# Performance analysis of wavelet scattering transform-based feature matrix for power system disturbances classification

**Naema M. Mansour[1], Ibrahim A. Awaad[2], Abdelazeem A. Abdelsalam[1]**
[1]Department Electrical Engineering, Faculty of Engineering, Suez Canal University, Ismailia, Egypt
[2]Department of Electrical, Faculty of Engineering, Sinai University-Arish Branch, Arish, Egypt

**Article Info**

***Article history:***

Received Mar 12, 2024
Revised Jul 30, 2024
Accepted Aug 6, 2024

***Keywords:***

Invariance scales
Long short-term memory
Number of wavelets per octave
Power system disturbance
Sampling frequency
Signal length
Wavelet scattering transform

**ABSTRACT**

Recently, the wavelet scattering transform (WST) was introduced as a powerful feature extraction tool for classification processes. It provides good performance in applications involving audio signals, images, medical data, and quadcopters for structural health diagnosis. It is also employed in several electrical engineering applications, such as the classification of induction motor bearing failures, electrical loads, and industrial robot faults. Despite its development, the performance of the wavelet scattering (WS) network constructed in the MATLAB environment to compute WST coefficients has not been highlighted in the literature so far. In this paper, the properties of the WST feature matrix are examined, and the parameters that have a significant impact on coefficient magnitudes and matrix dimensions are defined. With minimal configuration, a WS network could extract low-variance features from real-valued time series for use in machine learning and deep learning applications. The feature matrix, which contains zero, first, and second-level WST coefficients derived from various power system signal configurations, is constructed to be trained using long short-term memory (LSTM) networks. The simulation results demonstrate the efficacy of the proposed classifier with an accuracy approach of 100%. The MATLAB toolbox has been used to create different signals for the WS and LSTM networks. WST has proven to be a powerful tool for power system disturbance classification.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*

***Corresponding Author:***

Naema M. Mansour
Department Electrical Engineering, Faculty of Engineering, Suez Canal University
4.5 Km Ring Road, Ismailia, Egypt
E-mail: naima.mansour@eng.suez.edu.eg

## 1. INTRODUCTION

Wavelet scattering transform (WST) which is used to compute a locally translation-invariant representation of the real-valued signals that are stable to time-warping deformations and suitable for many signal processing and machine learning applications, was developed by Bruna and Malat. The translation-invariant representation is computed by cascading wavelet convolutions, modulus operators, and low-pass filters, which average the amplitude of iterated wavelet coefficients. The strength of traditional signal-processing tools and the depth of a deep neural network are combined in WST. It provides feature vectors that are robust to noise, time-shift invariant, and stable against time-warping deformations, which have recently provided a powerful result in classification tasks [1]–[5].

As a powerful feature extraction tool, WST has recently been introduced as a part of powerful classification algorithms for image and audio classification, which has led to its widespread use in various applications. In [6], the WST coefficients of the audio signals are used to provide an accurate classification of voiced and unvoiced sounds, and WST's translation-invariant image representation could be classified as introduced in [7]. Also, in the audio and image processing applications, a speaker identification system based on WST has been introduced in [8], indoor fingerprinting localization-based WST has been introduced in [9], and in [10], a WST-based machine learning for ground penetrating radar imaging has been introduced. In the medical field, WST has been used as a powerful feature extraction tool for X-ray COVID-19 detection [11], classification of interictal and preictal EEG signals [12], wearable electrocardiogram quality assessment [13], glaucoma detection [14]. Also, WST provides good performance in chemistry applications; the classification of organic molecules has been introduced by Hirn *et al.* [15]. In the electrical engineering field, the growth of WST applications is still very slow. By using WST, Rohan [16] has introduced a method by which the mechanical components of industrial robots could be diagnosed. In study [17], the different mixed faults of quadcopter structures could be accurately classified. Reaching up to 99.98% accuracy, a classification framework for nonintrusive appliance load monitoring-based WST feature extraction is proposed in [18]. Greater fault diagnosis accuracy than other methods in the literature; AlBader and Toliyat [19] proposed a fault diagnosis technique for analog circuits and rotating machinery bearings and gear faults based on WST features. Based on WST feature extraction, an algorithm for rotating machines bearing fault detection is introduced in [20], [21]. In all these applications, a feature matrix is implemented using one or multi-levels of WST according to the nature of the analyzed signal components.

In power system protection, feature extraction from current signals in various normal and up-normal conditions is considered a vital issue, with generating a discriminative feature matrix being the desired task to achieve accurate fault diagnosis, particularly in low fault current cases such as high impedance faults and isolated micro-grid faults. As a pre-processing tool required for artificial intelligence (AI) based classification algorithms, WST has not yet been introduced for power system disturbance diagnosis in the literature. Electric current signals, which contain extensive information in both time and frequency, are the principal sources of information in power system applications. Separating information embedded in the power system current signal could be achieved by WST, which provides significant power for classification. There are multiple factors that determine the characteristics of the WST network implemented in the MATLAB toolbox [22], [23]. Each factor's appropriate value needs to be carefully selected based on the input signal's characteristics. For the first time in the literature, this paper examines the impacts of the input signal sampling frequency, its length, the invariance scale, and the Q-factor of the filter banks on the WST feature matrix dimensions and their coefficient magnitudes created in the MATLAB toolbox. Furthermore, a new classification algorithm has been proposed for power system disturbance diagnosis that utilizes the WST as a feature extraction tool and the long short-term memory network (LSTM) network as a binary classifier. Designed to capture historical information of time series data, a LSTM is a type of deep neural network suitable for predicting long-term nonlinear series. The inability to remember long-term dependencies due to gradient is the shortcoming that is faced by recurrent neural networks (RNN). LSTM networks are explicitly designed to avoid this problem, making them ideal for learning features, identifying, and classification. The LSTM network has recently been proposed in the literature for identifying power system disturbances, with promising results [24], [25]. Zhang *et al.* [24] developed a line trip fault prediction algorithm based on LSTM and support vector machine (SVM) with accuracy reaching about 97%. Karan and Yeh [26] introduce a fault classification model for microgrids that employs DWT as a preprocessing tool and LSTM and convolutional neural networks (CNN) for data training. The analysis revealed that the LSTM network outperformed the CNN classifier and achieved high accuracy in classifying the faults. Belagoune *et al.* [27] introduces three unique classification methods based on LSTM with high accuracy for fault region identification, fault type, and fault location prediction of the transmission line. The suggested LSTM classifier in [28] provides an overall classification accuracy of 91.21% for detecting high impedance faults in solar PV integrated power networks. Branco *et al.* [29] introduced a fault prediction model based on LSTM using DWT. The wavelet LSTM model showed better results in all analyses compared with the standard LSTM mode. Omar *et al.* [30] introduced a fault classification on the transmission line using the LSTM network as a tool to classify different types of faults. Simulation results show promising classification accuracy of 100%, 99.77%, and 99.55% for ideal 30 and 20 dB noise, respectively. Cortes-Robles *et al.* [31] proposed an LSTM-based technique for classifying events that affect power quality (PQ) in power networks with distributed generation sources, with an accuracy of 99.75%. A comprehensive fault identification model-based LSTM is introduced in [32]. This model can accurately identify the short circuit, surge, residual current, and other faults in the distribution network lines. With an accuracy of 99.98% compared to 42.98% for ANN. Also, the authors of [25] introduced deep learning techniques for transmission line fault classification based on LSTM.

Although CNN-based classification models with learned filters can provide adequate accuracy, they have some limitations, such as a lack of understanding of their architecture and the need for a large dataset for training, which leads to longer training times, making training computationally expensive. In contrast, Bruna and Mallat in 2013 [7] introduce wavelet scattering (WS) networks as CNNs with fixed filters and weights, that provide a generic and fixed initialization of the first layers of a deep network, while the remaining layers are learned supervised. So, the WS network is regarded as a deep learning tool that is fast, well-understood, computationally inexpensive, and works with a small dataset of training samples. As demonstrated by Mallat, these properties guide the optimization of the network architecture to preserves high-frequency information for classification while avoiding useless computations. In the MATLAB toolbox, the pre-assigned factors affect the architecture of the WS network; by adjusting these variables, the dominant features of the input time signal could be correctly extracted. In power system analysis, the signal's features could be accurately retrieved by adjusting the sampling frequency, wavelet filter coefficients, and the number of wavelets per octave. Furthermore, the matrix dimensions can be adjusted, potentially adjusting the computational burden. Wavelet's multi-resolution property makes it an appropriate tool for analyzing irregular transient changes in voltage or current signals in the power system network during upnormal conditions. For all of these reasons, the WS network is appropriate for power system disturbance classification. Using the LSTM model without a pre-filtering stage made it difficult to make accurate predictions. As a preprocessing tool, WST is used to extract the feature matrix in this study. The suggested classifier is investigated and validated using a variety of WS networks and sine-wave signals with varying lengths and sampling frequencies. In this work, the MATLAB software platform is used for tasks involving the creation, analysis, and plotting of signals and WS networks and creating a WS feature matrix and LSTM network.

The paper is organized as follows: section 2 outlines the WST network architecture. Section 3 investigates the WST network's specified parameters. Section 4 examines the WST's performance, and the new LSTM-based classification model for power system disturbances is described and evaluated in section 5. Finally, the research results are concluded in section 6.

