows (W) of size 10, where for each $w \in W$, the range of data points is calculated. In the case of sub-signals of length 1800, this will reduce data to 180 points. The reason behind using a range to summarize the windows is that a range preserves the distortions and sudden jumps.

The second step is calculating p-centers, where each p-center is a sub-signal that represents all sub-signals for a specific time window across all $p \in P$. After calculating the p-center, each 1-minute window is represented by a 1 vector (sub-signal) that represents all $p \in P$. The process of creating the p-centers takes care of two issues. Firstly, it reduces the dimensionality of the data while keeping the useful information. Secondly, it emphasizes the change in the sub-signals produced by the PMUs that detected the fault, and this is achieved using a weighted approach, where the weights are extracted dynamically for each time window. Such a weighting scheme is necessary since there is no guarantee that there is a single PMU that will be significant for all faults. The significance of the PMU can be controlled by the geographical proximity of the PMU to the fault.

Since the geographical information was provided for neither the PMU nor the fault, another scheme must be adapted. To calculate smooth p-centers, the Differentiable Loss Function for Time-Series based on Soft-DTW [18] technique is used in our study. Several techniques are proposed in the literature to calculate centers of timeseries, such as Euclidean, Regularized Wasserstein Distance, and Dynamic Time Warping (DTW).

Such techniques tend to overfit and get influenced by noise. Instead, a smoothed formulation of DTW, namely Soft-DTW, is used. Soft-DTW considers all alignments and not just the optimal one, and computes in reasonable time by utilizing the fact that gradients of soft-DTW are differentiable w.r.t to all of its variables. This characteristic can be used to calculate centers of groups of timeseries which better summarizes the data and reduces the dimensionality. In the context of summarizing PMU data, Soft-DTW can calculate 1 vector (sub-signal) that represents all $p \in P$ for a certain time window and a certain signal. To preserve the characteristic that faults are not necessarily shown on all PMUs, and to minimize the chances of normal operation overcoming the fault information, weights are given to each vector (sub-signal) coming from each $p \in P$. Such weights cannot be predetermined as discussed before, thus; dynamic weights are needed. In this experiment, a weight relying on the standard deviation (σ) of the time window is used. The reasoning behind choosing σ is that despite that the signals have significant variation over long period of time, on a sub-subinterval signal level, the signals are less varying unless there is a fault. This step is performed separately for each channel.

IV. Experimental Evaluation

A. Prediction Models

Six models where used in the experiments. The introduced *Single-channel Convolutional Neural Network* (SC-CNN) using either voltage, current or frequency. The two introduced multi-channel models (*Parallel Channel Embedding* based CNN (PCE-CNN), *Simultaneous Channel Embedding* based CNN (SCE-CNN)). Also, three traditional models where used: *Logistic Regression* (LR), *Multi-layer Perceptron* (MLP) and Support Vector Machine (SVM).

B. Experimental Setup

1) Data Split: The preprocessed 24-month dataset described in Section III was split into training and test sets. Namely, to capture the seasonal patterns that span throughout the year, all prediction models (described in Section IV) were trained on preprocessed signals from 2016 and tested on signals from 2017. This resulted in a total number of 704 training signals (out of which 176 were fault-related line outages), while the number of signals for testing was 848 (212 of them being fault-related line outages).

We considered 3 channels across all signals: positive sequence voltage magnitude (V), positive sequence current magnitude (I), and frequency (f). Note that, once all channels are considered, the total number of measurements (i.e., the total number of samples over all signals) in the dataset was 838080; out of which 380160 were contained in the training set, while the test set contained the remaining 547920.

2) Number of PMUs: Measurements from all available PMUs are used for training and testing using the data split. Section IV Sub-section D discusses the effect of number of PMUs on model performance.

3) Model Parameters: In conducted experiments, an embedding dimension of 30 was used for the models that employ a fully-connected layer to learn dense embeddings before the final output layer (these include all models except for LR). SVM uses a linear kernel and a value of 1 for the Regularization parameter.

