JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 40, 1031-1043 (2024)
DOI: 10.6688/JISE.202409_40(5).0007

Hybrid Model of Multi-Resolution Signal Transformation and Deep Neural Network in Power Quality Disturbances Classification

DAR HUNG CHIAM+, KING HANN LIM AND KAH HAW LAW[1]
Department of Electrical and Computer Engineering
Curtin University Malaysia
CDT 250, 98009 Miri, Malaysia
E-mail: chiamdh@postgrad.curtin.edu.my+; glkhann@curtin.edu.my
[1] *Electrical and Electronic Engineering Programme Area*
Universiti Teknologi Brunei
BE1410 Gadong, Brunei Darussalam
E-mail: kahhaw.law@utb.edu.bn

Industrial end-users always demand for better power quality to retain grid's efficiency and maintain machinery health. However, increasing complexity of multiple energy systems increase the risk of deterioration in the quality of power supplies. An effective power quality disturbances detection tool is highly needed to automatically detect the unusual event exhibiting inside the systems. Power quality disturbances can be categorised into two category, *i.e.* (a) short and fast transient disturbances and (b) long and slow disturbances, which complicates the process of classification using similar model. In this paper, a hybrid deep learning model consisting of multi-resolution transformation and deep neural network is proposed for power quality disturbances detection. Transformer network has been proposed to improve the classification and computation performance with its multi-head attention mechanism and parallel computing characteristics. The proposed hybrid model first transforms input signal into multiple frequency components using multi-level signal decomposition signal via wavelet transform. A layer of convolutional kernel is used to obtain the spatial and temporal features from the wavelet components. The process is followed by higher order latent feature extraction using transformer network which includes a layer of transformer encoder and a pooling mechanism. The proposed model is able to outperform other deep neural network models with better accuracy despite of the noisy condition.

Keywords: attention mechanism, convolutional transformer, multi-resolution signal transformation, power quality disturbance classification, deep neural networks

1. INTRODUCTION

The widely use of renewable energy generators and advanced electronics controllers [1, 2] in modern power grid have greatly improved the energy sharing and increased the complexity of network transmission. To further reduce network complexity, microgrid is formed to perform local power quality monitoring and control on multiple power sources

Received December 31, 2022; revised April 15, 2023; accepted May 22, 2023.
Communicated by Wei Kitt Wong.
+ Corresponding author.

and loads. However, the monitoring of power quality disturbances (PQDs) becomes tedious when there are multiple intermittent DC/AC resources such as renewable energy sources and plugged-in hybrid electric vehicles connected in the microgrid [3]. The occurrence of single or multiple disturbance at one short instance is not easily observed using human intervention. Some common examples of PQDs [4, 5] includes sag, swell, interrupt, harmonics, transients, spike, notch, oscillatory transient, flicker, combined Sag with harmonics, flicker with harmonics, and flicker with swell. Therefore, a real-time PQD identification and classification system is required in guiding the decision-making of the power management systems for instant mitigation of any PQ issues arises .

Classification of PQDs can generally be divided into three stages [6], *i.e.* signal processing, feature extraction, and classification. Signals are first segmented into specific signal cycles. As time-domain features are generally insufficient in providing discriminative features, signal transformations into frequency-domain features [7, 8] are applied to increase signal resolution. Wavelet transform has been proven to have better performance for hybrid network by decomposing original signals into time-frequency domain [9]. Feature extraction process is subsequently performed to extract the discriminate information from the transformed signals. Statistical feature extraction is normally performed on the transformed signals [10]. The extracted features are then being passed into a classifier for classification process. Commonly used classifiers include decision tree [11], extreme learning machine [12], and artificial neural networks [13]. On the other hand, optimal feature selection techniques are applied to select prominent features for the classification process [10]. Deep neural networks (DNN) automatic feature extraction methods are introduced to reduce the network complexity [5]. DNN methods do not require the computation of statistical features and simplify the process of optimal feature selection. DNN methods can extract feature automatically through the end-to-end learning process of neural networks from data augmentation [14].

In this paper, a hybrid model of multi-resolution signal transformation incorporated with a DNN method, convolutional Transformer network (ConvT) is proposed to tackle multiple combined PQDs with time-frequency information. Multi-resolution wavelet transform is used to transform time domain signal into time-frequency wavelet domain. Five coefficients representing five frequency components of the signals are encoded into latent feature via a dense layer and stacked into a matrix with five dimensions. Global and local spatial relationship of the latent feature is then extracted using 1D-convolutional layer. Transformer encoder with the aid of multi-head attention (MHA) mechanism is then being used to further encode the extracted feature. A pooling mechanism using MHA is used to relates the relationship between the features into classes of PQDs. Accuracy of the classifiers are used to measure the performance of the network.

