Power Grid Online Surveillance Through PMU-Embedded Convolutional Neural Networks

Shiyuan Wang [ID], *Student Member, IEEE*, Payman Dehghanian [ID], *Member, IEEE*, and Li Li, *Student Member, IEEE*

Abstract—**Power grid operation continuously undergoes state transitions caused by internal and external uncertainties, e.g., equipment failures and weather-driven faults, among others. This prompts an observation of different types of waveforms at the measurement points (substations) in power systems. Modern power systems utilize phasor measurement units (PMUs) and intelligent electronic devices embedded with PMU functionality to capture the corresponding peculiarities through synchrophasor measurements. However, existing PMU devices are equipped with only one synchrophasor estimation algorithm (SEA) and are, thus, not always robust to handle different types of signals across the network. This article proposes a PMU-embedded framework that ensures real-time grid surveillance and potentially enables adaptive selection of preinstalled SEAs in the PMU. Therefore, it ensures high-fidelity measurements at all times and irrespective of the input signals. Our proposed framework consists of: 1) a pseudocontinuous quadrature wavelet transform which generates the featured scalograms and 2) a convolutional neural network for event classification based on the extracted features in the scalograms. Our experiments demonstrate that the proposed framework achieves high classification accuracy on multiple types of prevailing events in power grids, through which an enhanced grid-scale situational awareness in real time can be realized.**

Index Terms—**Convolutional neural network (CNN), feature extraction, phasor measurement unit (PMU), waveform classification, wavelet transform (WT).**

I. INTRODUCTION

WITH the widespread deployment of synchrophasor technology in modern power grids, system monitoring and control settings have been revolutionized into a new era with high-resolution measurements [1]–[3]. Synchrophasor measurements, captured across the network via phasor measurement units (PMUs), have transformed many applications, e.g., power system model validation, state estimation, dynamic stability, online monitoring, protection and control functions, and postevent analysis [4], [5]. IEEE standard C37.118.1-2011 [6] has defined the PMU expected outputs, i.e., magnitude, phase angle, frequency, and rate of change of frequency (ROCOF), and the corresponding desirable accuracy. These outputs are obtained from synchrophasor estimation algorithms (SEAs) which are primarily driven by mathematical approximations. In most cases, and irrespective of the focused application, marketplace PMUs are typically furnished with only one SEA tool, each unleashing distinctive advantages and limitations, solely valid to one or a few certain applications [7]–[9]. Typically, the waveforms fed into PMUs have variant behaviors; for instance, phasor magnitudes and phase angles go through step changes during faults, and the waveform measurements could be noisy. Besides, unbalanced load, voltage surge or sag, harmonics, and frequency drift are also common phenomena in electrical power networks [10]–[12]. In dealing with the above-mentioned conditions, dynamic SEAs based on time-domain signal processing techniques were applied. Some research efforts have proposed a single PMU equipped with only one sophisticated SEA, which is deemed to respond to various prevailing conditions [13]–[15]. Laboratory tests and field observations have revealed how inefficient the PMU measurements could be, if this "one-size-fits-all" SEA is applied to capture both static and dynamic features and peculiarities [16]–[18]. To meet the growing demand for high-speed, low-latency, and yet absolutely accurate measurements from PMU sensors, a more efficient mechanism that provides online event detection and assists in selecting the right SEA at the right time is desired.

In pursuit of mechanisms for power grid event detection, a power quality-based event detection approach is proposed in [19], which utilizes an advanced metering infrastructure (AMI) network in the smart grid and is featured with a fault tolerant capability. However, this approach requires a grid-scale AMI installation. A multiple-event analysis approach though cluster-based sparse coding method is presented in [20], which achieves a wide-area situational awareness. However, this multiple-event detection scheme is only examined on three outage scenarios (generator trip, load trip, and line trip). A grid-scale situational awareness approach using sparse unmixing technique is introduced in [21], which only focuses on detecting the "trip events" and takes seconds to achieve the adequate results. A more advanced real-time event detection mechanism using energy similarity measure (ESM) on wide-area measurements is proposed in [22], which can swiftly detect multiple types of events across the power grid. However, this mechanism requires full observation of the grid through PMU measurements at each

bus, and depends on reliable communication channels. A method for detection and classification of multiple events through principal component analysis (PCA) of frequency measurements is proposed in [23], which only detects loss of generation and/or loads. To the authors' best knowledge, there is no scheme available in the literature that can handle multiple event detection and classification, particularly by relying only on the original waveforms captured on a single measurement point.

With the development of artificial intelligence (AI), many AI-based approaches have been proposed for adaptive sensing and control in power systems and for improved resilience [24], [25]. It has been shown that the machine learning techniques, such as neural network-based reinforcement learning (RL) [26], [27], can successfully handle the nonlinearity and unknown dynamics of the systems [28], [29]. In [30], a deep RL method is proposed for nonlinear derivation of control strategies for unmolded power systems. An RL-based maximum power point tracking (MPPT) algorithm is presented in [31] to optimize the rotor speed and electrical power of the permanent-magnet synchronous generators (PMSG). An RL-based optimal control for hybrid energy storage systems is introduced in [32]. Also, RL can be used for detecting false data injection attacks in power grid's voltage control systems [33]. A cooperative RL algorithm is proposed in [34] that solves the economic dispatch problem in microgrids with distributed energy resources.

Motivated by the successful implementation of machine learning techniques in power systems and the fact that transitions of events in the power grid could be revealed through waveform patterns in the voltage and current signals [35], [36], we propose a novel online surveillance framework that classifies various events. In the proposed framework, the waveform patterns and signatures are extracted via a time–frequency domain analytic—wavelet transform (WT) [37], [38]. Built on the WT-extracted features, i.e., scalograms, a machine learning mechanism based on convolutional neural networks (CNNs) [39], [40] is employed to detect and classify the events through pattern recognition process. This article's main contributions are highlighted as follows.

1) We propose a pseudocontinuous quadrature WT (PCQ-WT) that effectively captures the power waveform features corresponding to different events.
2) A CNN mechanism is developed to classify the scalograms generated by PCQ-WT, achieving a high event classification accuracy (92.84 ± 1.20%) for 12 distinct modes of grid prevailing conditions.
3) The combination algorithms of the two steps process the input signals in real time (2.24 ± 0.54 ms), which leaves a large time margin to the subsequent SEA modules within PMUs to function.
4) The proposed framework shares the exact same input signals of the market-place PMUs with no additional investment; thus, it is an economically viable technology to be embedded within the existing PMUs.

This article is organized as follows. Section II introduces a background on continuous WTs (CWTs) and the pattern classification through CNNs. Section III details the proposed online surveillance framework, consisting of: 1) feature extraction from voltage and current signals via PCQ-WT and 2) event classification via CNN. Case studies and experiment results are analyzed in Section IV, and the conclusions are presented in Section V.