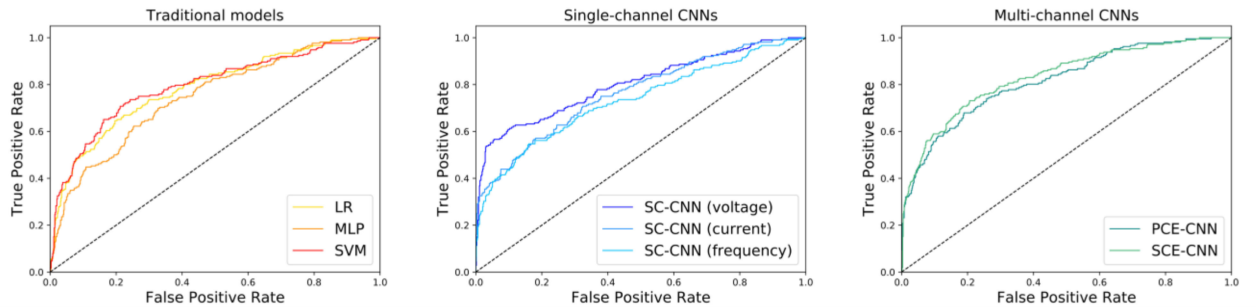


Fig. 2. ROC curves of traditional classifiers (left), single-channel CNNs (middle), and multi-channel CNNs (right).

TABLE II
FAULT DETECTION PERFORMANCE ACROSS MULTIPLE EVALUATION METRICS. THE LARGEST METRIC VALUES ARE BOLDED, WHILE '$*$' INDICATES COMPARABLE PERFORMANCE. FOR TRADITIONAL MODELS, WE REPORTED ONLY FOR BEST PERFORMING INPUT (V)

Models		Input	Accuracy	AUC
Traditional	LR	V	0.6320	0.7846
	MLP	V	0.6615	0.7455
	SVM	V	**0.8066***	0.7937
Single channel	SC-CNN	V	0.7464	0.8016*
	SC-CNN	I	0.7228	0.7654
	SC-CNN	f	0.6851	0.7305
Multi-channel	PCE-CNN	V, I, f	0.7971*	0.8151*
	SCE-CNN	V, I, f	0.7971*	**0.8316***

As for the CNN variants (SC-CNN, PCE-CNN, SCE-CNN), in the convolutional layer, a 1-D kernel of size 30 was used with a stride of 30. The number of output filters (i.e., the dimensionality of the convolutional embeddings) was set to 150. All prediction models were implemented in Python 3.7 and run on a Linux machine with 64GB of memory and a 20-core Intel(R) Xeon(R) Gold 6230 CPU @ 2.10GHz.

4) Evaluation Metrics: To evaluate the degree to which the models' predictions match the original fault occurrence labels, Accuracy (ratio of correctly classified fault-related outages) was measured. As for the relevance of the fault probability scores associated with the predictions, the Area under the Receiver Operating Characteristic curve (AUC) was calculated. Note that the values of both metrics range from 0 to 1, such that larger values indicate higher classification (and thus fault detection) performance.

C. Fault Detection Performance

The performances of the prediction models for the task of detecting faults were evaluated and compared using the classification metrics and data split outlined earlier. All available PMUs are utilized. The results obtained from this evaluation are presented in Table II where for LR, MLP and SVM the results are reported only for experiments based on voltage since LR, MLP and SVM showed best performance on voltage.

Discussion: From Table II, it can be observed that CNN-based variants perform comparable to, or in most cases, better than traditional classification models. This suggests that the convolution operations leveraged in CNN models seem to capture the sub-signal patterns relevant to fault detection.

Among the CNN variants, the multi-channel CNNs trained yield the largest accuracy and AUC. The corresponding ROC curve plots are presented in Fig. 2.

Between the single channel models, SC-CNN trained on voltage produces the highest accuracy and AUC. This is expected since the voltage signals would typically experience a visible drop in magnitude during the fault that can be observed from multiple PMUs in the vicinity of the fault, thus providing a strong identifier of the fault. This makes voltage the best candidate to detect faults. Otherwise, if the PMU placement was not as sparse, the rapid change in current would be much more prominent indicator of a fault. On the other hand, the impact on frequency at the rest of the network (outside of faulted line) may not be as prominent due to the frequency being highly regulated as a global property of the system. As mentioned earlier, the impact on the current is very predictable on the line where the fault has occurred, where the current goes high and then drop to zero when the fault is cleared. However, the impact on current magnitude at other locations in the network where PMU measurements are taken can be different, without a specific pattern that the algorithm can exploit for accurate prediction. For all faults, only one of the PMUs or none is near the fault location. Thus, the percentage of PMUs that would provide current measurement with precise characteristics is exceedingly small.

The multi-channel CNNs, on the other hand, further improve CNN-SC (voltage) accuracy by large margins ($\sim$5%) in additional to AUC lifts. Considering that both multi-channel variants obtain comparable performance, our suggestion is for PCE-CNN to be used when a large number of parameters is not an issue. In contrast, one should use SCE-CNN in case a slight decrease in performance is allowed for the benefit of having a simpler model. Additionally, training and inference times for CNN based model are within practical ranges. Data reduction techniques helped keeping training time under 20 seconds and inference times under a second, using the hardware described earlier. Overall, although voltage signal was the most informative for predicting fault-related line outages, multi-channel CNNs achieved the best trade-off between accuracy and AUC.