2. WAVELET DOMAIN DNN CLASSIFIER

Signal transformations are used to transform signals into different domain for better analysis. The use of wavelet transform shows marvelous performance in real-time PQD classification [4, 15]. In this paper, a hybrid model composing of multi-level wavelet decomposition with a layer of convolutional kernel and transformer mechanism is proposed as shown in Fig. 1. Details of the mechanism are explained in the following subsections.

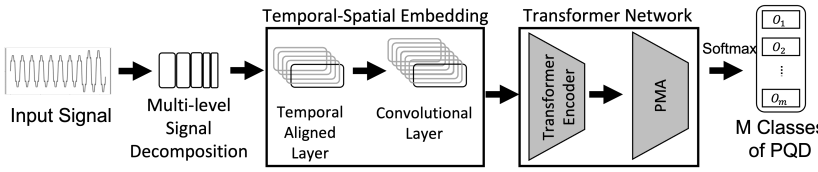

Fig. 1. Proposed hybrid model incorporating of multi-resolution signal transformation and convolutional transformer. The proposed model is named as WT-ConvT.

2.1 Multi-Resolution Signal Transformation via Wavelet Transform

Time-frequency domain conversion is carried out using wavelet transform (WT). Daubechies 4 (DB4) is chosen as mother wavelet as it can better recognize the short and fast transient [16]. The use of multi-level signal decomposition (MSD) at higher levels allows the detection of slow transients using the same mother wavelet [16]. Four level of decomposition is proposed which results in five decomposition coefficients, *i.e.* four detail coefficients and one approximate coefficients. Each decomposition coefficients represents different frequencies, *i.e.* global features ,and each frequency bands contain time domain components, *i.e.* local features.

WT decomposes signals into local representations in time-frequency domain. A discrete wavelet transform (DWT) can be achieved by dilating the translating mother wavelet discretely over the input signal. A generic mother wavelet can be expressed as,

$$\psi_{a,b}(t) = |a|^{-1/2}\psi(\frac{t-b}{a}). \tag{1}$$

A discretized mother wavelet can be achieved by setting $a = a_0^m$ and $b = nb_0a_0^m$, where $a_0 > 1$ and $b_0 > 0$ are constants, m and n are positive integers which represents the scaling and shifting parameters. The discretized mother wavelet $\psi_{m,n}(t)$ is defined as,

$$\psi_{m,n}(t) = a_0^{-m/2}\psi(\frac{t-nb_0a_0^m}{a_0^m}). \tag{2}$$

A dyadic-orthonormal wavelet transform can be obtained by selecting proper constant pair a_0 and b_0. An example would be $a_0 = 2$ and $b_0 = 1$. Dyadic orthonormal WT does not have redundant information among the decomposed signals, and allows signal decomposition into multiple scales or bands of time-frequency representations.

Daubechies' with four filter coefficients or DB4 mother wavelet is used for the decomposition as DB4 is good at detecting short and fast transient [16]. Higher decomposition level or MSD, *i.e.*, four decomposition level is used which allows the detection of slow transient disturbances. A multi-level wavelet transform decomposes signals into approximate coefficients and detailed coefficients. Approximate coefficients represents the smoothed signal, while detailed coefficients representing the detailed signal containing sharper edges or sharper magnitude transition. As a result, fast and short transient disturbances will be detected as lower level decomposition, *i.e.*, level 1; whereas long and slow transient disturbances will be detected at level 4 decomposition. Subsequently, MSD allows the use of single type of mother wavelet while covering different types of

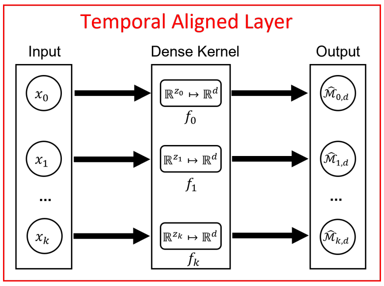

Fig. 2. Temporal aligned layer.

transients. The output of MSD can be noted as $\mathscr{M}$ and the I^{th} level can then be mapped as,

$$\mathscr{M} = [cA_I, cD_I, cD_{I-1}, ..., cD_1], \tag{3}$$

where $\mathscr{M} \in \mathbb{R}^Z$, $Z = [z_k]_{k=0}^K$, z_k is the dimension of individual vector in $\mathscr{M}$ and $K = I + 1$.

