Global attention-based LSTM for noisy power quality disturbance classification

Dar Hung Chiam, King Hann Lim, Kah Haw Law

Global attention-based LSTM for noisy power quality disturbance classification

Dar Hung Chiam* and King Hann Lim

Department of Electrical and Computer Engineering,
Curtin University Malaysia,
Miri, Sarawak, 98009, Malaysia
Email: chiamdh@postgrad.curtin.edu.my
Email: glkhann@cutin.edu.my
*Corresponding author

Kah Haw Law

Electrical and Electronic Engineering Programme Area,
Universiti Teknologi Brunei,
Gadong, Brunei Darussalam
Email: kahhaw.law@utb.edu.bn

Abstract: An increased dependency of digital control systems in the modern electrical network demand for a better quality of power signal. The occurrence of power quality disturbances (PQDs) in the network reduces the lifespan of power semiconductors and solid states switching devices. Global attention-based long short-term memory (LSTM) network is proposed to perform automatic time-series PQD detection and classification. Attention-based LSTM helps improve the noise immunity to extract salient features from noisy signal for PQD classification. The aim of this article is to analyse the performance of proposed attention-based LSTM under different noise conditions. Addictive white Gaussian noise is added to synthetic PQDs in different signal-to-noise ratio. These random generated noises are used to train and test the performance of proposed method, as well compared towards generic LSTM model. This work also shows the sensitivity of proposed method towards unknown noises that is not seen by the model during training phase.

Biographical notes: Dar Hung Chiam is currently pursuing his PhD in Electrical and Computer Engineering at the Curtin University Malaysia. He received his Bachelor of Engineering (BEng) in Electrical and Electronic

Engineering from the Curtin University Malaysia in 2018. His research interests include power quality analysis, machine learning and artificial intelligence.

King Hann Lim received his Master of Engineering (MEng) and PhD in Electrical and Electronic Engineering in 2007 and 2012 respectively. He is currently a staff member in the Department of Electrical and Computer Engineering at Curtin University Malaysia. His research area includes image/video processing, artificial intelligence, and intelligent transportation system. He is a IEEE senior member in 2017. He has published more than 70 journals and conference papers in the related areas of his research interest.

Kah Haw Law is working as an Assistant Professor under Faculty of Engineering at the Universiti Teknologi Brunei (UTB), Brunei Darussalam. He received his PhD in Electrical and Electronic Engineering from the University of Nottingham in 2015. His research interests include renewable control systems, electric power systems, power electronics and drive systems, and advanced power conversion/inversion systems. His current research work involves modelling and implementation of novel inverter topology for renewable power generation, specifically for STATCOM system on protocol design and performance aspects.

This paper is a revised and expanded version of a paper entitled 'Noise level evaluation on power quality disturbances classification' presented at 2021 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), Curtin, Malaysia, 7–9 July 2021.

---

1 Introduction

Power quality (PQ) monitoring is an essential service performed by many utility companies for their industrial and larger commercial customers (Aramwanid and Boonyaroonate, 2015). Unexpected variation of voltage or current from normal condition may frequently happen in electrical networks (Baraniak and Starzyński, 2020). These anomalies are commonly caused by the increased complexity structure of power systems involving the integration of renewable energy resources and microgrids (Chen et al., 2021). As stated in the IEEE standard 1159-1995 (EM Committee et al., 2009), PQ disturbances (PQDs) cover wide range of abnormal signal phenomena. These PQDs includes transient (impulsive and oscillatory), short duration variations (interruption, sag and swell), frequency variations, long duration variations (sustained under voltages and sustained over voltages) and steady state variations (harmonics, notch and flicker). Frequent occurrence of PQDs in the networks increases the risk of electric shortage and reduces the lifespan of electrical components. Advanced metering infrastructure (AMI) and phasor measurement units (PMU) allows continuous monitoring of the power systems performance. Smart meters and utility monitoring devices were introduced to make use of the massive raw data collected (Zhou et al., 2016). The backbones behind these smart metering structures are algorithms that can perform desired analysis. Machine learning has been introduced to perform automatic detection and classification of anomalies presence in power systems (Zhang and Liu, 2008; Janik and Lobos, 2006).

Figure 1   The proposed global attention-based LSTM network

Detection and classification of PQDs can be generalised into two stages, i.e., feature extraction and classification stage. Feature extraction stage generally detects the most distinctive features to characterise an occurrence of PQDs. Feature extraction process are normally aided with signal processing techniques. Some of the commonly applied signal processing techniques includes fast Fourier transform (FFT) (Heydt et al., 1999), discrete Fourier transform (DFT) (Szmajda et al., 2007), short time Fourier transform (STFT) (Jurado and Saenz, 2002), wavelet transform (WT) (Jurado and Saenz, 2002; Zhu et al., 2004), S-transform (ST) (Zhao and Yang, 2006), and wavelet-packet-transform (WPT) (Panigrahi and Pandi, 2009). Feature extraction process is performed by capturing the signal statistics that can be used to identify the class of the signals. Some of the commonly used statistical parameters includes minimum, maximum, entropy, standard deviation, mean, RMS, skewness, range, and kurtosis. These extracted features are used to ease the identification of different classes of PQDs during the classification process (Wang and Chen, 2019; Lee et al., 1997; Khokhar et al., 2017; Kumar et al., 2015; Bhavani and Prabha, 2017; Liu et al., 2019). The selection of these hand-crafted features requires expert knowledge to avoid information losses. The classification accuracy is highly dependent on the predefined filter specifications, selection of unique features and accurate pattern recognition methods (Balouji et al., 2018).

Distinctive features are extracted and subsequently used in decision making mechanism during classification process. Widely used classification techniques include support vector machine (SVM) (Naderian and Salemnia, 2017), expert system (Muthusamy and Ramanathan, 2018; Ahila et al., 2015), and artificial neural networks (Chen et al., 2016; Alshahrani et al., 2016). SVM can achieve high accuracy with least amount of training but the accuracy is highly dependent on the training samples. Expert systems can be used with limited data samples but the implementation cost is high with slow execution. Artificial neural networks have advantage in getting high accuracy for real time application. However, the presence of noise has high impact on the accuracy. The main focus of current research is to get highest accuracy with minimum of resources spent. In order to achieve real-time classification, deep neural networks are used for both feature extraction and classification in Wang and Chen (2019) and Deng et al. (2018). These automatic feature extraction processes removed the needs of complex feature selection stages and allowed the detection of unseen conditions for accurate classification. Previous work of Chiam et al. (2021) showed that single layer LSTM model trained with noise could achieve high classification accuracy of 84.87% under 20 dB SNR AWGN. However, training with real-time PQD integrated with real time noise would increase the complexity of the harvesting and training process.