## 2. ARCHITECTURE OF WST NETWORK

In 2012, Bruna and Malat [3] introduced the mathematical algorithm and structure of WST, which is similar to a CNN. WST decomposes the original data into a series of stages or layers of a tree structure, with the output from one layer being the input for the next. Each layer performs three basic transform operations: convolution, nonlinearity (modulus), and average pooling. As shown in Figure 1, WST analyzes the real-valued signal in three successive computation phases; the convolution step determines the signal's resemblance to wavelets of different frequencies and scales. The modulus non-linearity step extracts the amplitude modulation of the convoluted signal and allows for the extraction of higher-order modulations. Regarding the complicated signal:

$$f(t) = f_r(t) + j\, f_i(t) \tag{1}$$

The modulus operator is defined as (2):

$$|f(t)| = \sqrt{f_r(t)^2 + f_i(t)^2} \tag{2}$$

By squaring the real and imaginary parts of the signal and convolving with the invariance scale function, a modulated coefficient is generated. Finally, the averaging step, which can be considered a type of pooling, in which an implicit down sampling operation is performed to reduce the dimensions of the feature map. This enhances the model's resistance to shifts in the relative positions of the transients in the input signal and enhances its robustness against time-wrapping distortion. In contrast, the averaging process leads to the loss of accurate time localization. WST calculates higher order or layers coefficients by iterating over wavelet transforms and modulus operators. Bruna and Mallat [3] developed a computational implementation with a specific architecture of a WST similar to a deep convolutional network using a predetermined wavelet filter, as illustrated in Figure 2.

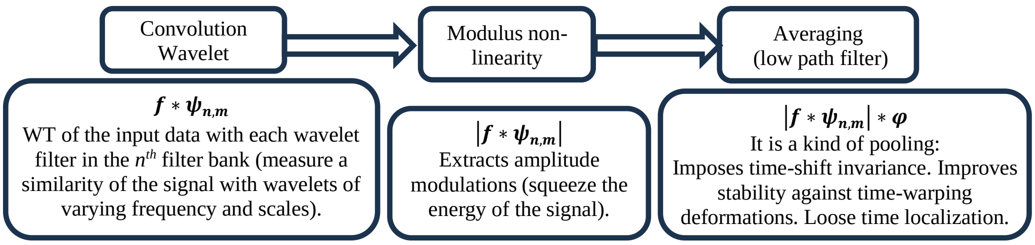


Figure 1. WST mathematical steps

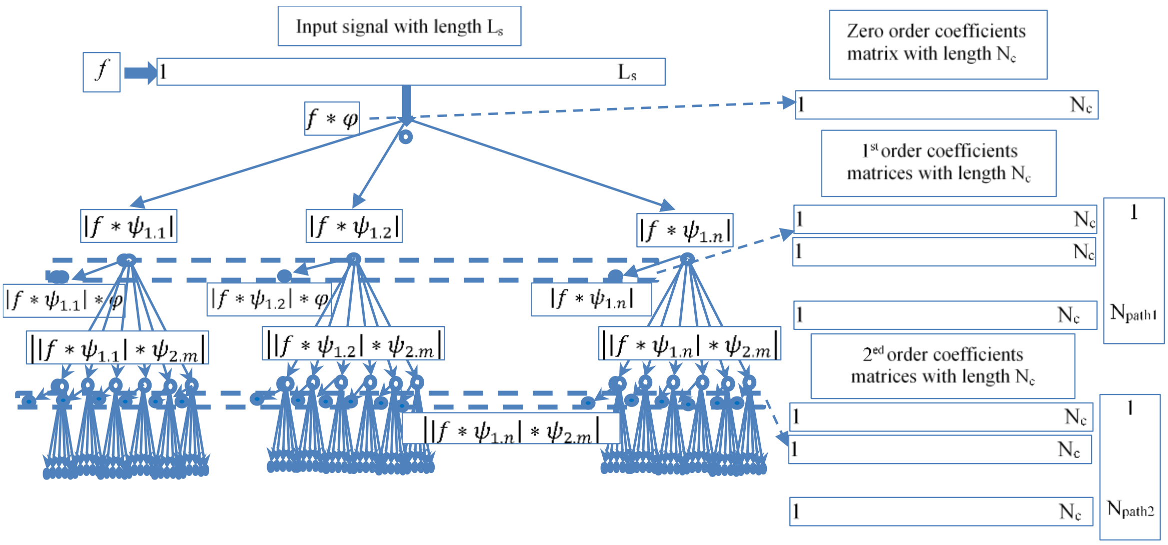


Figure 2. Wavelet time scattering network, $N_c$ number of coefficients in each path (depends on the sampling frequency of the input signal, input signal length, subsampling factor, and invariance scale), $N_{path1}$ the number of paths in the first order scattering level, $N_{path2}$ the number of paths in the second order scattering level ($N_{path1}$, $N_{path2}$ depends on the Q-factor of each level filter bank, and the sampling frequency of the input signal)

It provides coefficients that resemble the time-averaged values of the modulated magnitude of the input signal, providing informative signal invariants over potentially large time scales that are suitable for classifications. By this hierarchical implementation, the feature array of WST coefficients is computed as: The input signal $f(t)$ is convolved with the scale function $\varphi(t)$ to provide the zero-order scattering coefficient, i.e., the average of the input signal along the scale function according to (3).

$$S_o = f(t) * \varphi(t) \tag{3}$$

The input signal is convolved with the first filter bank of the continuous wavelet function $\psi_{1,n}$ to produce continuous wavelet transform coefficients (CWTC), then the modulus of the CWTC is computed and convolved with the scaling function to provide a set of first-order scattering coefficients as in (4):

$$S_1 = |f(t) * \psi_{1,n}| * \varphi(t) \tag{4}$$

where *n* is the number of filters in the first-order filter bank, in the MATLAB toolbox, its value is dependent on the values of the invariance scale, the number of filters per octave of the first-order WST, and the sampling frequency of the input signal. By repeating the same calculations, the second-order WST coefficients are computed; the modulus of the CWTC of the input signal with the first filter bank is convolved with the second CWT filter bank, then the modus is computed, and convolved with the scaling function as in (5):

$$S_2 = \left||f(t) * \psi_{1,n}| * \psi_{2,m}\right| * \varphi(t) \tag{5}$$

where *m* is the number of filters in the second-order filter bank, in the MATLAB toolbox, its value is dependent on the values of the invariance scale, the number of filters per octave of the second-order filter bank, and the sampling frequency of the input signal. The higher levels are computed to recover the information that has been lost in the previous levels. Andén and Mallat [2] proved that the first- and second-order coefficients carry more than 98% of the input signal energy, and the feature matrix that contains the feature captured in the first three levels is sufficient for classification issues [2].

## 3. WST NETWORK PREDEFINED PARAMETERS

As the coefficient order increases, Andén and Mallat [2] confirm that the energy of all scattering coefficients decays to zero, with the first and second-order coefficients carrying more than 98% of the input signal energy. Furthermore, their examination of audio signals demonstrated the ability to characterize non-stationary signals using the WST coefficient, as well as the ability to identify more sophisticated phenomena such as transient circumstances. Therefore, this study focuses on the zero, first, and second layer coefficients of WST.

In power systems, the dominant frequency is the fundamental frequency of normal operation, which is 50 or 60 Hz, as well as high-frequency components and some harmonics associated with the fundamental frequency in up-normal situations. Power system faults are typically characterized by variations in current magnitude, high frequencies due to the switching operations, harmonics, and DC exponential components. The unique characteristics of the power system signal make WST a suitable tool for capturing its key features. The zero-order coefficient contains the majority of the low frequency and DC components, whereas the dominant fundamental frequency can be described by the set of first-order coefficients, and the transient phenomena associated with faults and switching conditions can be described by the set of second-order scattering coefficients, which demonstrate wavelets with more narrow time support. Furthermore, the fault diagnosis can be achieved irrespective of its time occurrences due to the time-invariant effect of WST.

In the MATLAB programming environment, a WST scientific toolbox that provides tools for the analysis and classification of digital signals, such as sounds, images, and time series. The toolbox enables data-centric AI workflows by providing time-frequency transforms and automated feature extraction [22], [23]. To create a wavelet time scattering network with two filter banks, several parameters should be well defined in order to create a wavelet time scattering network capable of representing the features of the input signal. These parameters encompass:

- The wavelet function: The input signal's decomposition is achieved by using the wavelet function, by default, the Gabor (analytic Morlet) wavelet was utilized.
- The number of wavelet filters per octave (quality factor Q of the filter bank): The Q-factor value determines the center frequencies of the created wavelet filters; therefore, it should be specified so that all of the input signal's significant features can be effectively represented. The number of paths in each layer that can represent the characteristics of the input signal is determined by the Q-factor value. The architectural design of a WST network considers each path (the bold nodes in Figure 2) to be a row in the feature matrix, so the dimension of the feature matrix is strongly affected by this parameter. Contrary to the CNN, the filter weight of WST is constant and predefined. The first and second filter banks employ [8, 1] filters per octave by default.
- The duration of translation invariance (invariance scale IS): The invariance scale is measured in samples by default if the sampling rate is left unspecified, and in seconds if the sampling rate is specified. Its maximum value is limited by the length of the input signal. For instance, with 1,024 samples input signal and a sampling rate of 3,200 Hz, the highest possible value of the invariance scale is 0.32 seconds. In the MATLAB toolbox, if the IS value is not determined, the length of the input signal and the sampling frequency are needed to confirm the invariance scale value. If the value of the invariance scale is not predetermined, it is automatically calculated according to (6), where $L_s$ is the input signal's length per samples, and sampling frequency $f_s$.

$$IS = (L_s/f_s)/2 \tag{6}$$

  The larger the invariance scale, the larger the time support of the scale function, and accordingly, the more the down-sample since the average of the modulus of wavelet coefficients is calculated over a half-overlapping time window, the width of this time window is defined by scale invariance value. In essence, this factor is crucial in determining the dimensions of the feature matrix and plays a vital role in shaping its characteristics.
- The signal length (number of samples in the input signal): In the MATLAB toolbox, it is defined as an integer positive value that is either greater than or equal to 16. By default, a signal length of 1,024 samples are utilized. Its value determines the number of coefficients of each path, and hence the feature matrix dimensions.
- Sampling rate (sampling frequency of the input signal): It is chosen based on the frequency content of the input signal. Based on this value, the frequency span and hence the number of created wavelet filters are determined. The highest frequency passband is designed so that the amount of passband gain drops to half of its peak value at the Nyquist frequency. Thus, the number of wavelets in the filter bank, and the number of coefficients in each path are determined.

Careful selection of these parameters is critical since they have a significant impact on the WST coefficients' characteristics and the dimensions of the feature matrix. As a result, carefully choosing these variables is critical to obtaining relevant results.