II. BACKGROUNDS AND MOTIVATIONS

A. Power Waveform Modeling

To solve a classification problem, a mathematical representation of signals in power grid is needed. The three-phase time-domain sinusoidal signals fed into the PMU can be represented by the following:

$$x_{ph}(t) = A_{ph}(t)\cos\left(2\pi\int_0^t F(\tau)d\tau + \phi_{ph}(t)\right) \quad (1)$$

where $x_{ph}(t)$ is a one-dimensional (1-D) waveform measured from each phase; $A_{ph}(t)$, $F(\tau)$, and $\phi_{ph}(t)$ are the instantaneous magnitude, fundamental frequency, and phase angle in each phase, respectively. During both transient and steady-state operations, the waveform in each phase can be expressed by a summation of different orders of harmonic components. Thus, the actual waveform in each phase is

$$x_{ph}(t) = \sum_{h=1}^{H} A_{ph,h}(t)\cos\left(2\pi\int_0^t F_h(\tau)d\tau + \phi_{ph,h}(t)\right) \quad (2)$$

where h is the order of harmonics, and H is the maximum order of harmonics of interest. In a particular grid operation condition, different values of $A_{ph,h}(t)$, $F_h(\tau)$, and $\phi_{ph,h}(t)$ will appear in the three-phase power signal, which lead to different patterns and peculiarities. Therefore, an event can be detected and classified accordingly. To simplify the time-domain sinusoidal signal analysis without acquiring the rotating reference frame in Park transformation [41], Clarke transformation is applied to convert the three-phase signal from ABC to $\alpha\beta$-frame [42] by the following equations:

$$x_{\alpha\beta}(t) = x_\alpha(t) + jx_\beta(t) \quad (3)$$

and

$$\begin{bmatrix} x_\alpha(t) \\ x_\beta(t) \end{bmatrix} = \begin{bmatrix} \dfrac{2}{3} & -\dfrac{1}{3} & -\dfrac{1}{3} \\ 0 & \dfrac{\sqrt{3}}{3} & -\dfrac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} x_A(t) \\ x_B(t) \\ x_C(t) \end{bmatrix}. \quad (4)$$

Since power waveforms contain different frequency components, multiresolution waveform analysis techniques are suitable to extract the features, i.e., amplitude, frequency, and phase angle. The most commonly used technique is short-time Fourier transform (STFT) [43], [44] and wavelet analysis [45]–[47]. Studies show that the STFT has a higher computational burden, which leads to a lower time resolution than the wavelet analysis in time–frequency domain [48], [49]. Also, comparing the spectrograms in Fig. 1(a) and (b) with the scalograms of WT in Fig. 1(c)–(f), it can be seen that STFT outperforms CWT in frequency accuracy, while CWT provides more conspicuous results for feature extraction. Therefore, wavelet analysis is chosen as the mathematical tool of interest for feature extraction and online waveform monitoring.
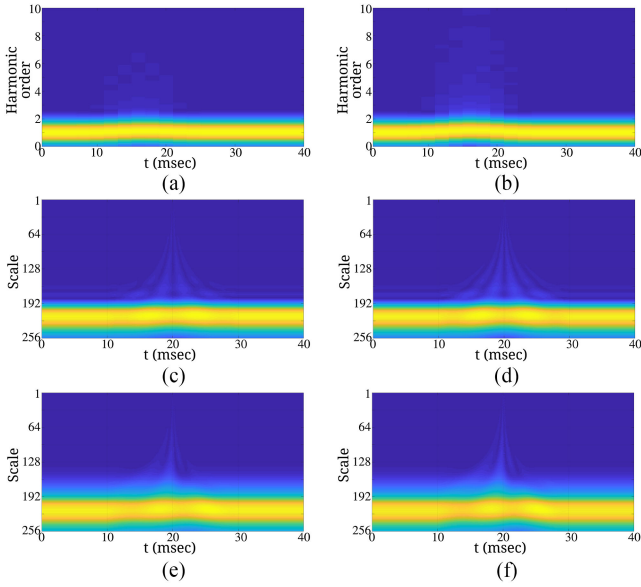
Fig. 1. Comparison of the STFT versus Morlet CWT versus DB4 CWT during: (a), (c), and (e) 2 Hz frequency jump; (b), (d), and (f) 40° phase jump. (a) STFT (b) STFT (c) Morlet (d) Morlet (e) DB4 (f) DB4.

B. CWT and Pseudo-CWT (PCWT)

The WT is obtained by computing the cross correlation between the signal of interest $x_{\alpha\beta}(t)$ and designated wavelets. This process is defined as follows:

$$X_{\alpha\beta}(\omega|a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x_{\alpha\beta}(t)\Psi^*\left(\frac{t-b}{a}\right) dt \quad (5)$$

where $\Psi(t)$ is the mother wavelet, * denotes the complex conjugate, a and b are the scaling factors and the time shift, and $\Psi(\frac{t-b}{a})$ is one of the "daughter wavelets" of $\Psi(t)$ [38]. With different selections of a and b, a wavelet bank is then determined. By selecting proper intervals for the *continuous* scaling factor along with the time shift, a CWT is achieved [40]. In a PMU, the real-time signals are sampled, and discrete signal processing is actually applied. Due to the limited computational capacity of hardware, the number of scaling factors are finite; therefore, the mathematical behavior of the CWT within the processor is *pseudo*continuous with a set of discrete noninteger scaling factors. So, here, the PCWT, with one of the discrete scaling factors, is defined as follows:

$$X_{\alpha\beta}[\omega|a_k,b_k] = \frac{1}{\sqrt{|a_k|}} \sum_{n=0}^{W-1} x_{\alpha\beta}[n]\Psi^*\left[\frac{nT_s - b_k}{a_k}\right] \quad (6)$$

where T_s denotes the sampling interval and W stands for the window (buffer) length. In a CWT, each daughter wavelet needs to cover a designated frequency range that reflects the features in time–frequency analysis. The central frequency of the daughter wavelets can be approximated by the following relationship with the scaling factor:

$$f = F_c/a_k \quad (7)$$

where F_c is the central frequency of the mother wavelet [37]. When a vector of scaling factors with length K is chosen, the wavelet bank Ψ and the extracted features at time instant n are expressed by the following:

$$\Psi^{K \times W} = \begin{bmatrix} \Psi\left[\dfrac{nT_s - b_1}{a_1}\right] \\ \vdots \\ \Psi\left[\dfrac{nT_s - b_k}{a_k}\right] \\ \vdots \\ \Psi\left[\dfrac{nT_s - b_K}{a_K}\right] \end{bmatrix} \quad \mathbf{X}_{\underline{\omega}}^{K \times 1}[n] = \begin{bmatrix} X_{\omega 1}(a_1, b_1) \\ \vdots \\ X_{\omega k}(a_k, b_k) \\ \vdots \\ X_{\omega K}(a_K, b_K) \end{bmatrix}.$$

$$(8)$$

To cover a sufficiently wide range of frequency and provide adequate pattern information for time–frequency analysis, the central frequency of mother wavelets F_C and the largest scaling factor a_K must satisfy the following condition:

$$\frac{F_c}{a_K} < f_0 < F_c \quad (9)$$

where f_0 is the frequency of the signal of interest. F_c needs to be chosen from a higher frequency range than the maximum frequency of interest and scaled down by a_k. Hence, as a_k increases, the corresponding frequency of the PCWT output decreases in the frequency domain.

Once the scaling factors are chosen, then (8) is able to generate the wavelet bank and a series time bin of $\mathbf{X}_{\underline{\omega}}^{K \times 1}$ along the time instant, i.e., a scalogram of PCWT is achieved.