Global attention-based LSTM is thus proposed for time-series PQD classification. The proposed method consists of an attention mechanism for highlighting salient features, a layer of LSTM architecture for feature extraction and a fully connected layer with softmax activation function for classification. Global attention mechanism is added in between input signal and LSTM layer to highlight the characteristics of input signals. The feature vector output from attention layer is used by LSTM to extract higher order

representation from the sequence of signal's magnitude. The entire automatic detection and classification process is demonstrated in Figure 1. Sliding window is used for the pre-processing of the raw power data. The classification performance of generic LSTM model and proposed global attention-based LSTM model is compared and studied. The comparisons of training with and without AWGN noise is done to analyse the strength and weaknesses of the proposed method. The analysis is further studied by introducing unknown noises to the models.

2 Global attention-based long short-term memory network

Long short-term memory (LSTM) is used with an attention layer to improve classification accuracy under noisy condition. The proposed method includes an attention layer between input data and LSTM layer. This attention layer is used to highlight the abnormalities presented in single-windowed (global attention) power signal before feeding into LSTM layer. The important features in the raw signal is first valued in attention layer, then passed into LSTM for higher dimensional feature extraction. The proposed model is summarised in Figure 2. Pre-processed data were passed into attention layer before feeding into LSTM for feature extraction. A fully connected (FC) layer is then used to classify the higher dimensional features output from LSTM layer into different classes of PQDs. The network input vector X retrieves the shifting signal as follows,

$$X = [b_i, ..., b_{i+T}], \tag{1}$$

where b_i contains the value of the original signal at the i-th position up to $i + T$ timestep, where T is the window size defined by the period of signal $t - 1$.

Figure 2   Global attention-based LSTM model with signal windowing
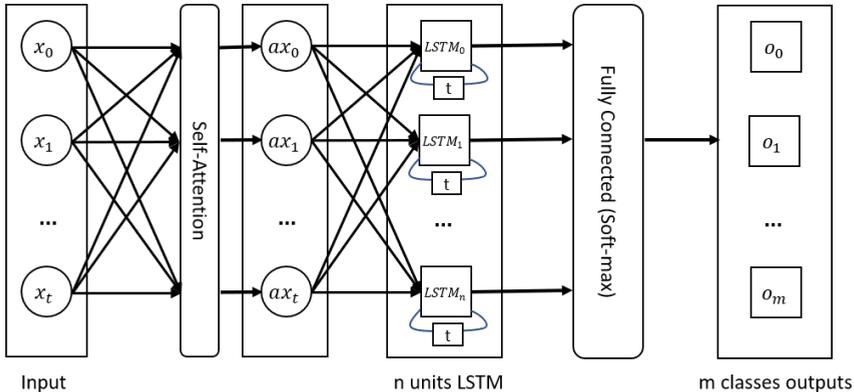
2.1 Global attention mechanism

Attention mechanism is proposed to highlight specific feature in the signal before feeding into LSTM for feature extraction process. Attention weights calculated is

multiplied to all data points present in each input sample, thus forming a global attention mechanism. This global attention mechanism is achieved by a dense layer with an softmax activation function as shown in Figure 3. Input signal x_t with T window size or timesteps is first passed into a dense layer to obtain the attention score, y_d as follows,

$$y_d = \sum_{t=0}^{T} w_{d,t} \cdot x_t, \tag{2}$$

where x_t represents the input signal, $w_{d,t}$ is the trainable weights kernel vectors of the dense layer. A softmax layer is used to normalise the attention score calculated into range between $(0, 1)$ as,

$$a_d = softmax(y_d) = \frac{e^{y_d}}{\sum_{j=0}^{T} e^{y_{d_j}}}. \tag{3}$$

Finally, the attention weights, a_t is being multiplied element-wise with the input signal to highlight the signal. This highlighted feature vector, $a_d x_t$ can be expressed as,

$$ax_t = a_d \odot x_t. \tag{4}$$

Attention score represents the important features through a dense layer. The attention weights are multiplied with input signal, highlighting the original signals with trained weight distribution. These highlighted feature vectors have higher noise immunity compared to raw signal without attention mechanism.

Figure 3   Self-attention mechanism (see online version for colours)
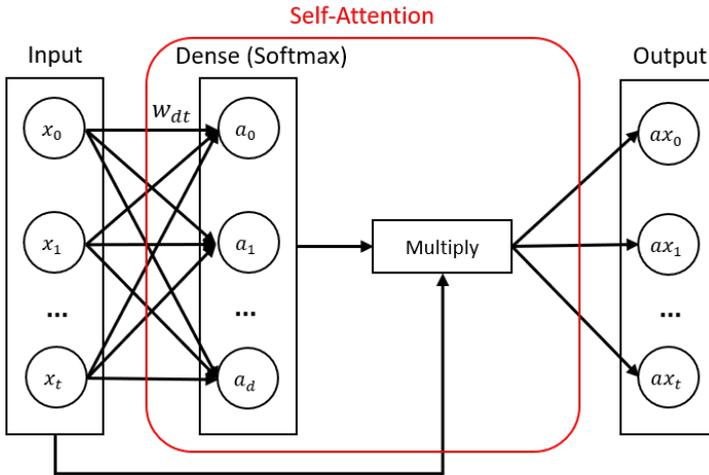
2.2   Long short-term memory

Feature vector output from attention layer is passed into LSTM layer for feature extraction. An LSTM architecture can be depicted as in Figure 4 (Hochreiter and