## 4. POWER SYSTEM SIGNALS ANALYSIS USING WST

In this section, the performance of WST is investigated for different power system signal configurations. This experimental study is based on a pure 50 Hz sine wave as the dominant power system frequency signal with different sampling rates and lengths. Different versions of this signal are then created to simulate different operating conditions in the power system. Different WS networks are implemented according to different values of the factors discussed in section 3. The ability of WST to extract characteristics from power system signals has been extensively researched.

### 4.1. Sampling frequency impacts

Unless the frequency limits in the WST toolbox are pre-defined, the frequency span and hence the number of wavelet filters that specify the number of WST paths are calculated based on the value of the input signal's sampling frequency. A constant signal length and a scale-invariance factor are used to examine the impacts of sampling frequency fluctuation. The following parameters are defined to create a WS network using the MATLAB toolbox if the input signal is a pure sine wave with a fundamental frequency of 50 Hz: the sampling frequency of 3,200 Hz, scale invariance of 0.32 sec, default signal length of 1,024 samples, and default quality factor [8, 1]. For the first filter bank, 48 decomposition filters were developed, and for the second filter bank, 8 filters, as illustrated in Figure 3.

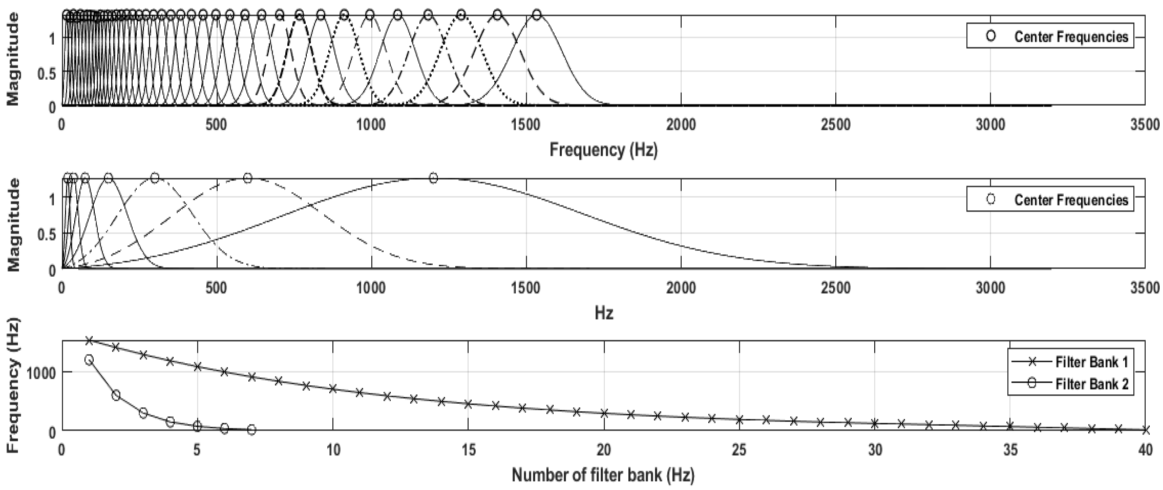


Figure 3. The bandwidth of the wavelets in the 1st and 2nd filter banks with a 3,200 Hz sampling frequency

It is evident that the 40th filter's center frequency for the first filter bank is 52 Hz, while the 39th and 41th filters' center frequencies are 57 and 46 Hz, respectively. Thus, as Figure 4 illustrates, the weight of the fundamental frequency seems to have a large energy for the 40th black node of the first WST coefficients. Figure 5 shows that the weight associated with the fundamental frequency is lowest in the 0th and 2nd order coefficients and highest in the 1st order coefficients. A total of 154 paths have been generated, with 1 being the 0th order, 48 being the 1st order, and 105 being the 2nd order shown by the black nodes in Figure 2. In this case, there are four coefficients in each path. Either the signal length or the invariance scale needs to be adjusted in order to modify the coefficient number in each path for this sampling rate. The number of paths created in the MATLAB toolbox is influenced by the frequency rate; as we mentioned in section 3, the number of paths decreases as the frequency rate does. When both the invariance scale and the signal length are constant, then each path's number of coefficients must also be constant.

However, because the length of the signal is determined by the number of samples rather than a time in the MATLAB toolbox, the time period increases by decreasing the frequency rate, and thus the number of coefficients in each path increases. As shown in Figure 6, if the sampling rate is halved with the same 1,024 samples as a signal length, the number of coefficients in each path doubles. Consequently, since the signal length in time is doubled, there are 109 paths created (1 for the zero-order, 40 for the first order, and 68 for the second order). The first-order scattering coefficients exhibit a prominent appearance of the fundamental frequency weight. Additionally, there are 72 total paths at an 800 Hz sampling rate, with 16 coefficients in each path. In summary, the highest possible sampling frequency with a signal length of 1,024 samples and an invariance scale value of 0.32 is 3,200 Hz. It can be observed that if the sampling frequency is decreased, the number of paths decreases as well, while the number of coefficients in each path increases; hence, the dimensionality of the feature matrix increases for the same signal length in samples, Q-factor, and invariance factor.

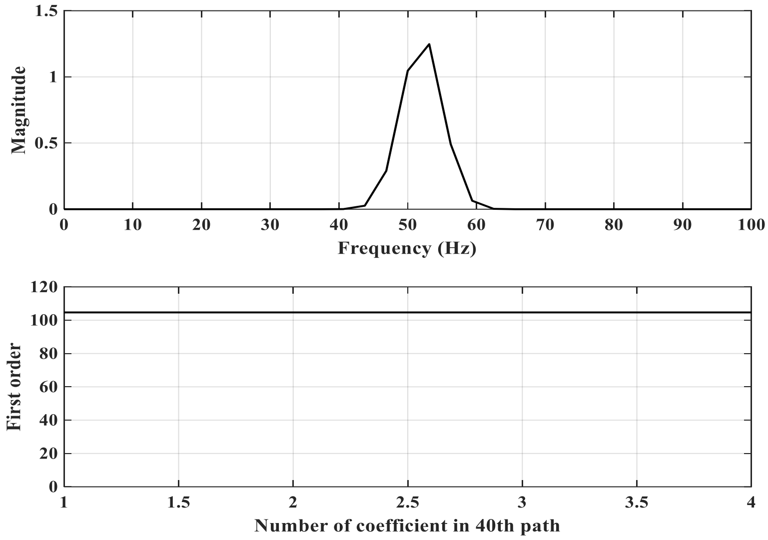


Figure 4. The bandwidth of the 40$^{th}$ wavelet and its WST 1$^{st}$ order coefficients with 3,200 Hz sampling frequency

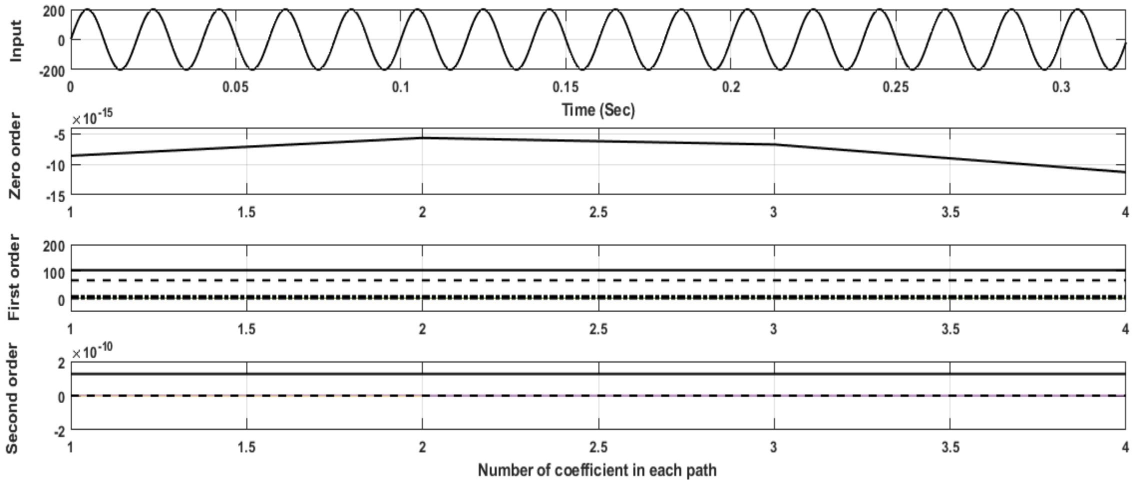


Figure 5. The input fundamental frequency signal, and the 0$^{th}$ (1 node), 1$^{st}$ (48 nodes), and 2$^{nd}$ (105 nodes) order WST coefficients for sampling rate 3,200 Hz, invariance scale 0.32 sec, and signal length 1,024 samples, feature matrix dimension (154×4)

### 4.2. Signal length impacts

Since the signal length in the MATLAB toolbox is actually determined by samples rather than time, any variation in its value with the other parameters held constant should be considered. By increasing the signal length, the number of paths remains constant, but the number of coefficients in each path increases with a constant Q-factor, sampling frequency, and invariance scale. As a result, the dimensionality of the feature matrix increases. Figures 6, 7, and 8 show the zero, first, and second scattering coefficient sets for signal lengths of 1,024, 2,048, and 4,096 for default values of Q-factor, invariance scale 0.32 sec, and sampling frequency of 1,600 Hz. The number of paths for all cases is 109, with 8, 16, and 32 coefficients in each path, respectively, and the amplitude of the coefficients in the feature matrix is slightly altered. Therefore, it is advised to shorten the signal length in order to lighten the computational load.