For each of the eight classifiers considered in this study the top 10 misclassified cases by each classifier were visually

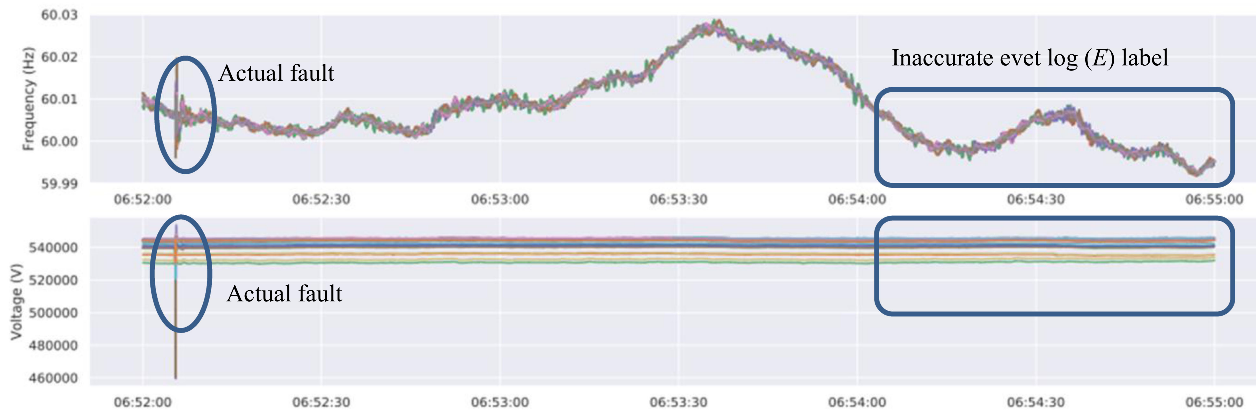


Fig. 3. Frequency and Voltage during the misclassified time window.

TABLE III
VISUAL INSPECTION OF THE TOP 10 MISCLASSIFICATIONS BY EIGHT MODELS

Models	Cases of misclassification	
	False labels	False predictions
LR, MLP, SVM	9	1
SC-CNN (voltage)	9	1
SC-CNN (current)	7	3
SC-CNN (frequency)	6	4
PCE-CNN	8	2
SCE-CNN	9	1

inspected to determine the potential cause of wrong classification. In Table III faults marked as false labels were not evident in PMU measurements but were labeled as fault-related line outage cases in the event log, while faults marked as false predictions were misclassified cases where visual inspection verified that these were actual faults. For a majority of the cases the fault was not present in the measurements according to the visual inspection. The reason for occurrence of these cases that appear misclassified but are correctly classified by the algorithms is that event log labels used for the study were not fully accurate. This is a consequence of the labels extracted from the utility's event log not accurately representing the start of the fault, nor the duration of the fault's visibility in the PMU measurements.

For example, the fault illustrated at Fig. 3 has started at 06:52:05 and effectively lasted until 06:52:08. However, the event log places this fault at 06:54. Thus, the label for time window between 06:54 and 06:55 was set to a value of 1 (fault) according to the event log, which is wrong since the complete fault ended before 06:54. During the time window 06:54 to 06:55 there is no visible fault as can be confirmed from Fig. 3.

D. The Effect of Number of PMUs on Model Performance

This section examines the effects of the number of PMUs on model performance. Fig. 4 shows the AUC for different number of PMUs using SCE-CNN. For each number of PMUs, 10 different random selections of PMUs were performed, then the model is trained and tested separately on each set of PMUs.

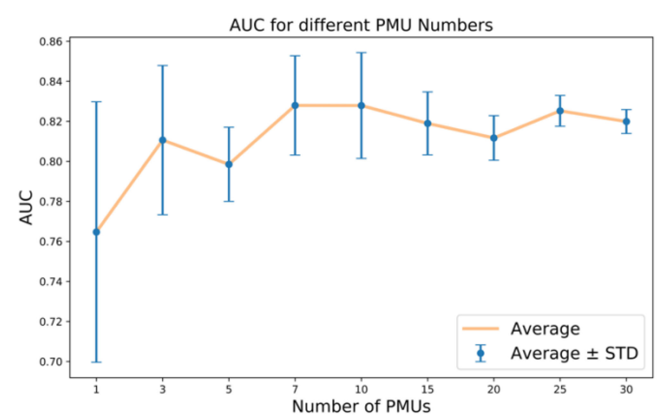


Fig. 4. Reported AUC for different number of PMUs.