Schmidhuber, 1997). The inputs to the LSTM cell consists of: previous memory or cell state, C_{t-1}, previous hidden state, h_{t-1}, and the current input, x_t. Sequential input passed into the LSTM will be processed through multiple gates which control the information flow. There are three 'gates' presence in LSTM cell: forget gate, f_t, input gate, i_t, and output gate, o_t. The information is first passed through forget gate, where unwanted information is erased, and a value of 1 signifies keeping all previous memory. Input gate decides which information to retain. The output of input gate will be filtered by the tanh activation function, producing a new candidate, $\tilde{c}$ for the cell state, while W represents trainable weights and b is the bias.

$$\begin{pmatrix} f_t \\ i_t \\ o_t \\ \tilde{c}_t \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \cdot [h_{t-1}, x_t] + b. \tag{5}$$

Figure 4 LSTM architecture
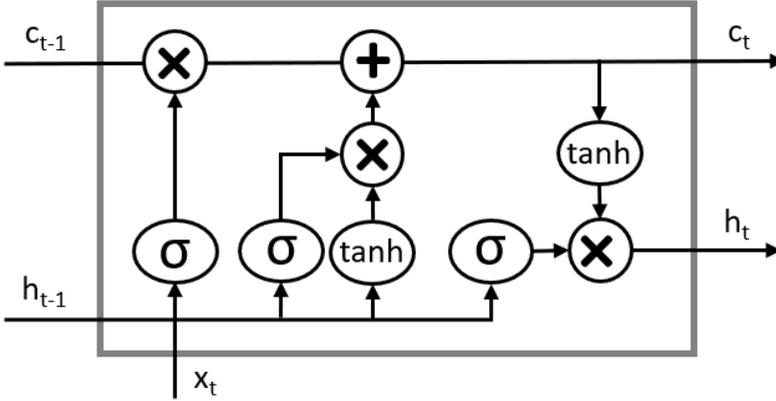
A new cell state, c_t is produced at every time step. The new cell state, c_t is achieved by forgetting irrelevant information while learning new information. The equation below shows the cell state updating mechanism. Previous cell state, c_{t-1} is multiplied element-wise with the forget gate to remove the unwanted information. At the same time, the candidate cell state, $\tilde{c}_t$ is multiplied element-wise with the input gate control.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t. \tag{6}$$

The third gate in an LSTM cell is the output gate, o_t. This output gate controls the output information from an LSTM cell. The information output or LSTM hidden state output, h_t is based on cell state. A tanh activation function is used to squeeze the cell state information into a range of (–1, 1). Then, a sigmoid activation function is used in output gate, o_t to decide the output content of the cell states. The output gate and hidden state output can be calculated as follows,

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{7}$$

$$h_t = o_t \odot \tanh(c_t). \tag{8}$$

The temporal features of the input signal will be extracted by LSTM layer. These temporal features representing the specific feature of the PQDs are encoded into higher dimension representation which will be classified into PQDs classes by the classifier dense layer. In our experiment, the pre-processed, 1-dimension single-period windowed power waveform which consists of 200 timesteps is used as input. The important characteristic of different classes of PQD is first highlighted in attention layer. The output from attention layer is fed parallel into 200 units of LSTM. The final hidden state output from LSTM representing extracted features are passed into fully connected (FC) layer for the classification process.

3   Data generation and pre-processing

PQD signals are generated using mathematical model (Liu et al., 2019; Tang et al., 2019) as listed in Table 1. Ten classes of single disturbance PQDs including class $P0$: normal, and eight classes of combined disturbances from class $P10$ to $P17$ are trained for automatic PQD classification. 200 random samples of three-period waveform are generated based on the modelling equations in Table 1. Sampling frequency of 10kHz is used in these experiments. Point-labelling is used where the magnitude of the signal waveform is compared with normal waveform and a difference with more than $\pm 0.005\%$ are labelled as specific disturbance class. The generated PQDs are displayed in Figure 6, where the square wave represents point labelling. Multiple disturbances are labelled according to the index. For example, class $P2$: swell is having an index of 2.0, thus label amplitude at 2.0; while class $P15$: swell+transient is having label amplitude of 2.4 at Swell+Transient region.

Figure 5   PQD windowing and labelling process (see online version for colours)
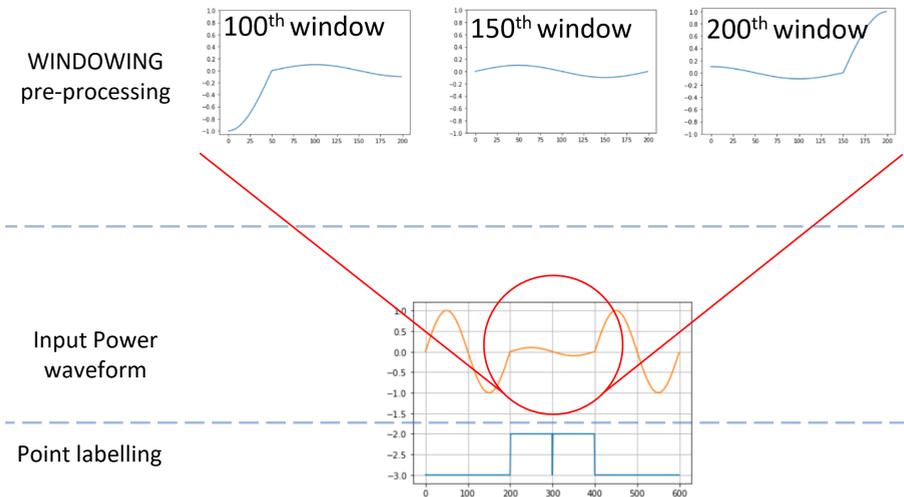
Table 1 PQD data generation