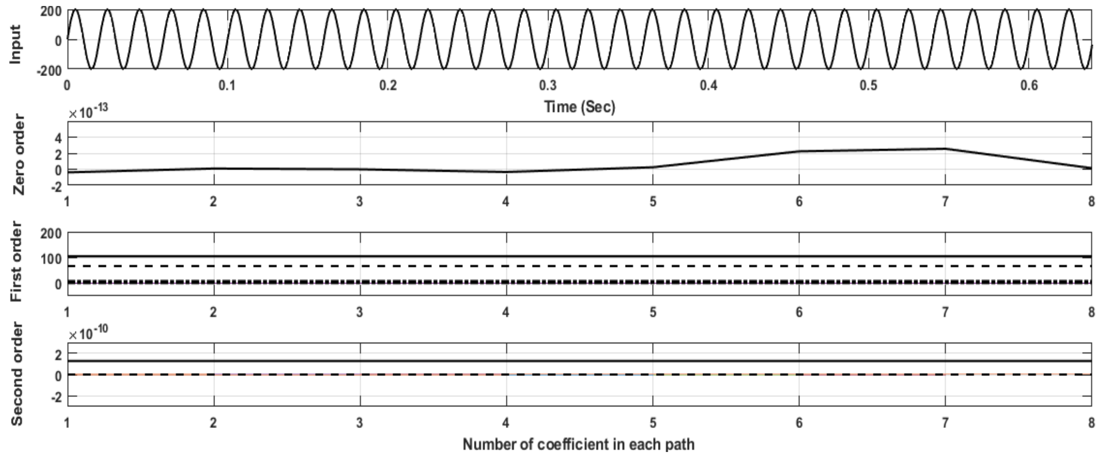


Figure 6. The input fundamental frequency signal, and the $0^{th}$ (1 node), $1^{st}$ (40 nodes), and $2^{nd}$ (68 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.32 sec, and signal length 1,024 samples, feature matrix dimension (109×8)

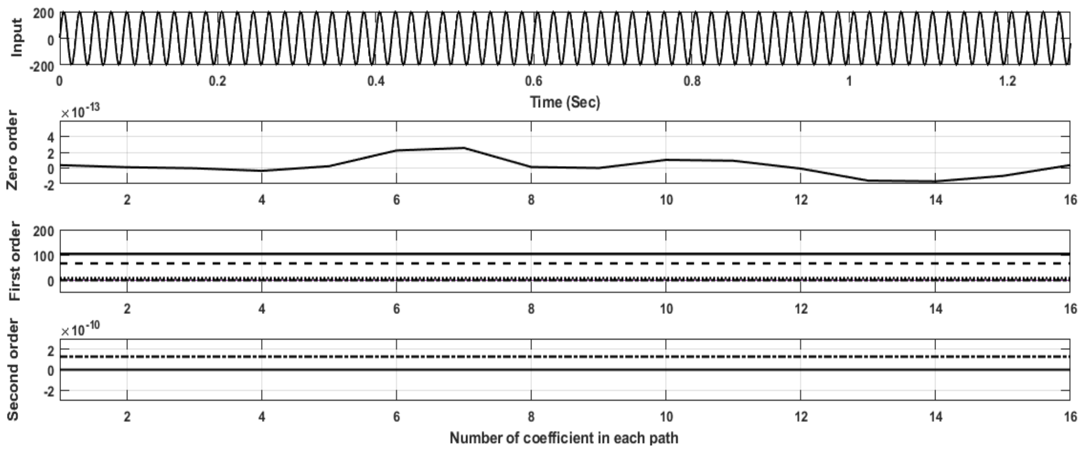


Figure 7. The input fundamental frequency signal, and the $0^{th}$ (1 node), $1^{st}$ (40 nodes), and $2^{nd}$ (68 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.32 sec, and signal length 2,048 samples, feature matrix dimension (109×16)

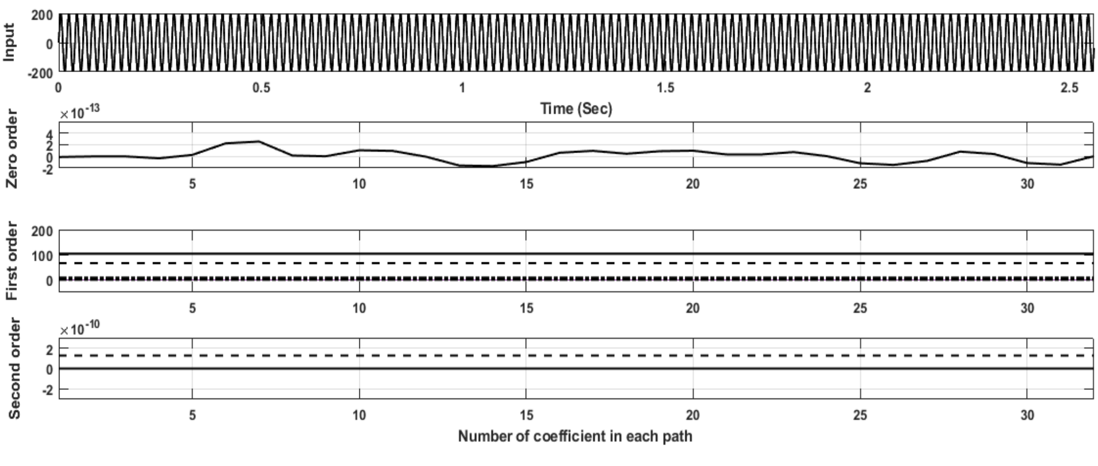


Figure 8. The input fundamental frequency signal, and the $0^{th}$ (1 node), $1^{st}$ (40 nodes), and $2^{nd}$ (68 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.32 sec, and signal length 4,096 samples, feature matrix dimension (109×32)

### 4.3. Quality factors impact

By dividing the frequency span into octaves, the Q-factor determines the number of the wavelet filters per octave that will be created by the WS network in the MATLAB toolbox, just as the sampling rate determines the frequency span that the wavelet filters must cover. The wavelet scattering networks with 154, 130, and 100 paths are created for the signal length 1,024, 0.64 sec invariance scale, 1,600 Hz sampling frequency, and different Q-factors for the first filter bank while keeping the second filter bank as default value ([8, 1]-[6, 1]-[4, 1]), as shown in Figures 9, 10, and 11. The number of created wavelets changes when the Q-factor for the first filter bank changes, as does the number of paths in the first-order WST coefficients, the number of paths in the second-order WST, and the dimensionality of the feature matrix. Furthermore, the change in feature coefficient magnitudes in each path is affected by the change in center frequency filters as well as the degree of fundamental frequency coverage with filter bandwidth. The first-order coefficients for the Q-factor [6, 1] are greater than those for the others [8, 1] and [4, 1]. So, the Q-factor should be carefully defined in accordance with the input signal's dominant frequency and its sampling rate.

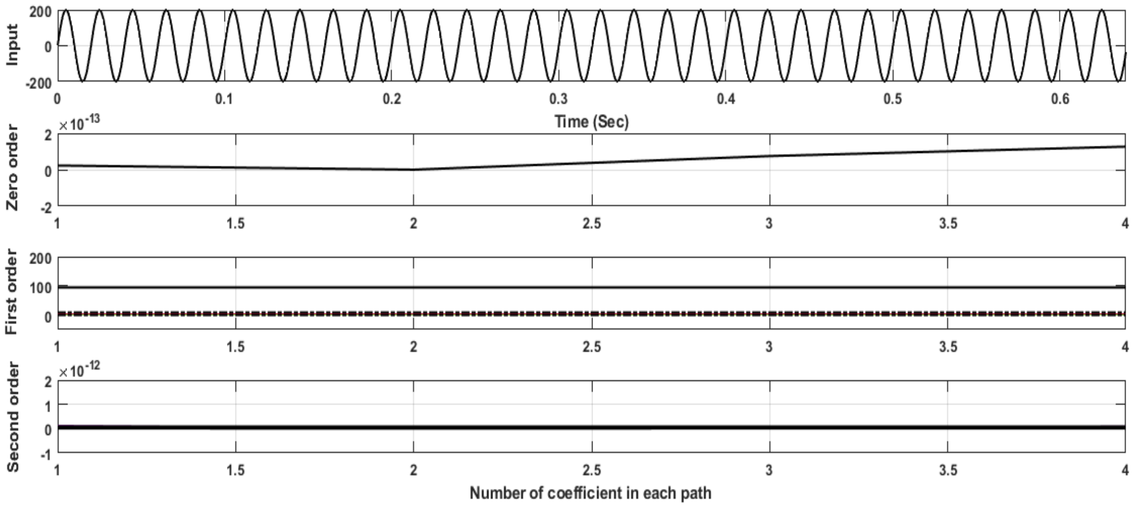


Figure 9. The input fundamental frequency signal, and the 0th (1 node), 1st (48 nodes), and 2nd (105 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.64, and Q-factor [8, 1] and default signal length, feature matrix dimension (154×4)

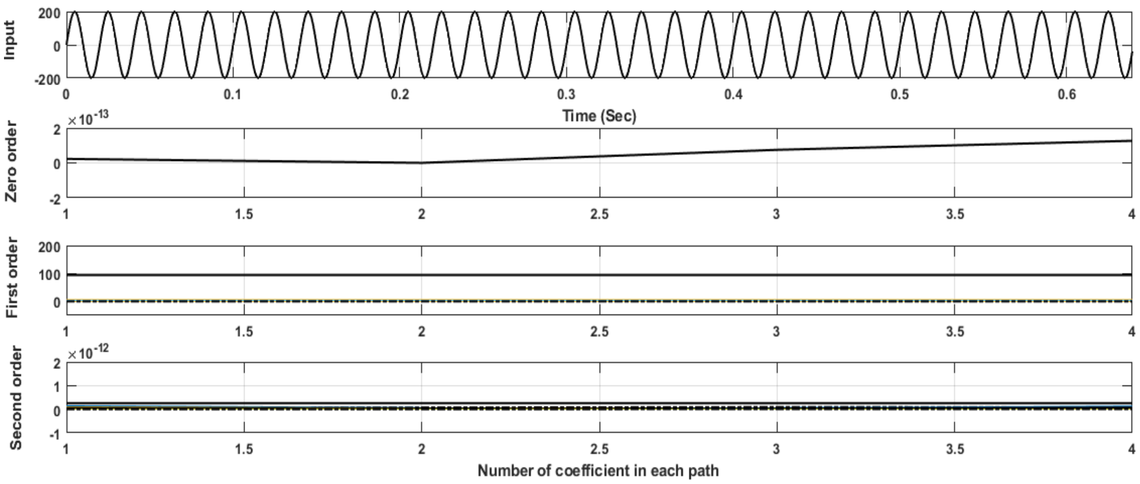


Figure 10. The input fundamental frequency signal, and the 0th (1 node), 1st (39 nodes), and 2nd (90 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.64, and Q-factor [6, 1] and default signal length, feature matrix dimension (130×4)