C. Convolutional Neural Networks

At this stage, we consider the obtained scalogram as 2-D images, and the process of event classification turns to an image classification. The conventional paradigm for image classification is to manually design the feature extractor and reduce the dimensionality of the data; the second phase is to employ a classifier to classify the lower dimensional features. This paradigm highly depends on the design of the feature extractor, where manually designing features for a complex task requires a great deal of human time and effort; it can take decades for an entire community of researchers. In contrast, CNNs are able to learn the feature extractor automatically and have been proven very successful in broad image-related tasks [39]. By definition, CNNs are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers [50]. In general, the convolution implementation is through cross correlations as defined by

$$\boldsymbol{s}^p(m,n) = \sum_u \sum_v \sum_w \boldsymbol{I}^u(m+v,n+w)\boldsymbol{K}^p(v,w) \quad (10)$$

where $\boldsymbol{s}^p(m,n)$ is the output of the convolutional layer at position (m,n) and pth channel, $\boldsymbol{I}^u$ is the uth channel of the image/data volume, and $\boldsymbol{K}^u$ is the uth convolutional kernel. A complex convolutional layer is composed of a small number of

complex layers [50] and is expressed by the following:

$$I_l = \text{pool}\left(\sigma(s)\right) \tag{11}$$

where I_l represents the output volume of the lth layer, $\sigma(\cdot)$ is the nonlinearity of the neurons, and $\text{pool}(\cdot)$ is a down sampling procedure. By stacking the convolutional layers, the abstraction capacity of the network generally increases.

The representations of the last convolutional layer are expanded to vectors and processed by the general fully connected layers. This transforms the representations with more nonlinearities into spaces with different (higher or lower) dimensions. The final layer of a CNN usually reduces the dimensionality of the representations to the number of classes; cross entropy [51] is then employed to measure the "goodness" of the classification (Kullback–Leibler divergence between the predicted distribution and the target distribution). Specifically, suppose a scalogram X_ω is mapped by the network to its representation (scores) $s = \mathcal{F}(\in \mathcal{X}_L) = [s_1, s_2, \ldots, s_N]^\mathsf{T}$, where $\mathcal{X}_L$ is the set of training samples labeled by L. The softmax nonlinearity is used to normalize the output s to a probability distribution $O = [o_1, o_2, \ldots, o_N]^\mathsf{T}$, where o_i is defined as [51]

$$o_i = P(c_p = i | X_{\underline{\omega}} \in \mathcal{X}_L) = \frac{e^{s_i}}{\sum_{l=1}^{N} e^{s_l}} \tag{12}$$

where c_p denotes the predicted label. Cross-entropy loss is defined by

$$J = -\log(o_L) = -\log \frac{e^{s_L}}{\sum_{l=1}^{N} e^{s_l}}. \tag{13}$$

Finally, gradients of the cross-entropy loss function with respect to the parameters in the CNN are used to train the CNN by gradient descent algorithm

$$K_{(q+1)} = K_{(q)} + \lambda \nabla_{K_{(q)}} J(K_{(q)}) \tag{14}$$

where K is the parameter of the CNN, λ is the learning rate, and q denotes the iteration number.

III. PROPOSED FEATURE EXTRACTION AND EVENT CLASSIFICATION BY CNNs

A. PCQ-WT-Based Feature Extraction

Gabor wavelets have been widely used in 2-D pattern recognition [52]–[54]. In order to simplify the design and increase the computational efficiency, a modified complex Gabor wavelet from [54] is adopted in this article and is written as follows:

$$\Psi(t) = \underbrace{\exp\left(j\omega_c(t-b)\right)}_{\text{Periodic component}} \cdot \underbrace{\exp\left(-\frac{(t-b)^2}{\alpha_0^2}\right)}_{\text{Gaussian envelope}} \tag{15}$$

where w_c is the central frequency. The Fourier transform of this Gabor wavelet is

$$\mathbf{F}_\Psi(\omega) = \alpha_0 \sqrt{\pi} \cdot \exp(-j\omega b) \cdot \exp\left(-\frac{\alpha_0^2}{4}(\omega - \omega_c)^2\right). \tag{16}$$

One can see that the Fourier transform of the Gabor wavelet is also a function on the theme of Gabor wavelet, although not

following the orthogonal property, since

$$|\mathbf{F}_\Psi(\omega_c \pm \epsilon)| \neq 0 \tag{17}$$

where ϵ is a small value; according to (16), this Gabor wavelet has a characteristic of predictable narrow bandwidth. By properly selecting α_0, one of the PCQ-WTs can cover a desired range of frequencies. And the time shift b plays no magnitude impact on (16) and (17). Therefore, to ease the derivation, let $b = 0$, then the CWT of the Gabor wavelet turns into

$$X_{\alpha\beta}(\omega_0 | a, b = 0) = \int_{-\infty}^{\infty} x_{\alpha\beta}(t) \Psi^*\left(\frac{t}{a}\right) \mathrm{d}\tau$$
$$= \int_{-\infty}^{\infty} \exp\left(j\left(\omega_o - \frac{\omega_c}{a}\right)t - \frac{t^2}{a^2\alpha_0^2}\right) \mathrm{d}\tau. \tag{18}$$

According to the Hubbard–Stratonovich transformation [55]

$$\exp\left(-\frac{\alpha}{2}x^2\right) = \sqrt{\frac{1}{2\pi\alpha}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\alpha} - jxy\right) \mathrm{d}y. \tag{19}$$

The CWT of Gabor wavelet in (18) becomes

$$X_{\alpha\beta}(\omega_0 | a, b = 0) = a\alpha_0\sqrt{\pi}\exp\left(-\frac{\alpha_0^2}{4}(a\omega_0 - \omega_c)^2\right). \tag{20}$$

It can be seen that when $\omega_0 = \omega_c/a$, (20) reaches its maximum and the dominant feature of the expected frequency is revealed. To make each frequency of interest share an equivalent maximal magnitude, make

$$a\alpha_0 = \omega_c/\gamma \tag{21}$$

where γ is a constant. By this expression and according to (15), the Gaussian envelope in Gabor wavelet is adaptive to different frequencies. The discrete form of the Gabor wavelet used in this article is

$$\Psi[n|a_k, b_k] = \exp\left(j\frac{\omega_c T_s(n - b_k)}{a_k}\right) \exp\left(-\frac{T_s^2(n - b_k)^2}{a_k^2\alpha_0^2}\right) \tag{22}$$

When applying the complex Gabor wavelet with a set of discrete scaling factors, we achieve the proposed PCQ-WT, and it can be written as

$$X_{\alpha\beta}(\omega_k | a_k, b_k = 0) = \sum_{n=0}^{W-1} x_{\alpha\beta}[n] \Psi^*\left[-\frac{T_s m}{a_k}\right]$$
$$= \sum_{n=0}^{W-1} x_{\alpha\beta}[n] \exp\left(-j\frac{\omega_c}{a_k}T_s n - \frac{T_s^2 n^2}{a_k^2\alpha_0^2}\right). \tag{23}$$

If one determines the frequencies of interest and designs the Gabor wavelet bank properly, a vector $\mathbf{X}_{\alpha\beta,\underline{\omega}}$ that consists of a set of PCQ-WTs can be obtained; it is able to conduct time–frequency analysis and generate scalograms, thereby extracting features from the waveform in frequencies of interest.