2.2 Temporal Aligned Layer

Temporal aligned layer is introduced to align the different length output coefficients of MSD. Functional block diagram of the temporal aligned layer is shown in Fig. 2. Temporal aligned layer is a collection of multiple single perceptron layers noted as $[f_k]_{k=0}^K : \mathbb{R}^Z \mapsto \mathbb{R}^{K\times d}$, where the single perceptron layer is a mapping function $f_k : \mathbb{R}^{z_k} \mapsto \mathbb{R}^d$, and K represent the total number of output coefficients from MSD such that $K = I + 1$. In this experiment $I = 4$-level of MSD is used, the total number of coefficients output $K = 4+1 = 5$. The temporal aligned MSD output, $\hat{\mathscr{M}}$ is obtained as follows,

$$\hat{\mathscr{M}} = [f_k(\mathscr{M}_k)]_{k=0}^K, \tag{4}$$

where $\hat{\mathscr{M}} \in \mathbb{R}^{K\times d}$ and fixed embedding output $d = 256$ is used.

2.3 Temporal-Spatial Embedding Layer

Temporal-Spatial Embedding Layer is designed in a mechanism of extracting temporal-spatial information using temporal aligned layer and a layer of 1D convolutional kernel. Temporal aligned layer is used to align the unequal length data retrieved from MSD output. The convolutional kernel used in this experiment includes kernel size = 5, stride = 1, input dimension = 5, and output dimension = 8. The small kernel size of 5 allows capturing of fine detail on the signals, while stride step of 1 ensures highest resolution of information being captured for the feature extraction step. The convolutional kernel used extracts temporal-spatial features from the aligned MSD features. The output feature matrix representing temporal-spatial features of the MSD are then being fed into transformer network for classification using multi-head attention on its salient features.

2.4 Transformer Encoder

Transformer network is introduced by Vaswani *et al*. [17] to allow parallel computation for time-dependent data. Transformer network follows encoder-decoder framework

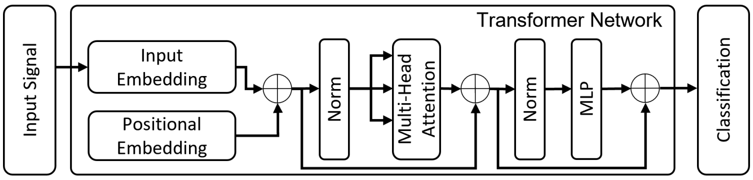

Fig. 3. Transformer encoder block.

[18] with the unique attention mechanism namely as multi-head attention (MHA). The input features are first normalised, then being projects into three three vectors, query Q, key K, and value V. A single head attention mechanism can be expressed as,

$$Attn(Q,K,V) = Softmax(\frac{A}{\sqrt{d_k}})V, \tag{5}$$

where $\frac{1}{\sqrt{dk}}$ is the scaling function to prevent gradient vanishing and exploding problem. $A = QK^T$ is the dot product between Q and K. K^T represents the transpose of K. Softmax function is the exponent of every single element of the input, y_i normalised with the sum of Z exponent element and can be depicted as follows,

$$Softmax(y_i) = \frac{e^{y_i}}{\sum_{j=0}^{Z} e^{y_j}}. \tag{6}$$

MHA can be achieved by multiple sets of trainable matrices, $W_q^{(i)}$, $W_k^{(i)}$, $W_v^{(i)}$ as [18],

$$Q^{(i)}, K^{(i)}, V^{(i)} = IW_q^{(i)}, IW_k^{(i)}, IW_v^{(i)}, \tag{7}$$

$$Head^{(i)} = Attn(Q^{(i)}, K^{(i)}, V^{(i)}), \tag{8}$$

$$MultiHead(Q,K,V) = Concat(Head^{(1)}, ..., Head^{(n)})W_O, \tag{9}$$

where i indexes the number of heads n and W_O represents the trainable parameter with size $\mathbb{R}^{d_k \times d}$. In Fig. 3, transformer encoder takes in input signals and embed it with positional embedding. A data normalisation is done before passing into the MHA. Output data from the MHA is normalised and processed before outputting as encoded latent feature.