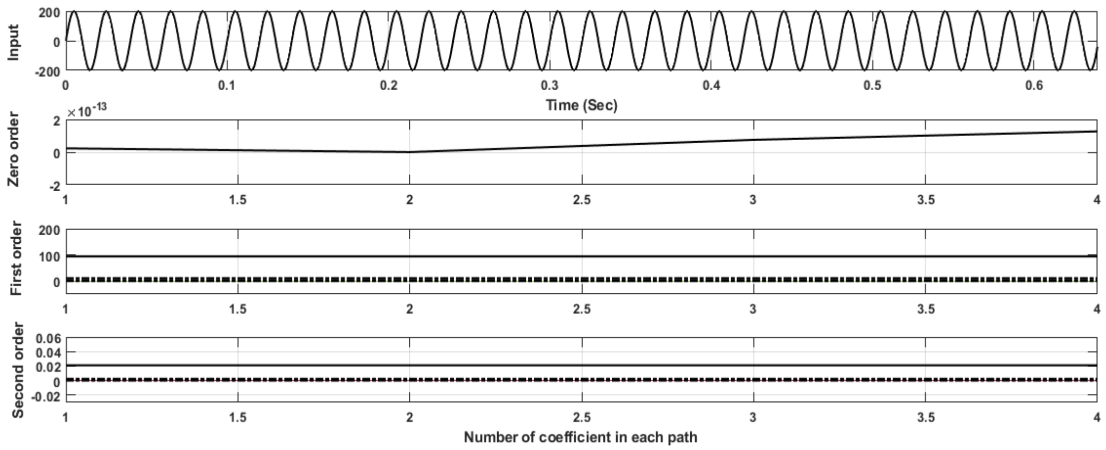


Figure 11. The input fundamental frequency signal, and the 0th (1 node), 1st (27 nodes), and 2nd (72 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.64, and Q-factor [4, 1] and default signal length, feature matrix dimension (100×4)

### 4.4. Invariance scale impacts

The invariance scale parameter is critical in determining how the WS network is created in the MATLAB toolbox. For time series data, the invariance scale is a time duration T, along which the modulus of the CWT coefficients is averaged. Referring to Figure 3, the logarithmic spacing of the higher center frequencies and the linear spacing of the lower center frequencies of the filter banks, indicating that a wavelet's time support cannot exceed the invariance scale (the coarsest-scale wavelet does not exceed the invariance scale determined by the time support of the low-pass filter). As a result, the invariance scale value has a substantial effect on the spacing of the center frequencies of the low-frequency wavelets in the filter banks. For a sampling frequency of 1,600 Hz and the default values of the Q-factor and signal length, the number of paths increases while the number of coefficients in each path decreases due to down sampling; as invariance scale values increase, the dimensionality of the feature matrix decreases; see Figures 12, 13, and 14, which show the WST coefficients of the input signal with three different invariance scale values of 0.16, 0.32, and 0.64 sec.

The amplitudes of the coefficients of the paths of first-order WST increase with noticeable values when the invariance scale decreases, owing to decreasing the time support of the low-pass filter equation $\varphi(t)$. Lowering the invariance scale value increases the number of coefficients in each path and represents signal frequencies with more equal points, hence enhancing the classification process. For power system signal diagnosis, increasing the amplitudes of the feature matrix parameters is beneficial; thus, low values of the invariance scale are advised, with the wavelet's time support of the dominant frequency not exceeding the invariance scale.

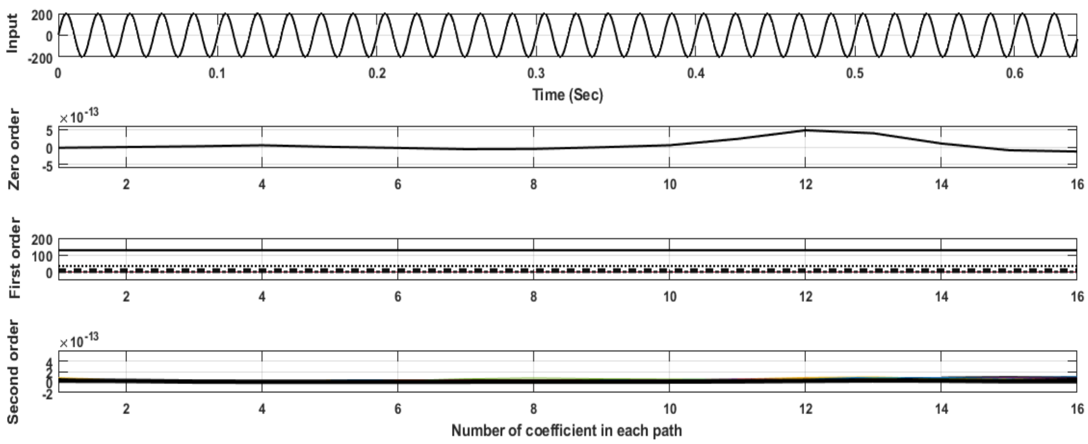


Figure 12. The input signal, and the 0th (1 node), 1st (32 nodes), and 2nd (39 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.16, and default Q-factor and signal length, feature matrix dimension (72×16)

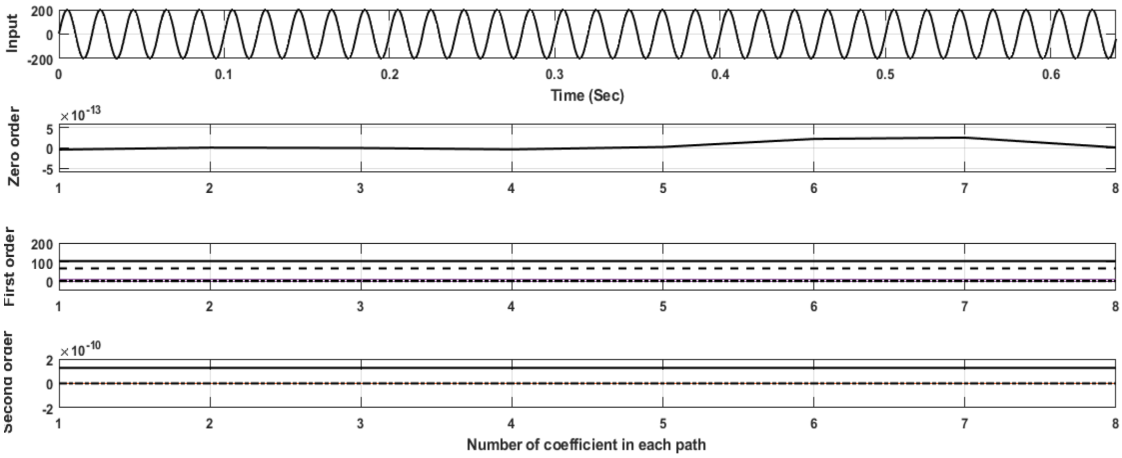


Figure 13. The input signal, and the $0^{th}$ (1 node), $1^{st}$ (40 nodes), and $2^{nd}$ (68 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.32, and default Q-factor and signal length, feature matrix dimension (109×8)

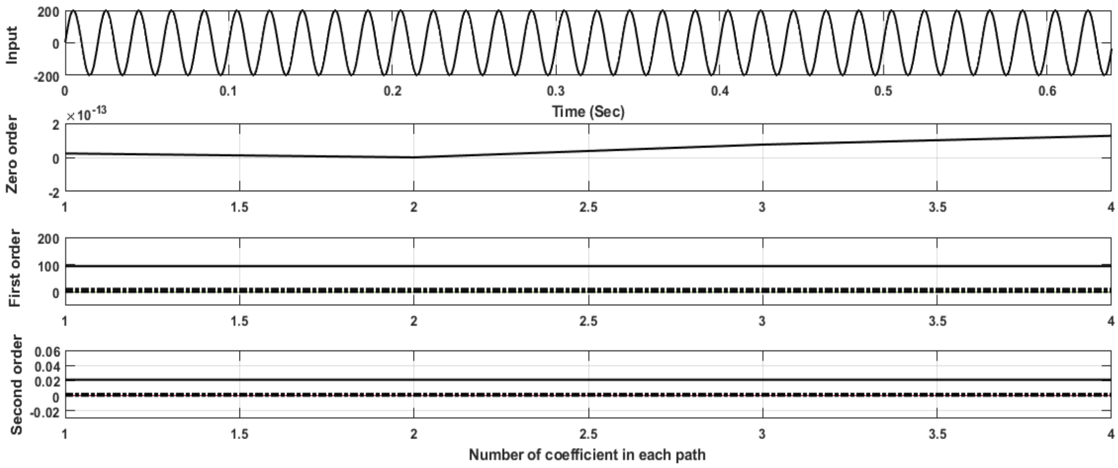


Figure 14. The input signal, and the $0^{th}$ (1 node), $1^{st}$ (48 nodes), and $2^{nd}$ (105 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.64, and default Q-factor and signal length, feature matrix dimension (154×4)

### 4.5. Input signal with DC component

Power system faults are frequently associated with DC components, especially those occurring near power sources. If the input signal contains a fundamental frequency with a peak value of 200 and a constant DC component with a magnitude of 60, the presence of a constant DC component is clearly visible in zero-order WST coefficients but has no discernible effect on the amplitudes of first and second order WST coefficients, as shown in Figure 15. When compared to the input signal's performance with exponentially decayed DC components (whose peak is 60 and time constant is 0.5 sec), the zero-order WST coefficient was able to separate the decaying DC component from the fundamental signal, as shown in Figure 16. The zero-order coefficient pattern clearly distinguishes between the constant and decaying DC components. In power system fault diagnostics, the values of zero-order coefficients in the feature matrix can easily distinguish between failures associated with high DC component values and other ones, i.e. the fault's closeness to the power sources.