Labels	PQD	Mathematical equations and parameters
P0	Normal	$y(t) = A[1 \pm \alpha(u(t - t_1) - u(t - t_2))]\sin(\omega t)$ Parameters: $\alpha \leq 0.1; T \leq t_2 - t_1 \leq 3T$
P1	Sag	$y(t) = A[1 - \alpha(u(t - t_1) - u(t - t_2))]\sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T$
P2	Swell	$y(t) = A[1 + \alpha(u(t - t_1) - u(t - t_2))]\sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T$
P3	Interrupt	$y(t) = A[1 - \alpha(u(t - t_1) - u(t - t_2))]\sin(\omega t)$ Parameters: $0.9 < \alpha \leq 1; T \leq t_2 - t_1 \leq 3T$
P4	Impulse transient	$y(t) = A[1 - \alpha(u(t - t_1) - u(t - t_2))]\sin(\omega t)$ Parameters: $0 \leq \alpha \leq 0.414; T/20 \leq t_2 - t_1 \leq T/10$
P5	Spike	$y(t) = \sin(\omega t) + \mathrm{sign}(\sin(\omega t)) \times [\sum_{n=0}^{8} K \times [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.02n))]]$ Parameters: $0 \leq t_1, t_2 \leq 0.5T; 0.01T \leq t_2 - t_1 \leq 0.05T; 0.1 \leq K \leq 0.4$
P6	Harmonics	$y(t) = A[\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i^2 = 1$
P7	Oscillatory transient	$y(t) = A[\sin(\omega t) + \alpha^{e^{-\frac{(t-T/2)}{\tau}}} \sin \omega_\pi(t - t_1)(u(t_2) - u(t_1))]$ Parameters: $0.1 \leq \alpha \leq 0.8; 0.5T \leq t_2 - t_1 \leq 3T; 8 \text{ ms} \leq \tau \leq 40 \text{ ms}; 300 \leq f_n \leq 900 \text{ Hz}$
P8	Notch	$y(t) = \sin(\omega t) - \mathrm{sign}(\sin(\omega t)) \times [\sum_{n=0}^{8} K \times [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.02n))]]$ Parameters: $0 \leq t_1, t_2 \leq 0.5T; 0.01T \leq t_2 - t_1 \leq 0.05T; 0.1 \leq K \leq 0.4$
P9	Flicker	$y(t) = [1 + \alpha_f \sin(\beta\omega t)]\sin(\omega t)$ Parameters: $0.1 \leq \alpha_f \leq 0.2; 5 \text{ Hz} \leq \beta \leq 20 \text{ Hz}$

Source: Liu et al. (2019) and Tang et al. (2019)

Table 1 PQD data generation (continued)

Labels	PQD	Mathematical equations and parameters
P10	Sag + harmonics	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))][\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i{}^2 = 1$
P11	Swell + harmonics	$y(t) = A[1 + \alpha(u(t-t_1) - u(t-t_2))][\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i{}^2 = 1$
P12	Interrupt + harmonics	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))][\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.9 < \alpha \leq 1; T \leq t_2 - t_1 \leq 3T; 0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i{}^2 = 1$
P13	Harmonics + notch	$y(t) = A[[\sin(\omega t)) - \text{sign}(\sin(\omega t)) \times [\sum_{n=0}^{8} K \times [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.02n))]]] + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t)$ $+ \alpha_7 \sin(7\omega t)]$ Parameters: $0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i{}^2 = 1; 0 \leq t_1, t_2 \leq 0.5T; 0.01T \leq t_2 - t_1 \leq 0.05T; 0.1 \leq K \leq 0.4$
P14	Sag + transient	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))][1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0 \leq \alpha \leq 0.414; T/20 \leq t_2 - t_1 \leq T/10$
P15	Swell + transient	$y(t) = A[1 + \alpha(u(t-t_1) - u(t-t_2))][1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0 \leq \alpha \leq 0.414; T/20 \leq t_2 - t_1 \leq T/10$
P16	Sag + oscillatory transient	$y(t) = A[1 - \alpha(u(t_2) - u(t_1))] \sin(\omega t) + \alpha^e \dfrac{-(t-T/2)}{\tau} \sin \omega_\pi (t - t_1)(u(t_2) - u(t_1))$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.1 \leq \alpha \leq 0.8; 0.5T \leq t_2 - t_1 \leq 3T; 8 \text{ ms} \leq \tau \leq 40 \text{ ms}; 300 \leq f_n \leq 900 \text{ Hz}$
P17	Swell + oscillatory transient	$y(t) = A[1 + \alpha(u(t_2) - u(t_1))] \sin(\omega t) + \alpha^e \dfrac{-(t-T/2)}{\tau} \sin \omega_\pi (t - t_1)(u(t_2) - u(t_1))$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.1 \leq \alpha \leq 0.8; 0.5T \leq t_2 - t_1 \leq 3T; 8 \text{ ms} \leq \tau \leq 40 \text{ ms}; 300 \leq f_n \leq 900 \text{ Hz}$

Source: Liu et al. (2019) and Tang et al. (2019)