2.5 Pooling Multi-head Attention

Pooling multi-head attention (PMA) mechanism [19] is introduced as alternative for averaging or max-pooling in aggregating feature vectors. MHA is applied on a learnable set of k seed vectors, $S \in \mathbb{R}^{k \times d}$. Let transformer encoder output be $Z \in \mathbb{R}^{n \times d}$, k seed vectors PMA can be represented as follows,

$$y_k(Z) = MultiHead(S,Z,Z). \tag{10}$$

The output of PMA is the predicted result of each class based on their time-series pattern. $Softmax(y_k)$ is implemented to confine the M classes into probability.

3. RESULT AND DISCUSSION

All experiments are conducted using Pytorch. There are 76.8k 10-periods PQD samples with 3.2 kHz sampling rate randomly generated based on mathematical modelling equations [8, 20]. A total of 4800 10-periods PQD samples are generated for each disturbance classes as shown in Table 1. Random generated additive white Gaussian noise (AWGN) with signal-to-noise ration (SNR) of 20-50dB and noiseless condition are randomly added into the generated data to further improve the generalisation of the models towards random noise conditions. The samples are then partitioned into 90% training samples and 10% validation data. Additional five sets of 16k samples comprising of 20dB, 30dB, 40dB, 50dB, and noiseless signals are generated for the testing process. These 16k samples consist of 1000 samples per disturbance class. All models in this experiment are trained with 200 epochs under Pytorch environment.

This research consists of three experiments. Experiment #1 studies on the classification performance of transformer encoder network as compared to Deep LSTM network. The transformer network is optimized using temporal-spatial embedding layer in Experiment #2. Finally, The computation complexity of the proposed models are compared and analysed with several literature models in Experiment #3. Average classification accuracy is used as the main evaluation matrix as follows,

$$Acc_n = \frac{TP_n}{\sum_{j=0}^{m} S_j}, \tag{11}$$

where the classification accuracy for each disturbance class, Acc_n can be attained from the true positive, TP_n divided by the total number of samples for m classes of PQD, S_j.

3.1 Experiment #1: Transformer Encoder Network

Transformer encoder network is proposed to substitute the use of Long Short Term Model (LSTM) [?] for parallel computation in time-dependent series. Instead of using Deep LSTM, the input signal is fed into two layers of transformer encoder network for automatic sequential feature extraction. The output from transformer encoder network is used for classification via a dense layer with softmax activation function. Transformer encoder network achieve better overall classification performance when compared to Deep LSTM model as shown in Table 2. Transformer encoder network gives advantage especially in classifying fast disturbance class P8-Notch under high noise condition. The

Table 1. 16 classes PQDs generated using mathematical equations [5, 20].

Label	Class Description	Label	Class Description
P0	Normal	P8	Notch
P1	Sag	P9	Flicker
P2	Swell	P10	Sag+Harmonics
P3	Interrupt	P11	Swell+Harmonics
P4	Impulse Transient	P12	Interrupt+Harmonics
P5	Spike	P13	Flicker+Harmonics
P6	Harmonics	P14	Flicker+Sag
P7	Oscillatory Transient	P15	Flicker+Swell

classification performance of transformer encoder maintains at 100% on P8 when 20dB AWGN is introduced. However, it can be noticed that Deep LSTM is severely impacted with performance drop to 66.67%. This shows that Deep LSTM model is less effective in extracting distinct features of fast disturbance under high noise condition. Besides that, It can be noticed that both networks perform worse in class P10-Sag+Harmonics on high noise 20dB SNR test. The classification performance of Deep LSTM network on class P10 is notably weak regardless of the level of noise introduced. Confusion matrix shows that the P10 in transformer encoder network is confused with class P9-Flicker. This can be explained with high level of AWGN distorted the magnitude of the input signal with flicker effect. Deep LSTM on the other hand shows mutual confusion between P10 with P0-Normal. The confusion of disturbance class with normal signal also presents for class P11 and P12. These confusions are more critical as it classify disturbance signals as normal signal. This also shows that LSTM model struggle to identify the characteristics of combined disturbance.

3.2 Experiment #2: Hybrid Multi-Resolution Convolutional Transformer Versus Deep CNN [6]

The use of multi-level signal decomposition using wavelet transform is introduced in this experiment to increase the resolution of input signal with wavelet domain representations. The input PQD signal is transformed into five wavelet coefficients using 4-level signal decomposition wavelet transform. Temporal-spatial embedding layer is used to extract spatial and temporal relationship between frequency bands. The five wavelet coefficients with different length are then pass through individual dense layers to produce five equal-length latent encoding, which are then stacked vertically to form a 5×256 feature matrix. An 1D convolutional kernel is used to extract spatial and temporal features

Table 2. Performance comparison of Deep LSTM with the transformer encoder model tested with 20-50dB AWGN and noiseless conditions.