### 4.6. Change of the input signal magnitude

Power system faults are typically accompanied with an increase in electrical current magnitudes; hence, this section investigates the change in input signal magnitude. If the peak value of the fundamental frequency input signal is doubled at 0.86 sec, as shown in Figure 17, increasing the input signal magnitude results in increasing the first-order coefficient amplitudes, and the time of change has no effect on coefficient value. So, in power system classification, the WST feature matrix could produce good classification results between normal and abnormal conditions regardless of the time of fault occurrence.

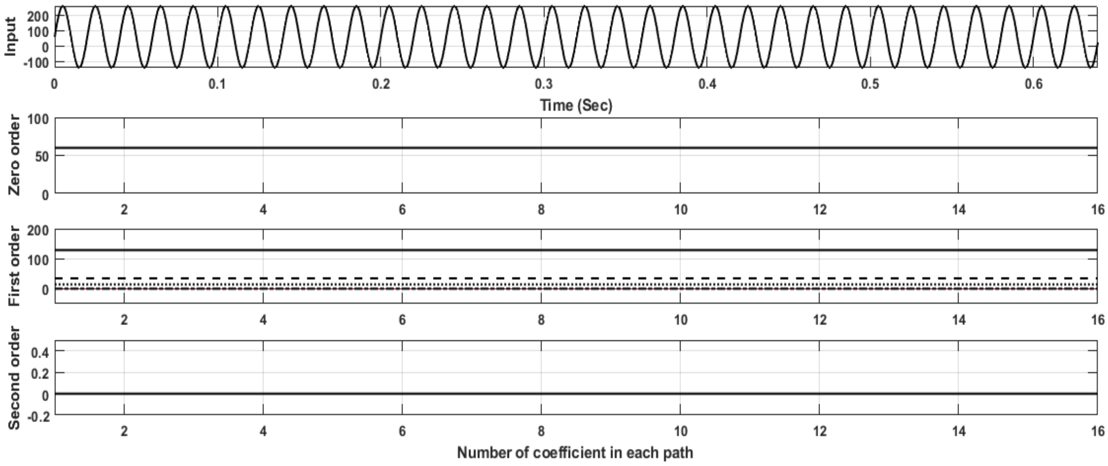


Figure 15. The input signal, and the 0th (1 node), 1st (32 nodes), and 2nd (39 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.16, and default Q-factor and signal length, feature matrix dimension (72×16)

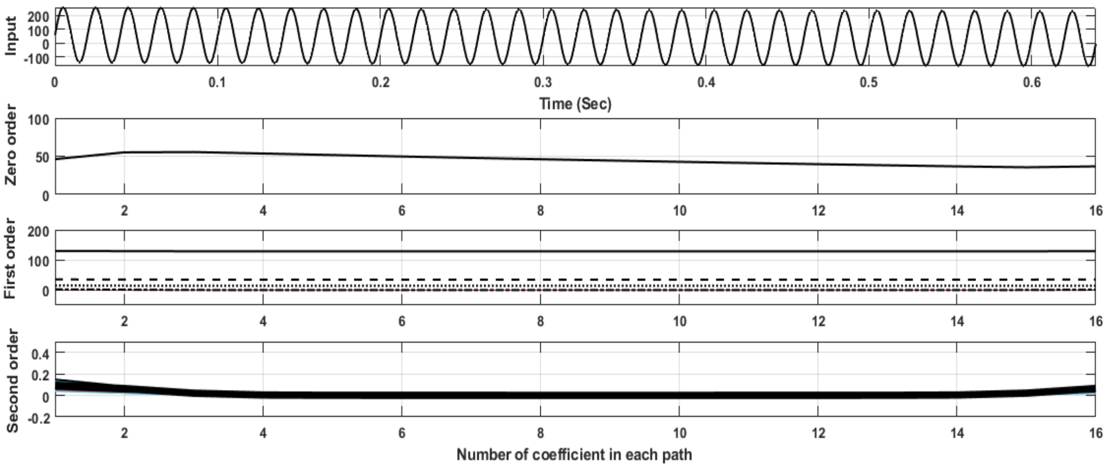


Figure 16. The input signal, and the 0th (1 node), 1st (32 nodes), and 2nd (39 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.16, and default Q-factor and signal length, feature matrix dimension (72×16)

### 4.7. Variation of signal frequency

Reasonable-magnitude harmonics in the power system are a common issue during fault conditions, especially with the high penetration of power electronics-based resources. The ability of WST to discriminate between a pure fundamental signal and one that contains harmonics is tested in this section. As in Figure 18, a nonstationary signal with the 2nd, 3rd, 5th, and 7th harmonics all having the same peak value. A WS network with 0.16 invariance scale, 1,600 Hz sampling frequency, signal length of 2,048 samples, and default Q-factor is created. The four frequency components emerge correctly in the first-order coefficients, each with the same amplitude and reflecting the four paths (black nodes). Each path displays the frequency that falls within its bandwidth with the highest energy value (each harmonic has about 5 coefficients in its path, with the rest 32 being zeros). As a result, the WST feature matrix in this situation could be completely different from the one for an input signal with only a fundamental frequency. Furthermore, when these harmonics are superimposed on the fundamental, as shown in Figure 19, the four harmonics have equal 32 coefficients in their paths in the first-order WST coefficients, which distinguishes them from the other two cases.

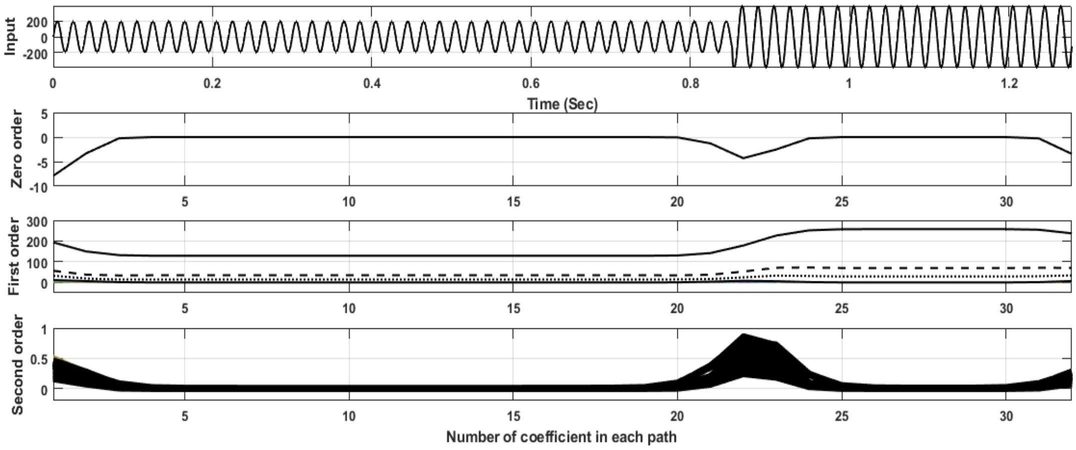


Figure 17. The input signal (step change at 0.86 sec), and the 0th (1 node), 1st (32 nodes), and 2nd (39 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.16, and default Q-factor and signal length, feature matrix dimension (72×32)

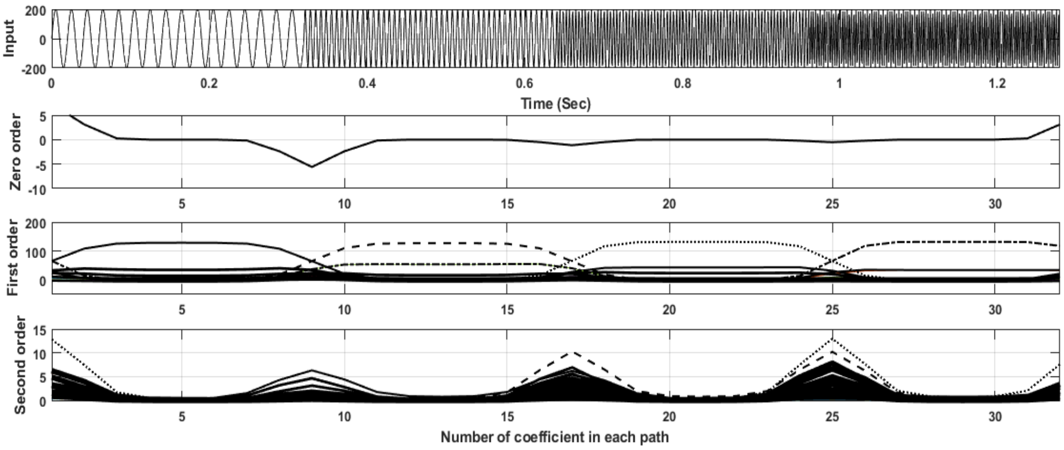


Figure 18. The non-stationary input signal (50, 150, 250, 350 Hz), and the 0th (1 node), 1st (32 nodes), and 2nd (39 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.16, default Q-factor and signal length 2,048 samples, feature matrix dimension (72×32)