Figure 6 Waveform and labelling for all PQDs, (a) P0: normal (b) P1: sag (c) P2: wwell (d) P3: interrupt (e) P4: impulse transient (f) P5: spike (g) P6: harmonics (h) P7: oscillatory transient (i) P8: notch (j) P9: flicker (k) P10: sag + harmonics (l) P11: swell + harmonics (m) P12: interrupt + harmonics (n) P13: harmonics + notch (o) P14: sag + transient (p) P15: swell + transient (q) P16: sag + oscillatory transient (r) P17: swell + oscillatory transient (see online version for colours)
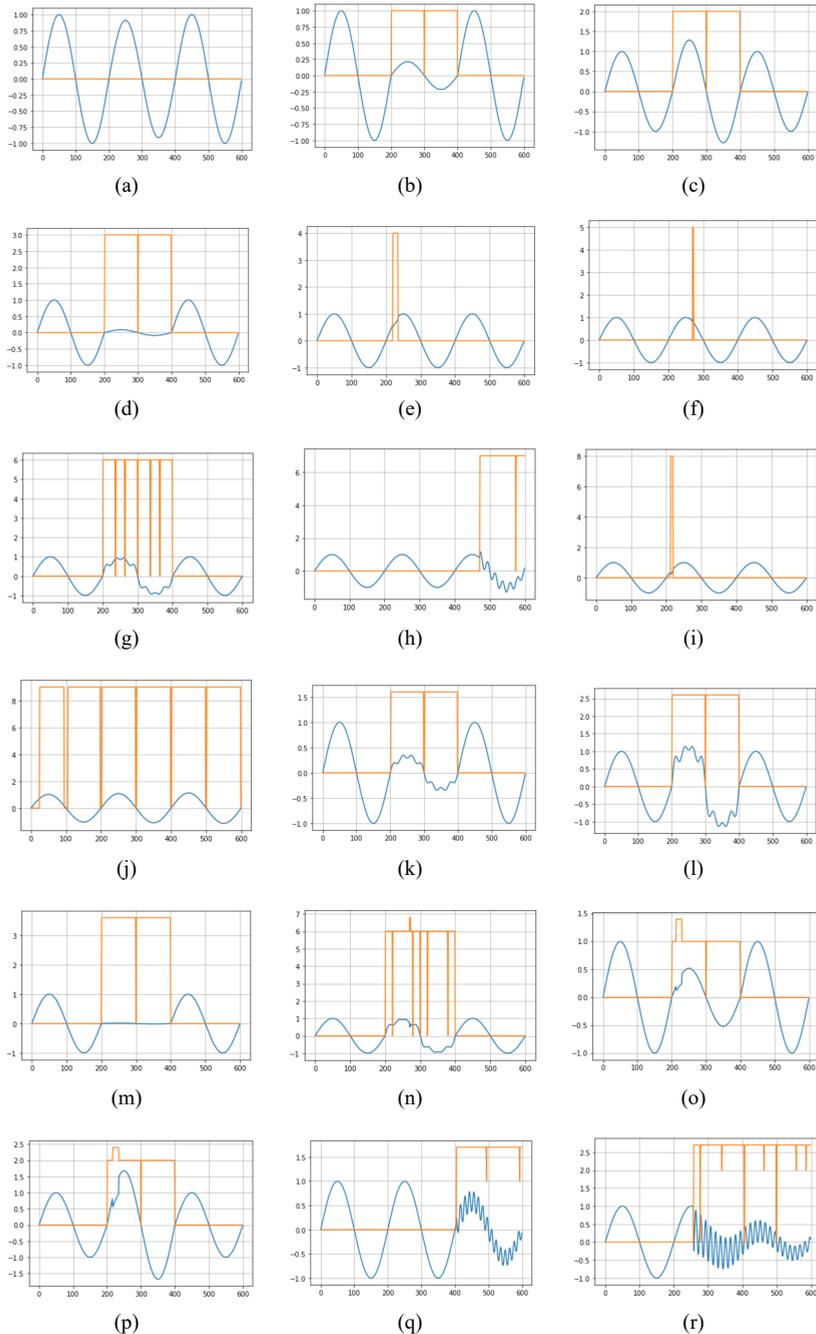
The two stages in data pre-processing include data segmentation and data formatting. Sliding window technique is used to segment the voltage signal as shown in Figure 5. A single period window size equivalent to 0.02 seconds of the 50 Hz voltage waveform is used. The time-step of the sliding window or 'stride' used is kept at one. Window-labelling is labelled according to the occurrence of the point-labelling. Each windowed data consists of two information, magnitude of signal and 18 categories one-hot-encoded window labelling. Data formatting is done before the input of windowed data into the proposed global attention LSTM architecture. Normalisation of the signal magnitude has been carried out during this data formatting stage. Normalisation has been done by dividing the signal magnitude with maximum amplitude which can be expressed as,

$$V(t) = \frac{v(t)}{\max\limits_{t \in n} v(t)},$$ (9)

where n represents number of windows and t represents window size. The formatted input, $V(t)$ is the normalisation of magnitude $v(t)$ over maximum magnitude present in the entire data sample.

4   Experiment setup

The classification was conducted using Keras with Tensorflow backend. The formatted input vector is partitioned into 70% training data, 15% validation data and 15% testing data. Two experiments are conducted to analyse PQD detection and classification using LSTM network. The first experiment is carried out to analyse the performance of the proposed global attention-based LSTM network model. Signal-to-noise (SNR) of 20–50 dB addictive white Gaussian noise (AWGN) are added into the testing samples to evaluate the model under noisy condition. Second experiment is carried out to study the training performance with added randomised 20–50 dB and noiseless samples during the model training session. The trained model is first tested under four seen conditions, which are 20–40 dB SNR AWGN and noiseless condition. The model is further exposed to three unseen noisy conditions, noise A with 15 dB SNR AWGN, noise B with 20–25 dB SNR positive-uniformly distributed random noise, and noise C with 15–30 dB SNR uniformly distributed random noise. The calculation of SNR can be depicted as follows,

$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}}.$$ (10)

The main evaluation matrix used in these experiments is the classification accuracy. The classification accuracy of individual class Acc_n is the true positive, TP_n over the total test samples for m classes of PQD, S_j as,

$$Acc_n = \frac{TP_n}{\sum_{j=0}^{m} S_j}.$$ (11)

Weighted accuracy (WAcc) is used to overcome the imbalanced data sample used. The weightage of each class is calculated by dividing the total number of samples of an

individual class over the total number of samples. Weighted accuracy is calculated by multiplying the individual class accuracy with its weight, which can be depicted as,

$$WAcc = \sum_{j=0}^{m} W_j \times Acc_j. \tag{12}$$

5 Experiment 1 – performance comparison of LSTM network versus global attention-based LSTM

Experiment 1 is evaluated using noiseless training dataset. The performance analysis is compared among four groups of testing samples which includes the original noiseless signal and AWGN with SNR ranging from 20–40 dB. Global attention mechanism is proposed to highlight the input signal. The input signal and output attention vector can be depicted as in Figure 7. Figure 7 shows the highlighted feature vector for oscillatory transient and Interrupt under 20 dB SNR and no noise condition. Specific feature vector are highlighted for each of the disturbance classes regardless of the AWGN level. This shows that the attention layer successfully learn on how to highlight the specific feature for each of the disturbance waveform. From 20 dB SNR to noiseless condition, there is only slight changes in the magnitude of the highlighted features. This slight changes indicates that the attention mechanism is having a good immunity against noisy condition.