B. Event Classification by CNNs

Pursuing development of an event detection mechanism in power systems, one needs to understand that the scalograms of
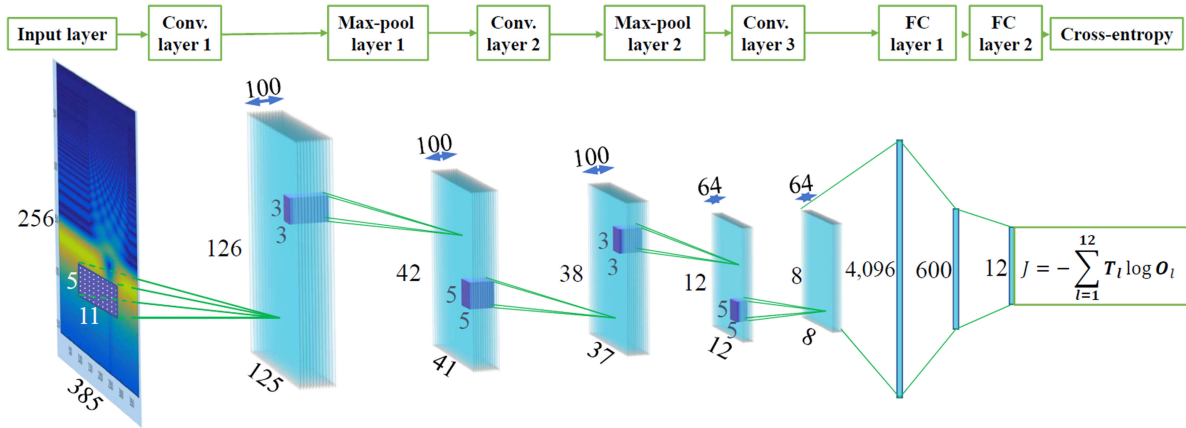
Fig. 2. Proposed architecture for CNN-based classification of the scalograms extracted from PCQ-WT. J is the cross-entropy loss, T_l is the true probability distribution over labels, and O_l is the predicted probability distribution by the networks [51].
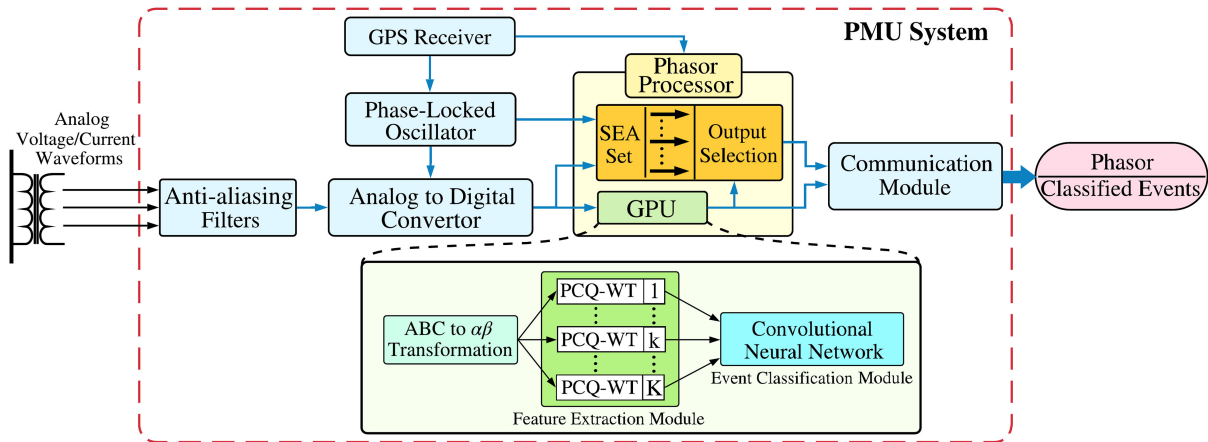

Fig. 3. Diagram of the proposed analytical framework embedded within PMU.

the waveforms generated by PCQ-WT convey valuable information on the events; the process of event detection is, therefore, converted to a supervised classification problem on the scalograms. However, the classification for the 2-D scalograms is challenging due to their high dimensionality. Specifically, every frame of the obtained scalogram has $scales \times time\,bins$ dimension (usually hundreds by hundreds); such high-dimensional data are prohibitive for most of the conventional pattern classification approaches. We treat the PCQ-WT scalograms as 2-D images and propose a CNN-based architecture to classify the events concealed in the scalograms (images). As the scalogram classification is not with very high abstraction level, we did not transfer any very deep neural networks to the task; instead, the proposed CNN has a simple architecture that meets the requirements of a synchrophasor, yet with very fast test times.

Our proposed CNN contains five layers—three convolutional (Conv.) layers and two fully connected (FC) layers. The architecture of the CNN can be seen in Fig. 2, and its specifications will be introduced in Section IV-C. The overall proposed online surveillance framework embedded within PMUs is

demonstrated in Fig. 3. This framework can work as an standalone event detection and classification tool within PMU, or it can assist the phasor processor to select a proper SEA, if a suite of multiple SEAs were equipped and available within the PMUs.

IV. CASE STUDY AND EXPERIMENTS

A. PCQ-WT Parameter Settings

The sampling frequency we used in this article is $Fs = 9600$ Hz. Up to the 50th order (3000 Hz) of harmonic is considered; therefore, mathematically, a frequency spectrum that ranges from 1 to 3000 Hz is requisite. We chose the mother wavelet's central frequency as 0.32 times of F_s, i.e., $F_c = 0.32$ (normalized) for PCQ-WT. The scaling factor should be in the range of [1, 3072]. However, calculating 3072 PCQ-WTs at the same time could lead to a high computational burden and large memory demands; hence, a down sampling of the scaling factor is much preferred. As the power waveform has its main energy concentration at around 60 Hz, this frequency has gained the most attention during monitoring. In designing the PCQ-WT,

Test Name	Input Range	Test Name	Input Range
Signal to Noise Ratio (SNR)*	40dB	Frequency Jump	-5Hz to 5Hz
Magnitude Jump	0.1-2pu	Phase Jump	$\pm\pi/18$ radians
Harmonic Distortion	0.5%-10%THD; order up to 50th	Out-of-Band Interference	10Hz to 120Hz; level 0.01-0.1pu
Amplitude Modulation	0.1Hz to 5Hz; level 0.005-0.1pu	Angle Modulation	0.1Hz to 5Hz; level 0.005-0.1pu
Frequency Ramp	±0.01Hz/s to ±1Hz/s, within ±5Hz	Single-line-to-ground (SLG) fault	Magnitude drop 0.2-1pu
Line-to-line (LL) fault	Magnitude drop 0.1-1pu**	Line-to-line-to-ground (LLG) fault	Magnitude drop 0.1-1pu

*Occurs in all test signals
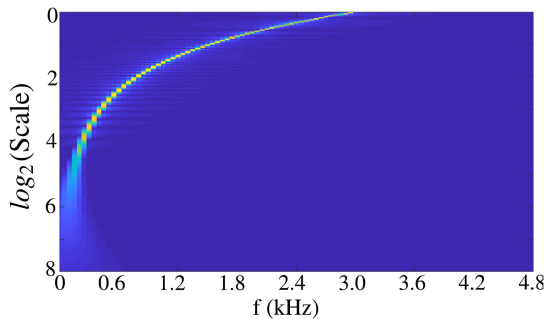
**Phase shift occurs at lines with faults

Fig. 4. Spectrum of the proposed wavelet bank, the scaling factor is plotted by $\log_2$.

we chose to compress the high-frequency range and neglect the low-frequency portion; a dyadic scaling factor range $[1, 256]$ is chosen, i.e., scaling factor $a = 2^i$, where i is sampled among 256 uniform intervals in the range of $[0, 8]$. Finally, we set the value of the constant $\gamma = 2$ and chose a fixed window size with 20 ms (192 time bins) associated with a fixed time shift $b_k = 100$. By setting these parameters, the PCQ-WT used in our experiments is obtained according to (22). The corresponding spectrum of the Gabor wavelet bank is shown in Fig. 4. The final scalogram fed into the CNN has a duration of 40 ms (385 time bins) including one historical window.