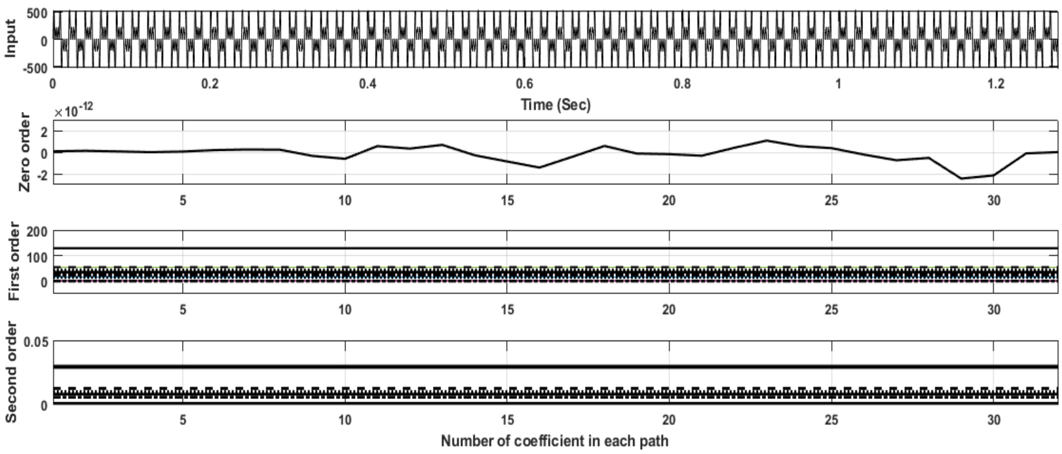


Figure 19. Multi harmonics signal (50, 150, 250, 350 Hz), and the 0th (1 node), 1st (32 nodes), and 2nd (39 nodes) order WST coefficients for sampling rate 1,600 Hz, invariance scale 0.16, default Q-factor and signal length 2,048 samples, feature matrix dimension (72×32)

## 5. PRPOSED LSTM-BASED CLASSIFIER USING WST FEATURE MATRIX

Despite its significant capabilities for discriminating between different power system signal configurations, as demonstrated by the preceding analysis in section 4, WST has yet to be introduced as a preprocessing approach for identifying power system disturbances in the literature. In this study, WST, a feature extraction tool, is introduced for the first time as a preprocessing tool for AI-training of power system disturbances classification. The LSTM network has been introduced in the literature for power system disturbance classifications with promising results, for more details about LSTM network, Sherstinsky introduced its fundamentals in [33].

The LSTM-based classification model using WST is modeled and tested. The workflow of the proposed classification model is shown in Figure 20. Using the MATLAB toolbox, the signals are created, WS network and its feature matrix are implemented, and finally, the training and testing process is carried out.

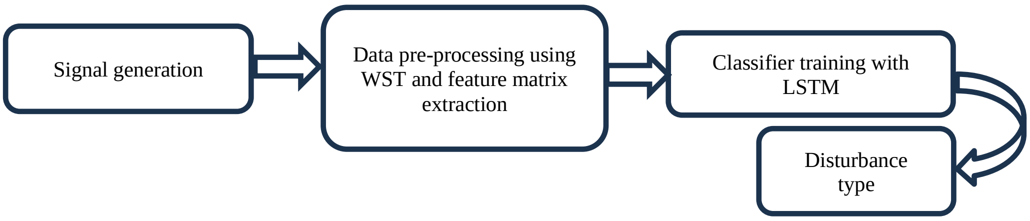


Figure 20. Signals classification model using LSTM

### 5.1. Signal generation

As a power signal simulation, a sine wave is generated with a fundamental frequency of 60 Hz, a peak value of 200, and a sampling frequency of 1,600 Hz. To cover the majority of power system scenarios, a dataset of 250 alternative variants of this signal with varying magnitudes and frequency contents is created. Pure fundamental frequency signal, fundamental frequency signal with DC components, fundamental frequency signal with DC exponential, nonstationary signal containing fundamental, third, fifth, and seventh harmonics, and a third, fifth, and seventh harmonics superimposed on the fundamental frequency signal are all generated. To evaluate the accuracy of the training model, 70% of the data set is used for training and 30% for testing.

### 5.2. Feature extraction

Due to its appealing properties described in section 4, WST is used to construct performance metrics for the training process of the created data set. The parameters for creating a WS network in the MATLAB toolbox are as follows: 1,600 Hz sampling rate, three layers or orders (zero, first, and second), [8, 1] Q-factors for the two filter banks, a 0.16 sec invariance scale, signal length 1,024 samples, and the analytic Morlet wavelet filter. A matrix with 72 paths and 32 coefficients for each path is created. For a 250-signal data set, a feature matrix of 72×32×170 dimensions is built for training and a 72×32×80 matrix for prediction. A sampling frequency has been chosen according to the input signal's fundamental frequency (50 Hz) and the highest frequency that is generated to be superimposed on the fundamental ($7^{th}$ harmonic). To realize the Nyquist criterion and ensure that all signal frequencies are represented in the WST coefficients with high magnitudes, a sampling frequency of 1,600 Hz is adopted. The [8, 1] filter Q-factor is appropriate for providing sufficient wavelet bandwidths capable of covering all signal frequencies. Furthermore, the 0.16 sec invariance scale value is appropriate for 1,024 samples signal length (the scale invariance is half the signal length), resulting in an averaging window of adequate length to ensure the time-invariance property. Finally, the Morlet wavelet was chosen because it can evaluate signals with both short high-frequency transients and long low-frequency components, and it is commonly used for time-frequency analysis of non-stationary time series data. As a result, it is appropriate for power system transients caused by faults.

### 5.3. Results and discussion

WST developed a data set containing the zero, first, and second-order coefficients. These feature matrices are regarded as inputs to the classification model based on the LSTM network. The LSTM network is employed to characterize the long-term dependencies hidden in the time series of the power system fault signals. Figure 21 depicts the LSTM model's confusion matrix, which clearly shows that the training strategy produced excellent results. After 150 iterations, the accuracy reached nearly 100% and the loss was practically zero, with a computation time of roughly 0.24 seconds.

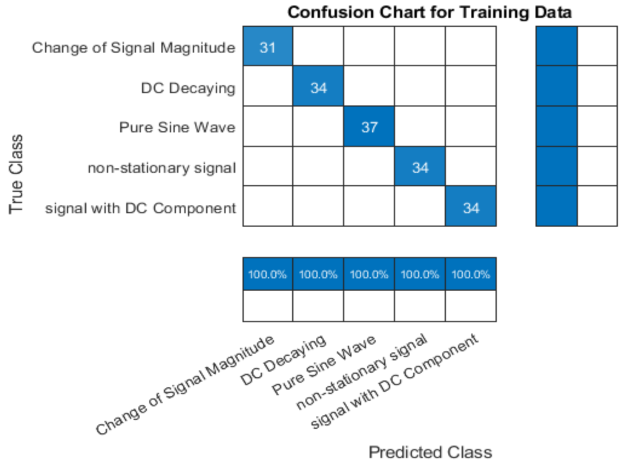


Figure 21. The testing confusion matrix of the signal classification model

Although the DWT-based classifier introduced in the literature [28], [29], [34], [35] delivers acceptable accuracy, DWT is not shift-invariant, so a shift in the input signal does not manifest itself as a simple equivalent shift in the DWT coefficients at all levels. This means that any change in the fault inception angle for the same fault type could be represented by various DWT coefficients at all levels. The shift-invariant property of WST could overcome this shortage, allowing the classifier's accuracy to improve. So, our proposed classifier has demonstrated promising results in increasing the prediction performance of LSTM. Our results demonstrated that the WST will be a promising tool for power system fault diagnosis, especially since the power system fault classification algorithms based on the WST feature matrix have yet to be introduced in the literature. In our study, a classifier model achieves 100% classification accuracy for ideal created power system signals. Even when the dynamic response of a real power system model is addressed, the accuracy will remain satisfactory; this is our future work.

## 6. CONCLUSION

The WST introduced by Mallat in 2012, applies a family of wavelet filters with various scales and translation to an input signal. The WST's main advantage is the use of wavelet filters, which preserve the signal's time and frequency localization. Thus, after applying consecutive convolutions and modulus operations to the time signal, its important features could be extracted. Importantly, the WST gives a well-defined set of scattering coefficients, which can be employed in a robust classification analysis. It is similar to that of CNN; however, it eliminates the need for neural network tuning or training.

Extracting a discriminative feature matrix is the most important stage in the power system fault diagnosis and classification with high accuracy. WST is introduced recently as a powerful feature extraction tool for classification purposes. Although spreading of its applications, the details of the WST feature matrix have not been addressed previously. In this paper, a detailed analysis of the characteristics of the WST feature matrix is introduced, and the factors required to be defined to create a suitable WS network for power system applications in the MATLAB toolbox are studied. The length of the signal, the sampling frequency, the Q-factor of the wavelet filters, and the invariance scale factor should be defined according to the nature of the input signal and its frequency components to implement the WS network capable of extracting the signal details. These four parameters greatly affect the dimension of the feature matrix (number of paths and coefficients of each path) and the coefficient magnitudes, so they should be carefully chosen.

With all parameters held constant, the sampling rate variation substantially affects the number of paths generated by the WS network in the MATLAB toolbox since it adjusts the frequency span that the WS network will cover. If the decomposed signal has high frequencies, then large sample rates are necessary. While, for variation of signal length with the other parameters constants, the number of coefficients in each path is varied. In other words, as its value increased, so did the number of coefficients in each path, allowing the matrix dimension to be changed.

In power system applications, carefully choose the value of the Q-factor of the wavelet filter banks to generate enough wavelets with center frequencies and bandwidth capable of covering all of the important frequency components of the input signal with sufficient energy. Also, the invariance scale value should be carefully chosen to ensure that the wavelet's time support for the dominant frequency does not exceed its value. This value should be chosen in relation to the signal sampling rate for constant signal length; a higher sampling rate needs a low invariance scale value. According to its network implementation, low-frequency components can appear with high coefficient magnitudes in the WST zero-order scattering coefficients. As a result, high-amplitude zero-order WST coefficients could be used to discriminate the faults nearing the power source that are rich with DC components. Furthermore, faults associated with multiple harmonics and stationery and non-stationary signals, such as those seen in renewable energy-based microgrids, could be discriminated with reasonable amplitudes in the WST coefficients, simplifying the classification procedure.

The WS network is recognized as a deep learning technology that is fast, well-understood, computationally inexpensive, and works with a small dataset of training samples to overcome the limitations of the CNN. To demonstrate the capabilities of WST feature matrices to be used as input for the classifier model, a well-structured classification approach based on WST and LSTM is proposed. The proposed algorithm is implemented, and the resulting data set is tested. It was able to classify various power system signals with excellent accuracy, achieving nearly 100% with a modest mathematical burden.