Figure 7 Oscillatory transient time-series signal with its attention output at (a) 20 dB level (b) no-noise level; and interrupt time-series signal with its attention output at (c) 20 dB level (d) no-noise level (see online version for colours)
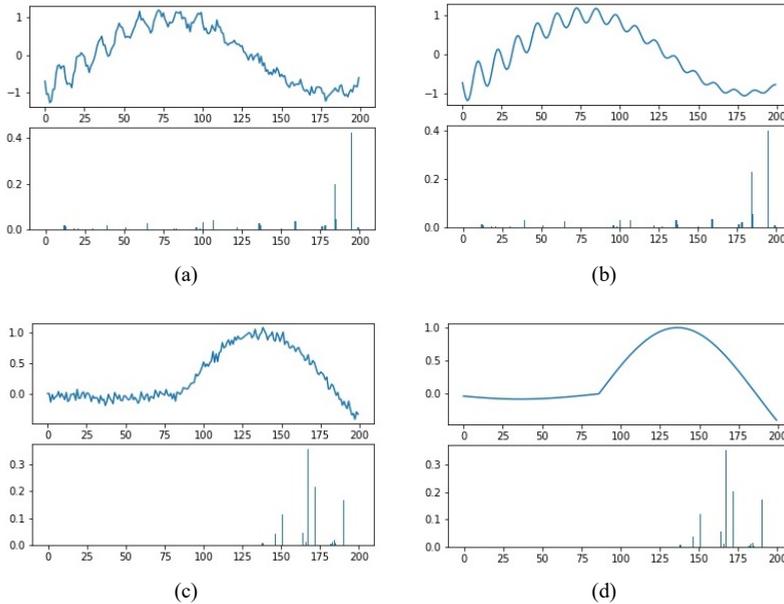
PQD classification using LSTM model tested under 20 dB SNR AWGN gives the WAcc of 18.99%. LSTM model classifier are biased towards class *P*7: oscillatory transient and class *P*11: swell+harmonics where both are having high similarities with the nature of noisy signal. The LSTM model without Attention mechanism is thus prone to noise and only perform better under noiseless condition. On the other hand, the classification performance of the proposed global attention-based LSTM model tested under 20–40 dB SNR AWGN increased significantly to 51.33%, 78.37%, and 87.54% respectively, as shown in Table 2. The proposed global attention-based model is affected by the 20 dB noise but still maintain the WAcc above 50%. This result proves that the attention mechanism has been successfully highlighted the important features under noisy condition. The added attention mechanism generalised the model by highlighting the signal with attention vector.

Although the proposed attention mechanism shows better overall results, it is not deniable that it comes with some weaknesses. From Table 2, it can be noticed that class *P*16: sag+oscillatory Transient and *P*17: swell+oscillatory transient are having difficulty in classification. It can be noticed that these two classes are very similar in nature, except only magnitude difference. Class *P*12: interrupt+harmonics has harmonics event occurred with the interrupt period. The Harmonics magnitude level is relatively small compared to the normal signal condition before or after the *P*12 event. The proposed attention mechanism is good at highlighting overall picture of the input signal, but insensitive to small magnitude changes.

Table 2 Classification performance of LSTM model trained with noiseless synthetic PQD data and tested with 20–40 dB AWGN and noiseless conditions

Attention		Without attention				With attention			
SNR Class		*20 dB*	*30 dB*	*40 dB*	*Noiseless*	*20 dB*	*30 dB*	*40 dB*	*Noiseless*
P0		0.00	0.00	6.25	99.86	14.78	71.46	97.67	99.31
P1		0.67	36.04	55.29	96.34	57.23	84.88	91.75	92.50
P2		4.58	44.18	62.92	98.61	67.65	89.91	96.58	97.67
P3		9.30	22.72	45.58	99.15	44.91	69.58	83.01	87.20
P4		0.37	28.67	96.68	98.47	40.90	77.18	85.48	86.66
P5		0.15	25.69	72.79	98.92	49.58	77.57	83.63	84.39
P6		0.00	4.86	91.07	99.36	61.75	89.86	94.84	95.63
P7		99.33	98.74	90.74	96.82	72.23	78.73	79.47	79.37
P8		0.62	30.74	68.68	97.72	43.06	66.16	72.60	73.91
P9		0.01	3.65	51.83	99.28	58.21	95.98	99.02	99.17
P10		8.98	57.87	89.87	96.45	79.62	90.99	92.34	92.53
P11		82.04	94.38	97.11	98.51	92.99	97.52	98.14	98.22
P12		57.41	86.48	93.35	94.96	47.26	59.83	65.20	66.82
P13		41.71	97.76	99.09	99.27	67.85	78.29	81.17	81.15
P14		18.23	80.24	98.37	97.45	66.17	81.44	84.80	85.04
P15		25.87	56.59	97.60	98.73	76.32	86.16	88.11	88.21
P16		71.93	78.48	83.47	88.14	54.59	55.30	54.98	55.03
P17		58.38	65.44	71.55	87.07	35.14	35.40	35.45	35.52
WAcc		18.99	38.54	62.53	97.69	51.33	78.37	87.54	88.54

6 Experiment 2 – noise analysis on the proposed global attention-based LSTM network

In experiment 2, the models trained with dataset consisting of noiseless and 20–50 dB SNR AWGN. The model can learn the temporal features from noisy signal if the noisy dataset is listed during the training process. The AWGN added into the training samples are learnt as a new type of feature posed on the samples. Subsequently this trained model is tested with added 20–40 dB AWGN, and three types of unseen noises, which are labelled as noise A, noise B, and noise C. Data with noise A is generated randomly with 15 dB AWGN to evaluate the model performance on samples that are not exposed during the training process. Data inserting with noise type B assimilates the situation where an unknown uniform positive noise has offset into the DC level of the signal. On the other hand, data added with noise type C has the unknown uniformly randomised noise with the magnitude ranging from [–0.3, 0.3]. This is to simulate a scenario where some parasitic components occurred in the system that cannot be picked up easily.