B. Test Waveform Specifications

The test power waveform modeling and parameter specification are selected according to [6] and [16], which describe the standard PMU testing and operating environments. With the 11 types of test waveforms as well as a normal waveform, a total 12 types of waveforms are simulated. The parameters for each type of simulated waveform are specified in Table I. We have made the following assumptions in the simulations.

1) The signal parameters are randomly selected in the designated ranges to simulate the uncertainty of event intensity. The time of occurrence for each event is also randomly located within a 40-ms run-time window which indicates that the event could happen at any time.
2) The measurement noise is assumed to be the white Gaussian and the signal-to-noise ratio (SNR) for all test signals is 40 dB. These assumptions are commonly observed in a PMU operational environment.

3) This article focuses on the transient event detection by using the original power waveforms. Only one transient event is assumed to occur at a given time instant, and the cascaded events caused by one peculiarity are not considered at this stage of the presented research.
4) We assume that the PMU is always functioning in its normal condition despite the faults in the grid; there is no bad data or congestion in the processor; the inner clock of PMU is synchronized with GPS, and the sampling rate is thus a constant.

C. Proposed CNN Configuration

As can be seen in Fig. 2, the CNN developed for the scalogram classification process has the following architecture: Input(256×385)–Conv($100, 5 \times 11$)–Max-pool(3×3)–Conv($100, 5 \times 5$)–Max-pool(3×3)–Conv($64, 5 \times 5$)–FC(600)–FC(12). Unlike the ordinary images which have homogeneous units on the two axes, the axes of scalograms are with different units. We chose a wide-shaped kernel in the first convolutional layer that could extract more information for transitions of the scalogram along the time axis. The stride of the convolution operation in the first layer is $(2, 3)$; other convolutional layers' strides are $(1, 1)$. Besides the last FC layer, batch normalization [56] is used in each Conv and FC layers. Dropout [57] was adopted in the third convolutional layer and the first FC layer to prevent overfitting. Rectified linear unit (ReLU) was chosen as nonlinearities in the neural network. The CNN used cross entropy as the loss function.

We experimented 120 000 samples of the wavelet scalograms corresponding to the 12 types of events in power grids; these events are simulated in the MATLAB environment as follows: 96 000 samples are used for training, 12 000 samples for validation, and 12 000 for testing. We used Adam [58] as the optimizer, which has the initialized learning rate of 1×10^{-3}, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The CNN was trained 120 epochs, and the learning rate was decayed 1/10 in every 30 epochs. The best validated model was recorded and tested.

D. Experimental Results and Analysis

Three patterns extracted by PCQ-WT from three selected events are demonstrated in Fig. 5. One can see from Fig. 5(a) that the magnitude of the highest energy concentration which stands for the fundamental frequency in the scalogram remains almost
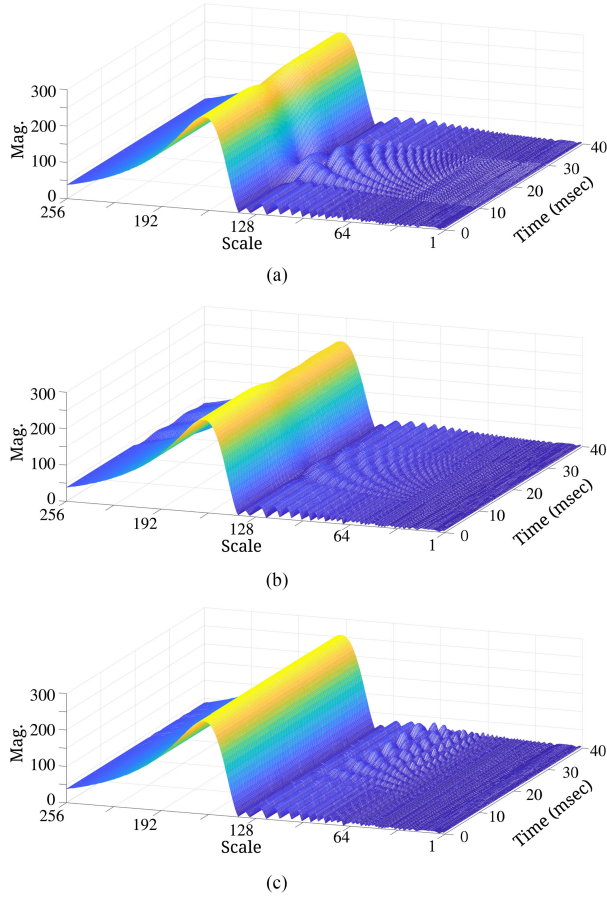
Fig. 5.   Test waveform simulation results: (a) 2 Hz frequency jump occurs at $t = 10$ ms; (b) high resistance line-to-line fault happens at $t = 10$ ms; (c) harmonic waves with orders of 5, 7, 9, and 11 are injected at $t = 10$ ms.

constant except at $t = 15$ ms (which is 5 ms after a frequency jump occurs). Among the high-frequency range—scaling factor from 1 to 128—the pattern appears almost immediately as the event happens. Fig. 5(b) shows a pattern caused by a high-resistance line-to-line fault. This pattern has different features compared to that in Fig. 5(a), and the highest energy concentration part has a significant drop after the fault happens. Similarly, this drop is around 5 ms after the fault; the features in the high-frequency range also show up immediately. When harmonics exist, one can see that the features in the high-frequency range are still unique in Fig. 5(c) and the fundamental frequency feature remains unchanged. Similar to the events in Fig. 5(a) and (b), the harmonic injection can be detected almost immediately.

As there is no such event detection mechanism that can simultaneously handle many types of events mentioned in Section IV-B, we here compare the proposed framework with traditional time–frequency analytics—WT and STFT, and a CNN module to form the test benchmarks.

1) The traditional WT method is implemented onto our proposed PCQ-WT algorithm. The scale factor is chosen as 2^i, where i is chosen from integers, i.e., $i = 0, 1, 2, .., 8$. This is aiming to ensure the same feature-extraction performance and the same frequency-coverage interval as

TABLE II
TRAINING TIME (HH:MM:SS) AND TESTING ACCURACY FOR DIFFERENT FEATURE-EXTRACTION METHODS

Scheme	Training Time	Best Accuracy
PCQ-WT*	01:37:29	92.84%
QWT	00:28:32	83.87%
STFT	01:34:47	80.97%

*Proposed

shown in Fig. 4. Here, we name this approach "QWT." As the QWT scalograms have the size of 9×385, the corresponding CNN architecture is adjusted to Input(9 $\times$ 385)–Conv(100, 3×11)–Conv(100, 3×3)–Conv(64, 3×3)–FC(600)–FC(12). This architecture aims to adapt to the new input size but with minimum change from the CNN specified in Section IV-C. Similar to our proposed CNN architecture for PCQ-WT, batch normalization was used between the two adjacent layers, two drop-out layers were applied in the third convolutional layer and first FC layer. ReLU was chosen as the activation function.