Fig. 4 shows that the variation of performance (shown in the standard deviation of AUC) decreases as the number of PMUs increase. This behavior is expected since fault can be observed from multiple PMUs in the vicinity of the fault. Since the PMUs are randomly selected without a prior knowledge of any spatial information or the proximity to the fault locations, the choice of PMU can affect the performance of the model. From Fig. 4, one can see that the performance of the model stabilizes at the range of 20 to 25 PMUs. This suggests that a smaller number of PMUs can be used with an acceptable model accuracy.

V. CONCLUSION

We implemented an end-to-end fault detection model for the electric grid using measurements from a reduced set of Phasor Measurement Units (PMUs). Several results were achieved:

- Novel and effective data preprocessing techniques are used to lower data dimensionality and reduce noise
- The proposed model is compared to three base models: Single Channel CNN (SC-CNN), Parallel Channel Embedding CNN (PCE-CNN), and Simultaneous Channel Embedding CNN (SCE-CNN). Each model introduced a different flavor of utilizing the data.

- The impact of the selected models is shown by utilizing a two-year recording of PMU measurements collected from a reduced set of PMUs, along with their event log. Through automated data preprocessing and reduction techniques, the models detect fault-related line outage faults with good performance.
- Convolutional Neural Networks based models achieved the best performers. Faults were predicted with good performance (AUC of 0.8016 for SC-CNN, 0.8151 for PCE-CNN and 0.8316 for SCE-CNN). The best performance was obtained using SCE-CNN.
- The experiments show the effects of the quality of the event log on the prediction results by manually examining a certain number of events (Table III). An acceptable accuracy can be stabilized using a lower number of PMUs than originally available.

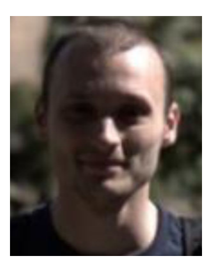
VI. Disclaimer

This paper 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.

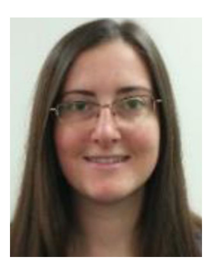
References

[1] C. Aggarwal. *Data Mining: The Textbook*. Berlin, Germany: Springer, 2015, pp. 293–333.
[2] North American SynchroPhasor Initiative, "Data mining techniques and tools for synchrophasor data," NASPI White Paper, NASPI-2018-TT-007, pp. 293–333, Jan. 2019.
[3] M. Biswal, Y. Hao, P. Chen, S. Brahma, H. Cao, and P. DeLeon, "Signal features for classification of power system disturbances using PMU data," in *Proc. Power Syst. Comput. Conf.*, 2016, pp. 1–7.
[4] D. Kim, T. Chun, S. Yoon, G. Lee, and Y. Shin, "Wavelet-based event detection method using PMU data," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1154–1162, May 2017.
[5] M. Biswal, S. Brahma, and H. Cao, "Supervisory protection and automated event diagnosis using PMU data," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1855–1863, Aug. 2016.
[6] L. Xie, Y. Chen, and P. R. Kumar, "Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.
[7] M. Rafferty, L. Xueqin, D. Laverty, and S. McLoone, "Real-time multiple event detection and classification using moving window PCA," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2537–2548, Sep. 2016.
[8] O. Dahal, S. Brahma, and H. Cao, "Comprehensive clustering of disturbance events recorded by phasor measurement units," *IEEE Trans. Power Del.*, vol. 29, no. 3, ppp. 1390–1397, Jun. 2014.
[9] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, "Feature-based classification of time-series data," *Int. J. Comput. Res.*, vol. 10, no. 3, pp. 49–61, 2001.
[10] L. Ye and E. Keogh, "Time series shapelets: A novel technique that allows accurate, interpretable and fast classification," *Data Mining Knowl. Discov.*, vol. 22, no. 1/2, pp. 149–182, 2011.
[11] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances," *Data Mining Knowl. Discov.*, vol. 31, no. 3, pp. 606–660, 2017.
[12] S. Wang, P. Dehghanian, and L. Li, "Power grid online surveillance through PMU-embedded convolutional neural networks," *IEEE Trans. Ind. Appl.,* vol. 56, no. 2, pp. 1146–1155, Mar./Apr. 2020.
[13] W. Wang *et al.*, "Frequency disturbance event detection based on synchrophasors and deep learning," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3593–3605, Jul. 2020.
[14] 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. Power Syst.*, vol. 34, no. 6, pp. 4640–4651, Nov. 2019.
[15] M. Kesici, C. Saner, M. Mahdi, Y. Yaslan, and V. Genc, "Wide area measurement based online monitoring and event detection using convolutional neural networks," in *Proc. 7th Int. Istanbul Smart Grids Cities Congr. Fair*, 2019, pp. 223–227.
[16] W. Li and M. Wang, "Identifying overlapping successive events using a shallow convolutional neural network," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4762–4772, Nov. 2019.
[17] M. Kezunovic, P. Pinson, Z. Obradovic, S. Grijalva, T. Hong, and R. Bessa, "Big data analytics for future electricity grids," *Electric Power Syst. Res.*, vol. 189, 2020, Art. no. 106788.
[18] M. Cuturi and M. Blondel, "Soft-DTW: A differentiable loss function for time-*series*," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 894–903.