(a) Deep LSTM

Class \ SNR	20dB	30dB	40dB	50dB	noiseless
P0	85.40	99.80	99.60	99.70	99.63
P1	93.30	97.00	96.60	96.80	96.80
P2	97.20	97.10	98.40	98.60	97.90
P3	98.20	99.70	99.90	99.90	100.0
P4	99.90	100.0	100.0	100.0	100.0
P5	98.90	99.50	99.30	99.30	99.90
P6	98.60	99.60	99.30	99.70	99.50
P7	97.80	99.80	99.10	99.50	99.80
P8	66.70	95.90	100.0	100.0	100.0
P9	94.50	99.20	99.90	99.90	99.70
P10	56.50	74.80	85.80	88.40	90.53
P11	83.30	96.30	98.00	97.00	99.49
P12	74.00	93.00	95.80	96.50	98.87
P13	89.50	100.0	100.0	100.0	100.0
P14	94.10	97.50	96.60	97.00	96.20
P15	87.80	97.00	96.90	98.20	98.10
Acc	88.48	96.64	97.83	98.16	98.54

(b) Transformer Encoder

Class \ SNR	20dB	30dB	40dB	50dB	noiseless
P0	71.30	89.50	94.90	97.90	98.03
P1	93.70	96.20	96.40	94.70	95.60
P2	95.90	98.00	98.50	98.60	98.90
P3	99.00	99.50	99.30	99.10	99.70
P4	99.90	100.0	100.0	100.0	100.0
P5	96.60	97.20	97.00	96.20	97.90
P6	99.90	100.0	100.0	100.0	100.0
P7	98.50	99.30	98.60	98.70	98.60
P8	100.0	100.0	100.0	100.0	100.0
P9	83.00	92.70	95.40	94.20	94.30
P10	62.40	93.40	96.40	97.40	99.08
P11	80.20	96.30	98.10	97.20	100.0
P12	83.80	95.50	96.80	97.40	100.0
P13	100.0	100.0	100.0	100.0	100.0
P14	98.90	98.90	99.50	99.30	98.70
P15	99.80	99.40	99.30	99.30	99.20
Acc	91.43	97.24	98.14	98.13	98.74

from the feature matrix. The settings of 1D convolutional kernel are kernel size = 5, stride = 1, input dimension = 5, output dimension = 8. The output features are then passed into a layer of transformer encoder which highlights the important information through MHA mechanism. Finally, a pooling mechanism using MHA is applied to aggregate the features before passing through softmax activation function for classification.

The proposed WT-ConvT is compared with deep CNN [5] to demonstrate the strength of the model in classifying PQDs under 20-50dB SNR AWGN and noiseless conditions. From Table 3 (b), it is noticed that WT-ConvT attained higher classification accuracy of 94.03% under 20dB SNR AWGN test. Overall, deep CNN model can only acquire classification accuracy of 90.56%. This shows that WT-ConvT is able to classify signals despite of higher noise condition. However, it can be noticed in Table 3b that WT-ConvT perform slightly poor for P0-Normal under high noise condition of 20dB and 30dB SNR AWGN. In Fig. 4 (b), it can be noticed that WT-ConvT has confusion on P0 with P10-Sag+harmonics. This is due to the additive noise characteristics has amplified the normal signal into higher noise level after signal normalization.

On the other hand, it can be noticed that Deep CNN obtained lower accuracy on class P8-Notch on high noise 20dB AWGN test. From Fig. 4 (a), it can be noticed that the confusions are between P8 with P14-Flicker+Sag and P15-Flicker+Swell. As for analysis, notch is categorised as a fast transient disturbance, while flicker with sag or swell are from slow disturbance, and can only be detected from multiple periods of signals. This confusion shows that deep CNN performance can be seriously impacted with higher noise levels, where the confusion happened across different frequencies signals. WT-ConvT model uses MSD to decompose input signal into multiple frequency components. It allows lower and higher frequency components to be separated from the time-series data. As a result, Fig. 4 (b) shows that WT-ConvT model suffers less impact from P8-Notch due to wavelet component could help in splitting the low and high frequency noise

Table 3. Performance comparison of Deep CNN [5] with the WT-ConvT tested with 20-50dB AWGN and noiseless conditions.