2) The STFT benchmark uses the same frequency coverage as Fig. 4 and window duration to generate spectrograms of size (256×385). These spectrograms are then fed into the CNN with configuration as in Fig. 2 for training.

We used a workstation which has an Intel Core i7-8700 K CPU and Nvidia GeForce GTX 2080Ti GPU as the computational platform; MATLAB 2016b and Pytorch 0.4.1 [59] as the implementation tools for PCQ-WT and CNN, respectively. 10 800 samples (9600 for training and 1200 for testing) were adopted for recording and comparing the computational time of each scheme. The training time and the best test result of these three approaches are compared as summarized in Table II. The best test classification accuracy of the proposed framework is found to be 92.84%, while the best accuracy for the QWT approach is 83.87% and for the STFT benchmark is 80.97%. The gap of the training time between PCQ-WT and STFT is less than three minutes, because the scalogram of PCQ-WT and the STFT spectrogram have the same size. QWT takes approximately one-third the training time of PCQ-WT; however, considering that the size of PCQ-WT scalogram is 28 times that in QWT, and that the classification accuracy of our proposed approach is almost 10% higher, the training time cost for the proposed approach is acceptable.

Additional details are shown in the confusion matrices (see Fig. 6). Comparing the confusion matrices between PCQ-WT [see Fig. 6(a)] and QWT [see Fig. 6(b)], one can see that the proposed PCQ-WT has only a slightly lower accuracy in detecting harmonic distortion and amplitude modulation events, but PCQ-WT shows higher and improved performance in detecting the rest types of events. Especially, PCQ-WT has improved SLG fault and LL fault detection performance than the traditional WT method. In detecting the normal operating condition in power grids, the PCQ-WT reveals a better accuracy than QWT, since the PCW-WT extracts more patterns from the waveforms and provides detailed information in the selected frequency intervals. Additionally, the CNN module would utilize those additional

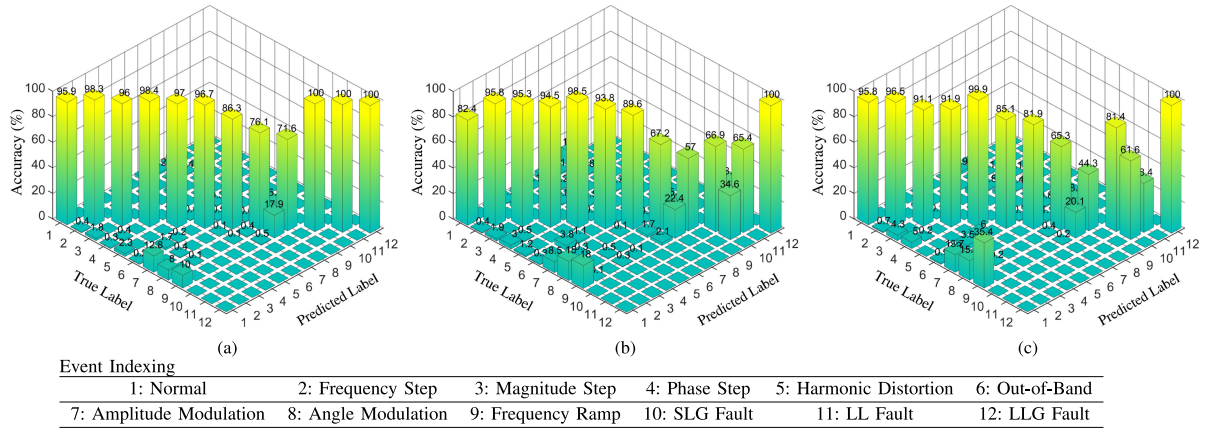| Event Indexing					
1: Normal	2: Frequency Step	3: Magnitude Step	4: Phase Step	5: Harmonic Distortion	6: Out-of-Band
7: Amplitude Modulation	8: Angle Modulation	9: Frequency Ramp	10: SLG Fault	11: LL Fault	12: LLG Fault

Fig. 6. Confusion matrices for the best validation results. (a) PCQ-WT. (b) QWT. (c) STFT.

features in the classification process. When comparing the PQC-WT with the STFT [see Fig. 6(c)], PCQ-WT also shows an improved performance, except for detecting harmonics. Nevertheless, one can see that the STFT approach has the highest accuracy in detecting harmonics among the three approaches, because the nature of STFT is to provide the harmonic information.

From Fig. 6(a), it can be seen that for most test waveforms, the proposed framework detects the event accurately. However, the frequency jump test case has the lowest accuracy; the reason is that the range of frequency jump in simulations is assumed to be within $[-5, 5]$ Hz. Thus, the values of frequency jump could be in the vicinity of zero; in such circumstances, the features in the scalogram are too weak to be detected by the CNN, and are hence, featured as "normal" signatures. In Fig. 6(a), the classification accuracy of the amplitude and angle modulation events are not ideal but still acceptable. The reasons for such an observation are twofold: 1) the physical duration in a scalogram proved to CNN is 40 ms, which is too short to capture the entire transition patterns and 2) the lower the modulation frequencies are, the smaller the modulation magnitudes appear. The above explanations are mathematically explained in the following:

$$x_{\alpha\beta}(t) = (1 + \underbrace{M_\Delta(t - t_\Delta)\sin(\omega_\Delta(t - t_\Delta)t))}_{\text{Modulation}}$$

$$\cdot \cos\left(2\pi \int_0^t F(\tau)d\tau + \phi(t)\right) \quad (24)$$

$$M_\Delta(t) = \begin{cases} C_M & t > 0, \quad C_M \in [0.005, 0.1] \\ 0 & t < 0 \end{cases}$$

$$\omega_\Delta(t) = \begin{cases} C_A & t > 0, \quad C_A/2\pi \in [0.1, 5] \\ 0 & t < 0 \end{cases} \quad (25)$$

where $M_\Delta(t)$ and $\omega_\Delta(t)$ are the modulation magnitude and frequency, respectively; and t_Δ is the event occurrence point. As the duration is 40 ms, i.e., $t_{\text{end}} = 40$ ms, and t_Δ should satisfy $0 < t_\Delta < 40$ so that the event can be shown in the scalogram, thus, $t_{\text{end}} - t_\Delta$ turns into a very small value. With small C_M

and C_A, the modulation can be written as

$$\text{Mod}(t) = C_M \sin(2\pi C_A(t - t_\Delta))$$
$$\approx 2\pi C_M C_A(t - t_\Delta)$$
$$\approx 0, t \in [t_\Delta, t_{\text{end}}]. \quad (26)$$

From (26), one can see that if the event occurs at a time that is very close to t_{end}, the modulation $\text{Mod}(t)$ is extremely small under the condition that C_M and C_A are small; meanwhile, the duration of modulation in the scalogram is very short. Thus, a scalogram with 40 ms duration would not contain sufficient features for the occurrence of amplitude modulation, and the classification accuracy for such events are relatively lower than the others. Similar arguments hold for analyzing the results under angle-modulation scenarios.

Therefore, increasing the window size and the modulation strength may be beneficial to the classification accuracy; this, however, sacrifices the overall performance, particularly, for real-time measurements. As the proposed scheme aims to be deployed as a fast online grid monitoring tool and an awareness mechanism to assist the PMUs selecting an appropriate SEA, we kept the parameters that are specified in Section IV-A.