Based on our results, we believe that using WST for feature extraction in power system protection applications will be a promising tool for fault detection and diagnosis, particularly for high impedance faults and the low current faults associated with the islanding operation of micro-grids. Also, the classification of transformer inrush and internal fault currents, and the transmission line fault prediction could be evaluated by WST. In future works, the WST feature matrix will be used for fault detection and diagnosis of the hybrid microgrids in both grid-connected and isolated modes, and the islanding detection applications.

## REFERENCES


[1] S. Mallat, "Group invariant scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, Oct. 2012, doi: 10.1002/cpa.21413.

[2] J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4114–4128, Aug. 2014, doi: 10.1109/TSP.2014.2326991.

[3] J. Bruna and S. Mallat, "Classification with scattering operators," in *CVPR 2011*, Jun. 2011, pp. 1561–1566, doi: 10.1109/CVPR.2011.5995635.

[4] J. Zarka, L. Thiry, T. Angles, and S. Mallat, "Deep network classification by scattering and homotopy dictionary learning," *arXiv preprint arXiv:1910.03561*, 2019.

[5] V. Lostanlen, G. Lafay, J. Andén, and M. Lagrange, "Relevance-based quantization of scattering features for unsupervised mining of environmental audio," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2018, no. 1, Dec. 2018, doi: 10.1186/s13636-018-0138-4.

[6] J. Andén and S. Mallat, "Multiscale scattering for audio classification," *International Society for Music Information Retrieval*, pp. 657–662, 2011.

[7] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013, doi: 10.1109/TPAMI.2012.230.

[8] W. Ghezaiel, L. Brun, and O. Lezoray, "Wavelet scattering transform and CNN for closed set speaker identification," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, Sep. 2020, pp. 1–6, doi: 10.1109/MMSP48831.2020.9287061.

[9] B. Soro and C. Lee, "A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization," *Sensors*, vol. 19, no. 8, Apr. 2019, doi: 10.3390/s19081790.

[10] Y. Jin and Y. Duan, "Wavelet scattering network-based machine learning for ground penetrating radar imaging: application in pipeline identification," *Remote Sensing*, vol. 12, no. 21, Nov. 2020, doi: 10.3390/rs12213655.

[11] A. Sami Al-Itbi, A. Bahaaulddin A. Alwahhab, and A. M. Sahan, "X-Ray Covid-19 detection based on scatter wavelet transform and dense deep neural network," *Computer Systems Science and Engineering*, vol. 41, no. 3, pp. 1255–1271, 2022, doi: 10.32604/csse.2022.021980.

[12] A. A. Susu, H. A. Agboola, C. Solebo, F. E. A. Lesi, and D. S. Aribike, "Wavelet time scattering based classification of interictal and preictal EEG signals," *Brain Res*, vol. 3, no. 115, 2020.

[13] F. Liu *et al.*, "Wearable electrocardiogram quality assessment using wavelet scattering and LSTM," *Frontiers in Physiology*, vol. 13, Jun. 2022, doi: 10.3389/fphys.2022.905447.

[14] H. A. Agboola and J. E. Zaccheus, "Wavelet image scattering based glaucoma detection," *BMC Biomedical Engineering*, vol. 5, no. 1, Mar. 2023, doi: 10.1186/s42490-023-00067-5.

[15] M. Hirn, S. Mallat, and N. Poilvert, "Wavelet scattering regression of quantum chemical energies," *Multiscale Modeling & Simulation*, vol. 15, no. 2, pp. 827–863, Jan. 2017, doi: 10.1137/16M1075454.

[16] A. Rohan, "Deep scattering spectrum germaneness to fault detection and diagnosis for component-level prognostics and health management (PHM)," *Sensors*, vol. 22, no. 23, Nov. 2022, doi: 10.3390/s22239064.

[17] W.-H. Lai, S.-T. Tsai, D.-L. Cheng, and Y.-R. Liang, "Application of wavelet scattering and machine learning on structural health diagnosis for quadcopter," *Applied Sciences*, vol. 11, no. 21, Nov. 2021, doi: 10.3390/app112110297.

[18] E. de Aguiar, A. Lazzaretti, B. Mulinari, and D. Pipa, "Scattering transform for classification in non-intrusive load monitoring," *Energies*, vol. 14, no. 20, Oct. 2021, doi: 10.3390/en14206796.

[19] M. AlBader and H. A. Toliyat, "Wavelet scattering transform based induction motor current signature analysisa," in *2020 International Conference on Electrical Machines (ICEM)*, Aug. 2020, pp. 1452–1457, doi: 10.1109/ICEM49940.2020.9270810.

[20] T. Bourgana, R. Brijder, T. Ooijevaar, and A. P. Ompusunggu, "Wavelet scattering network based bearing fault detection," *PHM Society European Conference*, vol. 6, no. 1, Jun. 2021, doi: 10.36001/phme.2021.v6i1.2875.

[21] R. N. Toma *et al.*, "Classification framework of the bearing faults of an induction motor using wavelet scattering transform-based features," *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22228958.

[22] "Wavelet scattering," *MathWorks*, 2024. https://www.mathworks.com/help/wavelet/ug/wavelet-scattering.html (accessed Jul. 21, 2024).

[23] L. Sifre, J. Anden, M. Kapoko, E. Oyallon, and V. Lostanlen, "ScatNet: a MATLAB toolbox for scattering networks," *DIENS*, 2013. http://www.di.ens.fr/data/software/scatnet/quickstart-audio/ (accessed Jul. 21, 2024).
[24] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2018, doi: 10.1109/ACCESS.2017.2785763.
[25] P. K. Shukla and K. Deepa, "Deep learning techniques for transmission line fault classification – a comparative study," *Ain Shams Engineering Journal*, vol. 15, no. 2, Feb. 2024, doi: 10.1016/j.asej.2023.102427.
[26] S. Karan and H.-G. Yeh, "Fault classification in microgrids using deep learning," in *2020 IEEE Green Energy and Smart Systems Conference (IGESSC)*, Nov. 2020, pp. 1–7, doi: 10.1109/IGESSC50231.2020.9285101.
[27] S. Belagoune, N. Bali, A. Bakdi, B. Baadji, and K. Atif, "Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems," *Measurement*, vol. 177, Jun. 2021, doi: 10.1016/j.measurement.2021.109330.
[28] V. Veerasamy *et al.*, "LSTM recurrent neural network classifier for high impedance fault detection in solar PV integrated power system," *IEEE Access*, vol. 9, pp. 32672–32687, 2021, doi: 10.1109/ACCESS.2021.3060800.
[29] N. W. Branco, M. S. M. Cavalca, S. F. Stefenon, and V. R. Q. Leithardt, "Wavelet LSTM for fault forecasting in electrical power grids," *Sensors*, vol. 22, no. 21, Oct. 2022, doi: 10.3390/s22218323.
[30] A. M. S. Omar, M. K. Osman, M. N. Ibrahim, Z. Hussain, and A. F. Abidin, "Fault classification on transmission line using LSTM network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 1, pp. 231–238, Oct. 2020, doi: 10.11591/ijeecs.v20.i1.pp231-238.
[31] O. Cortes-Robles, E. Barocio, E. Beltran, and R. D. Rodríguez-Soto, "Events classification in power systems with distributed generation sources using an LSTM-based method with multi-input tensor approach," *Electricity*, vol. 4, no. 4, pp. 410–426, Dec. 2023, doi: 10.3390/electricity4040022.
[32] S. Xiong, T. Gao, X. Zhao, and Z. Fei, "Multi-fault integrated classification and identification method based on LSTM," in *2023 IEEE 4th International Conference on Electrical Materials and Power Equipment (ICEMPE)*, May 2023, pp. 1–4, doi: 10.1109/ICEMPE57831.2023.10139519.
[33] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, Mar. 2020, doi: 10.1016/j.physd.2019.132306.
[34] K. Gu, Y. Zhang, X. Liu, H. Li, and M. Ren, "DWT-LSTM-based fault diagnosis of rolling bearings with multi-sensors," *Electronics*, vol. 10, no. 17, Aug. 2021, doi: 10.3390/electronics10172076.
[35] N. Qu, Z. Li, J. Zuo, and J. Chen, "Fault detection on insulated overhead conductors based on DWT-LSTM and partial discharge," *IEEE Access*, vol. 8, pp. 87060–87070, 2020, doi: 10.1109/ACCESS.2020.2992790.


## BIOGRAPHIES OF AUTHORS

**Naema M. Mansour** received the B.Sc. and M.Sc. degrees in electrical engineering from Minufiya University, Egypt in 2000 and 2006, respectively, and the Ph.D. degree in electrical engineering in 2011 from Minufiya University. Currently, lecture at Suez Canal University, Egypt. Research interests include power system protection, signal processing application in power system protection, distributed and renewable energy resources, utilization of optimization algorithms for power system protection enhancement, and power system control and management. She can be contacted at email: naima.mansour@eng.suez.edu.eg.

**Ibrahim A. Awaad** received the B.Sc. degrees in electrical engineering from Electrical Engineering Department, Faculty of Engineering Sciences, Sinai University, El-Arish, Egypt 2017. Currently he is assistant lecturer at Faculty of Engineering Sciences, Sinai University, Arish Branch, Arish 45511, Egypt. He is studying for a master's degree at Suez Canal University, Egypt. Research interests include power system protection, control, WST application, AI applications in power system, and micro-grid fault detection and diagnosis. He can be contacted at email: ebrahimawad92@gmail.com.

**Abdelazeem A. Abdelsalam** is professor at Suez Canal University, Egypt. He was a post doctorate fellow at University of Ontario Institute of Technology (UOIT), Canada. He received his B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from Suez Canal University, Egypt in 2001, 2005 and 2011, respectively. He is a member of the IEEE. He has authored or coauthored more than 60-refereed journal and conference papers and three book chapters, His research areas include power quality, D-FACTS technology, switched filter compensators, micro-grid interface and control and application of artificial intelligence techniques in power systems. He can be contacted at email: aaabdelsalam@eng.suez.edu.eg.