(a) Deep CNN

Class \ SNR	20dB	30dB	40dB	50dB	noiseless
P0	86.00	96.30	99.30	100.0	100.0
P1	93.80	98.30	98.40	97.80	98.10
P2	97.80	97.60	98.70	98.90	98.80
P3	100.0	99.80	99.80	100.0	100.0
P4	100.0	100.0	100.0	100.0	100.0
P5	95.80	98.70	98.70	99.10	99.20
P6	99.40	99.90	100.0	100.0	100.0
P7	100.0	100.0	100.0	99.90	100.0
P8	39.80	89.10	100.0	100.0	100.0
P9	95.50	100.0	100.0	100.0	100.0
P10	82.10	91.80	96.40	97.00	99.59
P11	88.70	97.50	98.20	97.20	100.0
P12	91.00	96.40	96.50	96.60	99.59
P13	88.50	100.0	100.0	100.0	100.0
P14	92.90	98.40	99.20	98.40	98.90
P15	97.70	99.20	98.30	99.20	98.90
Acc	90.56	97.69	98.97	99.01	99.57

(b) WT-ConvT

Class \ SNR	20dB	30dB	40dB	50dB	noiseless
P0	73.40	82.10	92.40	97.80	98.97
P1	93.00	96.60	97.10	96.40	97.10
P2	94.90	96.50	98.60	98.50	98.40
P3	99.30	98.90	99.30	99.20	99.00
P4	100.0	100.0	100.0	100.0	100.0
P5	97.70	98.50	97.70	98.20	98.30
P6	99.40	99.80	99.70	100.0	99.90
P7	98.80	99.50	99.40	99.40	99.40
P8	96.20	100.0	100.0	100.0	100.0
P9	85.50	98.50	99.90	99.90	99.90
P10	87.30	95.00	95.60	93.60	95.11
P11	92.90	97.60	98.20	97.20	100.0
P12	95.90	97.00	96.80	97.40	100.0
P13	99.10	100.0	100.0	100.0	100.0
P14	96.90	98.00	98.00	98.00	97.60
P15	94.20	96.20	95.00	97.40	95.00
Acc	94.03	97.14	97.98	98.31	98.67

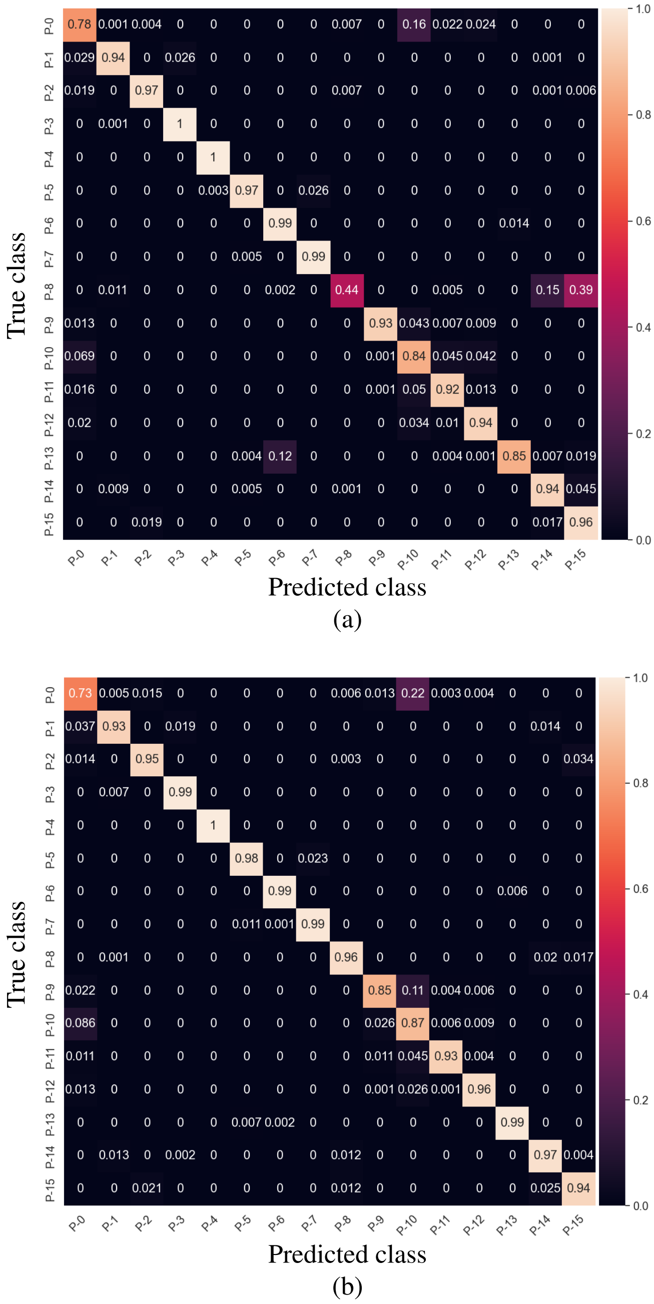

Fig. 4. Detail confusion matrix comparison on 20dB SNR AWGN test for (a) Deep CNN and (b) WT-ConvT.

to improve the generalization of feature extraction.