We also investigate the optimal selection of the parameter values in the CNNs. As there are thousands of values within the CNNs in this article, we here examined the PCQ-WT-connected CNN as a demonstration. Fig. 7 displays the optimal values in the first convolutional layer learned by the CNN, i.e., the 100 kernels of Conv. layer 1 in Fig. 2. These kernels indicate the extracted features for the scalograms, but since the scalograms are not natural images, the learned kernels do not present strong patterns for human vision.

As a time-critical infrastructure, a PMU (in a 60-Hz system) usually has to process the sampled signal and report the phasor measurements at the rate of 60 frames/s [6]. The computational time of the PCQ-WT and CNN should be, hence, investigated. Generating a 256×385 scalogram by PCQ-WT takes 1.20 ± 0.23 ms, and the CNN takes 1.04 ± 0.31 ms to process every scalogram. Therefore, the in-total computational time of our proposed framework is 2.24 ± 0.39 ms. In this timing experiment, the timer starts from when the data were loaded into
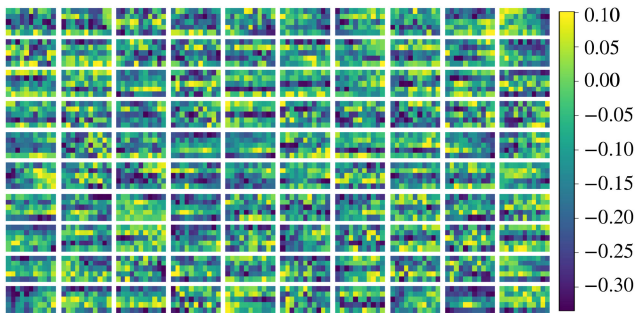
Fig. 7. One hundred learned kernels in the first convolutional layer of the proposed PCQ-WT scheme.

the CPU and GPU memory and ends when the classification results were given out. The test uses 32-bit precision. This short computational time satisfies the real-time constraints with large margin and leaves plenty of time for the subsequent SEA analytics to function.

V. CONCLUSION AND DISCUSSION

This article introduced a novel PMU-embedded analytic for power grid online surveillance that consists of the PCQ-WT feature extraction and CNN-based event classification mechanisms. The proposed framework aims to effectively extract the waveform features and efficiently classify multiple types of events in the grid. Our experiments and comparisons demonstrated that the proposed framework achieved improved accuracy for real-time event detection and classification. This framework would be a foundation for next-generation intelligent PMUs that could adaptively select an appropriate SEA and achieve higher phasor measurement accuracy. Future work would focus on the integration of SEAs in the proposed framework to leverage overall functionality of PMUs. Research on implementing other types of neural networks and machine learning techniques, such as [24] and [25], should be pursued to further improve the event detection and classification accuracy.

REFERENCES

[1] J. A. De la O Serna, "Synchrophasor measurement with polynomial phase-locked-loop Taylor–Fourier filters," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 328–337, Feb. 2015.

[2] M. H. Rezaeian Koochi, P. Dehghanian, S. Esmaeili, P. Dehghanian, and S. Wang, "A synchrophasor-based decision tree approach for identification of most coherent generating units," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2018, pp. 71–76.

[3] T. Becejac, P. Dehghanian, and M. Kezunovic, "Analysis of PMU algorithm errors during fault transients and out-of-step disturbances," in *Proc. IEEE PES Transmiss. Distrib. Conf. Expo.-Latin Am.*, 2016, pp. 1–6.

[4] N. H. Abbasy and H. M. Ismail, "A unified approach for the optimal PMU location for power system state estimation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 806–813, May 2009.

[5] S. Chakrabarti, E. Kyriakides, and D. G. Eliades, "Placement of synchronized measurements for power system observability," *IEEE Trans. Power Del.*, vol. 24, no. 1, pp. 12–19, Jan. 2009.

[6] *IEEE Standard for Synchrophasor Measurements for Power Systems*, *IEEE Std. C37.118.1-2011 (Revision of IEEE Std C37.118-2005)*, Dec. 2011.

[7] A. Z. Amanci and F. P. Dawson, "Synchronization system with zero-crossing peak detection algorithm for power system applications," in *Proc. IEEE Int. Power Electron. Conf.*, 2010, pp. 2984–2991.

[8] S. Das and T. Sidhu, "Robust algorithm to estimate fault synchrophasor from fault-transient synchrophasor in phasor data concentrator," *IET Gener., Transmiss. Distrib.*, vol. 9, no. 2, pp. 124–132, 2015.

[9] C. Thilakarathne, L. Meegahapola, and N. Fernando, "Static performance comparison of prominent synchrophasor algorithms," in *Proc. IEEE Innov. Smart Grid Technol. - Asia*, Dec. 2017, pp. 1–6.

[10] Á. Ortega and F. Milano, "Comparison of different PLL implementations for frequency estimation and control," in *Proc. IEEE 18th Int. Conf. Harmon. Qual. Power*, 2018, pp. 1–6.

[11] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1474–1483, Aug. 2011.

[12] P. K. Dash, K. R. Krishnanand, and M. Padhee, "Fast recursive Gauss–Newton adaptive filter for the estimation of power system frequency and harmonics in a noisy environment," *IET Gener., Transmiss. Distrib.*, vol. 5, no. 12, pp. 1277–1289, Dec. 2011.

[13] A. J. Roscoe, I. F. Abdulhadi, and G. M. Burt, "P and M class phasor measurement unit algorithms using adaptive cascaded filters," *IEEE Trans. Power Del.*, vol. 28, no. 3, pp. 1447–1459, Jul. 2013.

[14] I. Kamwa, S. R. Samantaray, and G. Joos, "Wide frequency range adaptive phasor and frequency PMU algorithms," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 569–579, Mar. 2014.

[15] P. Castello, J. Liu, C. Muscas, P. A. Pegoraro, F. Ponci, and A. Monti, "A fast and accurate PMU algorithm for P+M class measurement of synchrophasor and frequency," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 12, pp. 2837–2845, Dec. 2014.

[16] T. Becejac, P. Dehghanian, and M. Kezunovic, "Probabilistic assessment of PMU integrity for planning of periodic maintenance and testing," in *Proc. IEEE Int. Conf. Probabilistic Methods Appl. Power Syst.*, 2016, pp. 1–6.

[17] T. Becejac, P. Dehghanian, and M. Kezunovic, "Impact of the errors in the PMU response on synchrophasor-based fault location algorithms," in *Proc. North Am. Power Symp.*, 2016, pp. 1–6.

[18] C. Qian and M. Kezunovic, "A power waveform classification method for adaptive synchrophasor estimation," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1646–1658, Jul. 2018.

[19] C. Hou, S. Lin, S. Su, and W. Chung, "Fault tolerant quickest detection for power quality events in smart grid AMI networks," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2015, pp. 159–163.

[20] Y. Song, W. Wang, Z. Zhang, H. Qi, and Y. Liu, "Multiple event detection and recognition for large-scale power systems through cluster-based sparse coding," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4199–4210, Nov. 2017.

[21] W. Wang *et al.*, "Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1654–1664, Jul. 2014.

[22] R. Yadav, A. K. Pradhan, and I. Kamwa, "Real-time multiple event detection and classification in power system using signal energy transformations," *IEEE Trans. Ind. Inform.*, vol. 15, no. 3, pp. 1521–1531, Mar. 2019.