Mohammad Alqudah (Student Member, IEEE) received the B.Sc. degree in computer engineering from the Jordan University of Science and Technology, Irbid, Jordan, in 2012, and the M.S. degree in industrial and systems engineering from Binghamton University, Binghamton, NY, USA, in 2015. He is currently working toward the Ph.D. degree in computer and information sciences with the Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA, USA. His main research interests include machine learning for spatio-temporal and graph data.



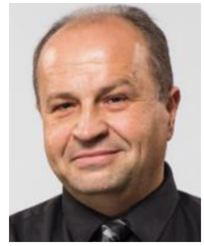
Martin Pavlovski (Student Member, IEEE) received the B.Sc. degree in electrical engineering and information technologies from the Saints Cyril and Methodius University of Skopje, Skopje, Macedonia, and the Ph.D. degree from Temple University, Philadelphia, PA, USA, in 2021. He was a Research Intern with the Macedonian Academy of Sciences and Arts, and a Visiting Scholar with the Center for Data Analytics and Biomedical Informatics, Temple University. His research focuses on machine learning from structured data.



Tatjana Dokic (Member, IEEE) received the M.Sc. degree in electrical and computer engineering from the University of Novi Sad, Novi Sad, Serbia, in 2012, and the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA, in 2019. She is currently an Assistant Research Engineer with Texas A&M University. Her main research interests include big data for power system applications, power system asset and outage management, weather impacts on power systems, time series analysis of PMU measurements, and smart grids.

Mladen Kezunovic (Life Fellow, IEEE) received the Dipl. Ing. from the University of Sarajevo, Sarajevo, Bosnia, in 1974, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Kansas, Lawrence, KS, USA, in 1977 and 1980, respectively. Since 1986, he has been with Texas A&M University, College Station, TX, USA, where he is currently the Regents Professor, Eugene E. Webb Professor, and the Site Director of Power Engineering Research Center consortium. For more than 25 years, he has been the Principal Consultant of XpertPower Associates, a consulting firm specializing in power systems data analytics. He has authored more than 600 papers, given more than 120 seminars, invited lectures, and short courses, and consulted for more than 50 companies worldwide. His expertise is in protective relaying, automated power system disturbance analysis, computational intelligence, data analytics, and smart grids. Dr. Kezunovic is a CIGRE Fellow, Honorary and Distinguished Member. He is a Registered Professional Engineer in Texas.

Yi Hu (Fellow, IEEE) received the B.Sc. degree in electrical engineering from Southeast University, Nanjing, China, in 1982, and the M.Sc. degree from Nanjing Automation Research Institute, Nanjing, China, in 1984, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Winnipeg, MB, Canada, in 1994. He was with NARI, ABB, TEKELEC, Catapult, and KEMA Consulting. Since 2008, he has been with Quanta Technology. He holds 13 U.S. patents and has authored more than 30 papers. His expertise is in power system protection and control, including system integrity protection schemes, advanced technology development and implementation, such as synchronized measurement technology and wide-area monitoring, protection and control (WAMPAC) systems, power system modeling/analysis, big data analytics, standard compliance testing and quality assurance.



Zoran Obradovic (Senior Member, IEEE) is currently a Distinguished Professor and a Center Director with Temple University, College Station, TX, USA, an Academician with the Academia Europaea (the Academy of Europe), and a Foreign Academician with the Serbian Academy of Sciences and Arts. He mentored 45 postdoctoral fellows and Ph.D. students, many of whom have independent research careers at academic institutions and industrial research labs. He is the Editor-in-Chief of the *Big Data Journal* and the Steering Committee Chair of the SIAM Data Mining Conference. He is also an editorial board member at 13 journals and was the general chair, program chair, or track chair for 11 international conferences. His research results were published at about 400 data science and complex networks articles addressing challenges related to big, heterogeneous, spatial and temporal data analytics motivated by applications in healthcare management, power systems, earth and social sciences.