3.3 Experiment #3: Comparative Analysis with DNN Methods

The complexity of the proposed transformer encoder network and WT-ConvT are compared with two literature DNNs methods as shown in Table 4. The comparison parameters a the best 20dB accuracy, time taken per epoch (TTPE) of the training on 76.8k samples, number of parameters and models size. As a result, the proposed WT-ConvT achieved the highest classification performance of 94.03%. The proposed model can achieve better performance with the slight increase in the number of parameters or model size while maintaining the TTPE similar to deep CNN method. Although Deep LSTM model requires least parameters, it requires double TTPE to train the network as compared to WT-ConvT. Deep LSTM also perform worse under high noise condition due to less distinct features extracted for combined disturbance classes. Transformer encoder network has the least TTPE, however, it requires more computation resources.

Table 4. Model accuracy and complexity comparisons.

Model	20dB	30dB	40dB	50dB	noiseless	TTPE	Parameters	Model Size
Deep LSTM [22]	88.48	96.64	97.83	98.16	98.54	68s	33936	0.136 MB
Deep CNN [5]	90.56	97.69	98.97	99.01	99.57	32s	164464	0.657 MB
Transformer	91.43	97.24	98.14	98.13	98.74	22s	835344	3.269 MB
WT-ConvT	94.03	97.14	97.98	98.31	98.67	32s	182928	0.725 MB

4. CONCLUSION

A hybrid model of multi-resolution signal transformation incorporating with a convolutional transformer network is proposed to retrieve generalized features in PQDs under noisy condition. First, wavelet coefficients are decomposed from time-series signal and subsequently converted into latent feature using a perceptron layer. Spatial and temporal relationship between multiple frequency components are extracted using single layer of multi-dimension 1D convolutional kernel. These features are then passed to transformer encoder for higher-order latent features aggregation using MHA pooling mechanism. Simulation experiments shows that the use of transformer encoder network has better distinct feature extraction compared to Deep LSTM network especially under high noise condition. Transformer encoder network achieves better classification accuracy of 91.43% under 20dB SNR AWGN test, which is higher when compared to Deep LSTM network with 88.48%. Besides that, transformer network allows fast parallel processing and significantly reduces the training time. The drawback of transformer encoder network is its high requirement on computation resources. WT-ConvT is thus proposed with reduced parameters and better classification performance of 94.03% on 20dB SNR AWGN condition as compared to the benchmark Deep CNN and Deep LSTM method. The proposed WT-ConvT model is able to highlight the salient difference between fast transient and slow disturbance via multi-level wavelet coefficients. As for future work, an improvement should be focused on optimizing the model size and reducing the required training time. Improvements should also be given on the detection of high noise slow disturbance, in which the additive noise causes huge deterioration to the signal condition.