[23] M. Rafferty, X. Liu, D. M. 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.

[24] C. Chen *et al.*, "Resilient adaptive and H_∞ controls of multi-agent systems under sensor and actuator faults," *Automatica*, vol. 102, pp. 19–26, 2019.

[25] C. Chen, H. Modares, K. Xie, F. L. Lewis, Y. Wan, and S. Xie, "Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4423–4438, Nov. 2019.

[26] H. R. Berenji, "A reinforcement learning-based architecture for fuzzy logic control," *Int. J. Approx. Reason.*, vol. 6, no. 2, pp. 267–292, 1992.

[27] C. Chen, Z. Liu, K. Xie, Y. Zhang, and C. P. Chen, "Asymptotic adaptive control of nonlinear systems with elimination of overparametrization in a Nussbaum-like design," *Automatica*, vol. 98, pp. 277–284, 2018.

[28] C. Chen, Z. Liu, K. Xie, Y. Zhang, and C. P. Chen, "Adaptive neural control of MIMO stochastic systems with unknown high-frequency gains," *Inf. Sci.*, vol. 418, pp. 513–530, 2017.

[29] K. Xie, C. Chen, F. L. Lewis, and S. Xie, "Adaptive asymptotic neural network control of nonlinear systems with unknown actuator quantization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6303–6312, Dec. 2018.

[30] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1653–1656, Mar. 2019.

[31] C. Wei, Z. Zhang, W. Qiao, and L. Qu, "An adaptive network-based reinforcement learning method for MPPT control of PMSG wind energy conversion systems," *IEEE Trans. Power Electron.*, vol. 31, no. 11, pp. 7837–7848, Nov. 2016.

[32] J. Duan, Z. Yi, D. Shi, C. Lin, X. Lu, and Z. Wang, "Reinforcement-learning-based optimal control of hybrid energy storage systems in hybrid ac–dc microgrids," *IEEE Trans. Ind. Inform.*, vol. 15, no. 9, pp. 5355–5364, Sep. 2019.

[33] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2158–2169, Mar. 2019.

[34] W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2192–2203, Jun. 2018.

[35] F. A. S. Borges, R. A. S. Fernandes, I. N. Silva, and C. B. S. Silva, "Feature extraction and power quality disturbances classification using smart meters signals," *IEEE Trans. Ind. Inform.*, vol. 12, no. 2, pp. 824–833, Apr. 2016.

[36] M. S. Manikandan, S. R. Samantaray, and I. Kamwa, "Detection and classification of power quality disturbances using sparse signal decomposition on hybrid dictionaries," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 27–38, Jan. 2015.

[37] S. Wang, P. Dehghanian, L. Li, and B. Wang, "A machine learning approach to detection of geomagnetically induced currents in power grids," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, Sep. 2019, pp. 1–7.

[38] S. Wang, P. Dehghanian, and Y. Gu, "A novel multi-resolution wavelet transform for online power grid waveform classification," in *Proc. Int. Conf. Smart Grid Synchronized Meas. Analytics*, 2019, pp. 1–8.

[39] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[40] S. Wang, L. Li, and P. Dehghanian, "Power grid online surveillance through PMU-embedded convolutional neural networks," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, Sep. 2019, pp. 1–8.

[41] R. H. Park, "Two-reaction theory of synchronous machines generalized method of analysis—Part I," *Trans. Am. Inst. Elect. Eng.*, vol. 48, no. 3, pp. 716–727, Jul. 1929.

[42] W. Duesterhoeft, M. W. Schulz, and E. Clarke, "Determination of instantaneous currents and voltages by means of alpha, beta, and zero components," *Trans. Am. Inst. Elect. Eng.*, vol. 2, no. 70, pp. 1248–1255, 1951.

[43] J. A. De la O Serna and J. Rodríguez-Maldonado, "Taylor–Kalman-Fourier filters for instantaneous oscillating phasor and harmonic estimates," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 4, pp. 941–951, Apr. 2012.

[44] M. Bertocco, G. Frigo, C. Narduzzi, C. Muscas, and P. A. Pegoraro, "Compressive sensing of a Taylor–Fourier multifrequency model for synchrophasor estimation," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 12, pp. 3274–3283, Dec. 2015.

[45] S. Wang, P. Dehghanian, and B. Zhang, "A data-driven algorithm for online power grid topology change identification with PMUs," in *Proc. IEEE Power Energy Soc. General Meeting*, Aug. 2019, pp. 1–5.

[46] D. P. Mishra, S. R. Samantaray, and G. Joos, "A combined wavelet and data-mining based intelligent protection scheme for microgrid," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2295–2304, Sep. 2016.

[47] K. Thirumala, M. S. Prasad, T. Jain, and A. C. Umarikar, "Tunable-Q wavelet transform and dual multiclass SVM for online automatic detection of power quality disturbances," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3018–3028, Jul. 2018.

[48] Y.-C. Su, K.-L. Lian, and H.-H. Chang, "Feature selection of non-intrusive load monitoring system using STFT and wavelet transform," in *Proc. IEEE 8th Int. Conf. e-Business Eng.*, 2011, pp. 293–298.

[49] S.-H. Ni, K.-F. Lo, L. Lehmann, and Y.-H. Huang, "Time–frequency analyses of pile-integrity testing using wavelet transform," *Comput. Geotechnics*, vol. 35, no. 4, pp. 600–607, 2008.

[50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[51] J. S. Bridle, "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," in *Proc. NATO Conf. Neurocomput.*, 1990, pp. 227–236.

[52] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 454–459.

[53] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[54] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.

[55] J. Hubbard, "Calculation of partition functions," *Phys. Rev. Lett.*, vol. 3, no. 2, 1959, Art. no. 77.

[56] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*.

[57] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*.

[59] A. Paszke *et al.*, "Automatic differentiation in pytorch," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 1–4.

Shiyuan Wang (S'18) received the B.Eng. degree in mechanical engineering from the University of Science and Technology Beijing, Beijing, China, in 2012, and the M.Sc. degree in electrical engineering from The George Washington University, Washington, DC, USA, in 2014, where he is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering.

His research interests include power system reliability and resiliency, smart grid and renewable energy, power grid harmonic analysis, and application of signal processing in energy analytics.

Payman Dehghanian (S'11–M'17) received the B.Sc. degree from the University of Tehran, Tehran, Iran, in 2009, the M.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 2011, and the Ph.D. degree from Texas A&M University, Texas, USA, in 2017, all in electrical engineering.

He is an Assistant Professor with the Department of Electrical and Computer Engineering, The George Washington University, Washington, DC, USA. His research interests include power system protection and control, power system reliability and resiliency, asset management, and smart electricity grid applications.

Dr. Dehghanian is the recipient of the 2013 IEEE Iran Section Best M.Sc. Thesis Award in Electrical Engineering, the 2014 and 2015 IEEE Region 5 Outstanding Professional Achievement Awards, and the 2015 IEEE-HKN Outstanding Young Professional Award.

Li Li (S'18) received the B.Sc. degree in control engineering from Jilin University, Changchun, China, in 2008, and the M.Sc. degree in electrical engineering from The George Washington University, Washington, DC, USA, in 2014, where he is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering.

His research interests include image processing, artificial intelligence, and machine learning.