Table 3 Classification performance of LSTM model trained with noisy synthetic PQD data and tested with 20–40 dB AWGN and noiseless conditions

Attention		Without attention				With attention			
Class	SNR	20 dB	30 dB	40 dB	Noiseless	20 dB	30 dB	40 dB	Noiseless
P0		92.14	98.85	99.64	99.75	55.05	97.46	99.32	99.41
P1		84.94	94.35	96.77	97.26	71.85	87.06	89.00	88.99
P2		89.59	95.62	97.13	97.45	82.08	94.60	96.35	96.47
P3		67.09	83.48	91.16	94.21	37.54	40.24	41.29	41.55
P4		66.77	88.25	94.04	94.79	52.19	75.05	78.06	78.44
P5		78.98	97.74	98.47	98.42	51.47	65.20	67.45	67.79
P6		88.99	98.29	99.08	99.25	75.67	89.60	91.51	91.87
P7		93.96	96.07	96.22	96.18	74.11	77.48	77.90	77.94
P8		76.22	97.02	97.80	97.80	47.70	61.80	63.10	63.67
P9		94.10	99.41	99.82	99.81	87.05	97.77	98.55	98.69
P10		87.56	94.10	95.42	95.59	84.76	90.85	91.58	91.64
P11		96.69	98.01	98.13	98.07	95.04	97.20	97.44	97.47
P12		67.64	82.81	91.55	95.06	67.28	78.64	81.63	82.15
P13		86.82	97.11	98.17	98.16	71.56	77.31	78.06	78.07
P14		66.54	84.47	92.14	93.49	62.60	69.94	70.06	69.94
P15		72.75	89.54	95.02	95.49	76.64	81.51	81.83	81.54
P16		81.17	86.46	86.54	86.46	44.03	44.22	44.43	44.53
P17		82.64	88.30	89.57	89.73	50.64	51.60	51.79	51.83
WAcc		84.87	94.22	96.43	96.97	66.67	82.89	84.30	84.45

The classification performance of both models with and without attention mechanism is compared in Table 3. Both models are set up using 100 units of hidden units and trained for 30 epochs. Both models are showing better classification performance compared to models trained under noiseless condition. The classification accuracy of LSTM model without attention improved from the previous 18.99% to 84.87% WAcc under 20 dB

SNR AWGN. However, the classification performance of Global attention-based LSTM model does not show drastic improvement when trained under noisy condition, from 51.33% to 66.67%. From Table 3, it can be noticed that class $P16$: sag+oscillatory transient and $P17$: swell+oscillatory transient are still having similar difficulty in classification. By comparing to previous train with noiseless experiment, it can also be noticed that class $P12$: interrupt+harmonics showing improved performance but reduce performance in class $P3$: interrupt. Training with added noise will change the nature of the signal being trained, especially on the magnitude of the signal. The proposed global attention mechanism is proved to be good in picturing global features, but insensitive to slight magnitude changes.

Three unknown noises were used to test the trained model. The comparison result is tabulated in Table 4. Unknown noise A is an AWGN with 15 dB SNR. Although the trained model can achieve good accuracy when training with SNR 20–50 dB, it does not perform well when unknown noise A having lower SNR level noise is used to test the model. The classification accuracy for noise A achieves 14.75% in model trained without noise and without attention mechanism. Training with mixture of noiseless and 20–50 dB noise without attention only improves the accuracy to 47.86%, which is still far from expected 84.87%. Comparatively, despite affecting by higher level of noise, model with attention mechanism can still achieve 36.01% and 48.07% under training without and with noise respectively.

Noise B contains positive uniform random noise ranging within 20–25 dB SNR. This positive uniform random noise simulates the DC level drift or offset when in the real-life application. It is presented as a new variant of noise to test the trained model. The model without attention achieved 47.07% and 52.09% under training without and with noise condition respectively. The proposed global attention-based LSTM model however achieved stable classification performance of 59.09% and 59.10% respectively for training without and with noise respectively. These classification performance are close to the average classification performance of testing with 20 dB SNR AWGN condition. This result shows that the proposed attention model is still reliable with different variant of noise, as long as the noise level are constant.

In addition, noise type C with uniform random amplitude from [–0.3, 0.3] is added to the original signal to further test the model's performance. This signal presented an uniformly distributed noise with 11–30 dB SNR range. LSTM model without attention mechanism achieve 20.06% and 53.94% under noiseless training and noisy training respectively. The classification performance of the proposed global attention-based model achieves consistence performance of 43.06% and 54.09% respectively. Overall, the proposed global attention mechanism has comparatively consistent, and better classification performance for unseen noise condition.

Table 4 Unseen noise performance comparison

Train Attention Noise	Without noise			With noise		
	A	B	C	A	B	C
Without	14.75	47.07	20.06	47.86	52.09	53.94
With	36.01	59.09	43.06	48.07	59.10	54.09

7   Conclusions

Global attention-based LSTM network is proposed in this paper for more generalised PQD classification. The presented method used global attention mechanism to highlight the important feature present in the input signal. Temporal feature of the highlighted signal is extracted using single layer LSTM. Two experiments using various noise level signals are conducted in this paper to evaluate the performance of global attention-based LSTM network model. Results from experiment 1 show the global attention mechanism aids in highlighting the global feature of an input signal. The classification performance under high noise condition of 20 dB SNR AWGN shows improvement from 18.99% without attention mechanism, to 51.33% aided with the proposed global attention mechanism. This shows that the proposed global attention mechanism is good in highlighting global features, while having less impact on the noises. Experiment 2 is conducted to further analyse the proposed model by training with noisy signal consisting random mixture of noiseless and AWGN with SNR levels of 20–50 dB. The performance of the model without global attention improves from 18.99% to 84.87% under 20 dB condition. However, the performance of the proposed global attention model only improves from 51.33% to 66.67%. The limitation of the proposed model is thus uncovered, where global attention mechanism are insensitive to condition with slight magnitude changes. Subsequently, the experiment continues by testing with three unseen noises. The results show the proposed global attention-based LSTM model having more generalised classification performance, where it can achieve averagely better classification accuracy on signals polluted with unknown noise. As a conclusion, global attention-based LSTM network model is a more general approach compared to typical LSTM model. In practical scenario, the harvest of large amount noisy data in the grid may posses high complexity in the experimental setup, which is not feasible in the real-time applications. As for future work, global attention can be aided with multiple resolution scaling tools to extract salient features from multiple resolution levels. A generative training mechanism such as generative adversarial network could be adopted to learn the distribution of samples using unsupervised manner to cover large set of data variation.

References

Ahila, R., Sadasivam, V. and Manimala, K. (2015) 'An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances', *Applied Soft Computing*, Vol. 32, pp.23–37 [online] https://doi.org/10.1016/j.asoc.2015.03.036.