REFERENCES

1. D. Razmi and T. Lu, "A literature review of the control challenges of distributed energy resources based on microgrids (mgs): Past, present and future," *Energies*, Vol. 15, 2022, p. 4676.
2. R. Smolenski, P. Szczesniak, W. Drozdz, and L. Kasperski, "Advanced metering infrastructure and energy storage for location and mitigation of power quality disturbances in the utility grid with high penetration of renewables," *Renewable and Sustainable Energy Reviews*, Vol. 157, 2022, p. 111988.
3. T. S. Mahmoud, B. S. Ahmed, and M. Y. Hassan, "The role of intelligent generation control algorithms in optimizing battery energy storage systems size in microgrids: A case study from western australia," *Energy Conversion and Management*, Vol. 196, 2019, pp. 1335-1352.
4. S. Khokhar, A. A. M. Zin, A. P. Memon, and A. S. Mokhtar, "A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network," *Measurement*, Vol. 95, 2017, pp. 246-259.
5. S. Wang and H. Chen, "A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network," *Applied Energy*, Vol. 235, 2019, pp. 1126-1140.
6. D. H. Chiam, K. H. Lim, and K. H. Law, "Detection of power quality disturbances using wavelet-based convolutional transformer network," in *Proceedings of IEEE International Conference on Green Energy, Computing and Sustainable Technology*, 2022, pp. 1-6.
7. P. Khetarpal and M. M. Tripathi, "A critical and comprehensive review on power quality disturbance detection and classification," *Sustainable Computing: Informatics and Systems*, Vol. 28, 2020, p. 100417.
8. Q. Tang, W. Qiu, and Y. Zhou, "Classification of complex power quality disturbances using optimized s-transform and kernel svm," *IEEE Transactions on Industrial Electronics*, Vol. 67, 2019, pp. 9715-9723.
9. J. Ku and B. C. Kovoor, "A wavelet-based hybrid multi-step wind speed forecasting model using lstm and SVR," *Wind Engineering*, Vol. 45, 2021, pp. 1123-1144.
10. S. Chamchuen, A. Siritaratiwat, P. Fuangfoo, P. Suthisopapan, and P. Khunkitti, "High-accuracy power quality disturbance classification using the adaptive ABC-PSO as optimal feature selection algorithm," *Energies*, Vol. 14, 2021, p. 1238.
11. T. Zhong, S. Zhang, G. Cai, Y. Li, B. Yang, and Y. Chen, "Power quality disturbance recognition based on multiresolution s-transform and decision tree," *IEEE Access*, Vol. 7, 2019, pp. 88 380-88 392.
12. I. S. Samanta, P. K. Rout, and S. Mishra, "Feature extraction and power quality event classification using curvelet transform and optimized extreme learning machine," *Electrical Engineering*, Vol. 103, 2021, pp. 2431-2446.

13. S. M. Qaisar, "Signal-piloted processing and machine learning based efficient power quality disturbances recognition," *PloS One*, Vol. 16, 2021, p. e0252104.
14. Y. Deng, H. Jia, P. Li, X. Tong, and F. Li, "A deep learning method based on long short term memory and sliding time window for type recognition and time location of power quality disturbance," in *Proceedings of IEEE Chinese Automation Congress*, 2018, pp. 1764-1768.
15. B. Panigrahi and V. R. Pandi, "Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbour algorithm," *IET Generation, Transmission & Distribution*, Vol. 3, 2009, pp. 296-306.
16. S. Santoso, E. J. Powers, W. M. Grady, and P. Hofmann, "Power quality assessment via wavelet transform analysis," *IEEE Transactions on Power Delivery*, Vol. 11, 1996, pp. 924-930.
17. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, Vol. 30, 2017, pp. 6000-6010.
18. H. Yan, B. Deng, X. Li, and X. Qiu, "Tener: adapting transformer encoder for named entity recognition," *arXiv Preprint*, 2019, arXiv:1911.04474.
19. J. Lee, Y. Lee, J. Kim, A. Kosiorek, S. Choi, and Y. W. Teh, "Set transformer: A framework for attention-based permutation-invariant neural networks," in *Proceedings of International Conference on Machine Learning*, 2019, pp. 3744-3753.
20. H. Liu, F. Hussain, S. Yue, O. Yildirim, and S. J. Yawar, "Classification of multiple power quality events via compressed deep learning," *International Transactions on Electrical Energy Systems*, Vol. 29, 2019, p. e12010.
21. D. H. Chiam, K. H. Lim, and K. H. Law, "LSTM power quality disturbance classification with wavelets and attention mechanism," *Electrical Engineering*, Vol. 105, 2022, pp. 259-266.
22. R. Machlev, A. Chachkes, J. Belikov, Y. Beck, and Y. Levron, "Open source dataset generator for power quality disturbances with deep-learning reference classifiers," *Electric Power Systems Research*, Vol. 195, 2021, p. 107152.

Dar Hung Chiam is currently pursuing his Ph.D. in Electrical and Computer Engineering at Curtin University Malaysia. He received his Bachelor of Engineering (B.Eng) degree in Electrical and Electronic Engineering from Curtin University, Malaysia in 2018. His research interests include power quality analysis, machine learning and artificial intelligence.

King Hann Lim received the M.Eng. and Ph.D. degrees in Electrical and Electronic Engineering from the University of Nottingham, in 2007 and 2012, respectively. He is currently a Professor with the Department of Electrical and Computer Engineering in Curtin University, Malaysia. His research areas include computer vision, artificial intelligence, and optimization. He has published more than 100 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 Universiti Teknologi Brunei (UTB), Brunei Darussalam. He received his Ph.D. degree in Electrical and Electronic Engineering from 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.