Alshahrani, S., Abbod, M. and Taylor, G. (2016) 'Detection and classification of power quality disturbances based on Hilbert-Huang transform and feed forward neural networks', *2016 51st International Universities Power Engineering Conference (UPEC)*, IEEE, pp.1–6.

Aramwanid, P. and Boonyaroonate, I. (2015) 'Power quality impact study and analysis of electrical power efficacy in sugar industry', *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp.1–4, IEEE.

Balouji, E., Gu, I.Y., Bollen, M.H., Bagheri, A. and Nazari, M. (2018) 'A LSTM-based deep learning method with application to voltage dip classification', *2018 18th International Conference on Harmonics and Quality of Power (ICHQP)*, IEEE, pp.1–5.

Baraniak, J. and Starzyński, J. (2020) 'Modeling the impact of electric vehicle charging systems on electric power quality', *Energies*, Vol. 13, No. 15, p.3951.

Bhavani, R. and Prabha, N.R. (2017) 'A hybrid classifier for power quality (PQ) problems using wavelets packet transform (WPT) and artificial neural networks (ANN)', *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, IEEE, pp.1–7.

Chen, S., Zhang, J., Wang, L., Zhang, H. and Li, L. (2021) 'Evaluation of power quality and reliability of distributed generation in smart grid', *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, Vol. 632, p.042015.

Chen, Z., Li, M., Ji, T. and Wu, Q. (2016) 'Detection and classification of power quality disturbances in time domain using probabilistic neural network', *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp.1277–1282.

Chiam, D.H., Lim, K.H. and Law, K.H. (2021) 'Noise level evaluation on power quality disturbances classification', *2021 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, IEEE, pp.1–6.

Deng, Y., Jia, H., Li, P., Tong, X. and Li, F. (2018) 'A deep learning method based on long short term memory and sliding time window for type recognition and time location of power quality disturbance', *2018 Chinese Automation Congress (CAC)*, IEEE, pp.1764–1768.

EM Committee et al. (2009) *IEEE Recommended Practice for Monitoring Electric Power Quality*, IEEE Std, pp.c1–81.

Heydt, G., Fjeld, P., Liu, C., Pierce, D., Tu, L. and Hensley, G. (1999) 'Applications of the windowed FFT to electric power quality assessment', *IEEE Transactions on Power Delivery*, Vol. 14, No. 4, pp.1411–1416.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural computation*, Vol. 9, No. 8, pp.1735–1780.

Janik, P. and Lobos, T. (2006) 'Automated classification of power-quality disturbances using SVM and RBF networks', *IEEE Transactions on Power Delivery*, Vol. 21, No. 3, pp.1663–1669.

Jurado, F. and Saenz, J.R. (2002) 'Comparison between discrete STFT and wavelets for the analysis of power quality events', *Electric Power Systems Research*, Vol. 62, No. 3, pp.183–190.

Khokhar, S., Zin, A.A.M., Memon, A.P. and Mokhtar, A.S. (2017) 'A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network', *Measurement*, Vol. 95, pp.246–259 [online] https://doi.org/10.1016/j.measurement.2016.10.013.

Kumar, R., Singh, B. and Shahani, D.T. (2015) 'Recognition of single-stage and multiple power quality events using Hilbert-Huang transform and probabilistic neural network', *Electric Power Components and Systems*, Vol. 43, No. 6, pp.607–619.

Lee, C., Lee, J., Kim, J-O. and Nam, S.W. (1997) 'Feature vector extraction for the automatic classification of power quality disturbances', *Proceedings of 1997 IEEE International Symposium on Circuits and Systems. Circuits and Systems in the Information Age ISCAS'97*, IEEE, Vol. 4, pp.2681–2684.

Liu, H., Hussain, F., Yue, S., Yildirim, O. and Yawar, S.J. (2019) 'Classification of multiple power quality events via compressed deep learning', *International Transactions on Electrical Energy Systems*, Vol. 29, No. 6, p.e12010.

Muthusamy, T.A. and Ramanathan, N. (2018) 'An expert system based on least mean square and neural network for classification of power system disturbances', *Int. J. Futur. Revolut. Comput. Sci. Commun.*, Vol. 4, No. 1, pp.308–313.

Naderian, S. and Salemnia, A. (2017) 'An implementation of type-2 fuzzy kernel based support vector machine algorithm for power quality events classification', *International Transactions on Electrical Energy Systems*, Vol. 27, No. 5, p.e2303.

Panigrahi, B. and Pandi, V.R. (2009) 'Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbour algorithm', *IET Generation, Transmission & Distribution*, Vol. 3, No. 3, pp.296–306.

Szmajda, M., Górecki, K. and Mroczka, J. (2007) 'DFT algorithm analysis in low-cost power quality measurement systems based on a DSP processor', *2007 9th International Conference on Electrical Power Quality and Utilisation*, IEEE, pp.1–6.

Tang, Q., Qiu, W. and Zhou, Y. (2019) 'Classification of complex power quality disturbances using optimized s-transform and kernel SVM', *IEEE Transactions on Industrial Electronics*, Vol. 67, No. 11, pp.9715–9723.

Wang, S. and Chen, H. (2019) 'A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network', *Applied Energy*, Vol. 235, pp.1126–1140 [online] https://doi.org/10.1016/j.apenergy.2018.09.160.

Zhang, Q-M. and Liu, H-J. (2008) 'Application of LS-SVM in classification of power quality disturbances', *Proceedings-Chinese Society of Electrical Engineering*, Vol. 28, No. 1, pp.106.

Zhao, F. and Yang, R. (2006) 'Power quality disturbance recognition using s-transform', *2006 IEEE Power Engineering Society General Meeting*, IEEE, p.7.

Zhou, N., Wang, J. and Wang, Q. (2016) 'A novel estimation method of metering errors of electric energy based on membership cloud and dynamic time warping', *IEEE Transactions on Smart Grid*, Vol. 8, No. 3, pp.1318–1329.

Zhu, T., Tso, S. and Lo, K. (2004) 'Wavelet-based fuzzy reasoning approach to power-quality disturbance recognition', *IEEE Transactions on Power Delivery*, Vol. 19, No. 4, pp.1928–1935.