TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

Deep Wavelet-based Convolutional Transformer Network in Power Quality Disturbances Classification

Dar Hung Chiam*, King Hann Lim, Jonathan Then Sien Phang, Basil Andy Lease
Department of Electrical and Computer Engineering
Curtin University Malaysia
CDT 250, 98009 Miri, Malaysia.
chiamdh@curtin.edu.my*; glkhann@curtin.edu.my

***Abstract*—Real time power quality monitoring is important to ensure stable functioning of the electrical appliances especially for the manufacturing sector. Deep-WT-ConvT is proposed to to better characterise and differentiate the minor differences between different types of power quality disturbances. However, the use of deep networks requires longer training time, and poses the risk of getting internal covariant shift issues due to distribution change in layer's input during training phase. This issue can be prevented by proper parameter initialisation and with lower learning rate, which slows down the training process. Batch normalisation (BN) layers are proposed to improve the classification performance of the PQD classifier network WT-ConvT. Results shows significant improvement on Deep-WT-ConvT model with accuracy improvement from 92.95% without BN layers to 94.44% with BN layers on 20dB SNR AWGN noise test.**

***Index Terms*—Power Quality Disturbances, Classification, Transformer network**

I. Introduction

Advanced in power generation technologies and increasingly sophisticated electrical appliances demands for stable quality of power supplies. Continuous power quality monitoring is required to alert on the abnormalities or power quality disturbances (PQD) occurred. Immediate and accurate mitigation actions must be carried out to protect the electrical appliances and minimize the downtime losses. Real-time classification of PQDs has thus became a critical challenge in this field of study [1]. The ability to identify and classify PQDs occurred allows locating the faults in terms of time, location, and narrowing the possible causes of the faults occurred.

Detection and classification of PQDs are performed manually by studying the historical power data. The advancement and evolution of the computing technologies improve the efficiency of the detecting mechanism with knowledge based methods. Statistical parameters such as RMS, minimum, maximum, average, and deviation are used as the statistical features for classification process. However, in order to extract useful statistical parameters, complex signal transformations are required. Extensive studies have been made on knowledge based methods in order to improve its performance in terms on classification accuracy and improving effectiveness of computing resources [2]–[10].

The introduction of artificial intelligence, especially in the field of machine learning and deep learning algorithms play a significant role in today's power system applications [11]. Model-based methods are proposed for automatic feature extraction and classification of PQD [12]–[14]. Different from knowledge-based method, feature extraction and classification process are performed in a closed-loop feedback system, which removes the need of manually selecting prominent statistical features to improve the classification performance [15]. Model-based methods with its automatic feature selections are further innovated with hybrid methods which combines signal transformation and neural networks [16]–[18].

The use of multiple layers of neural networks (or deep learning) allows automatic representations learning which produces discriminant features for the classification process [19]. However, the training of the network becomes a challenge as the designed network increases its depth. This is caused by the changes in distribution of the input of each layer during training. A lower learning rate and careful initialization of parameters is thus required. Batch normalization (BN) solves the saturating non-linearities or internal covariance shift issue by fixing the mean and variances of layer inputs [20].

In this paper, a deep wavelet-based convolutional transformer network (Deep-WT-ConvT) is proposed as depicted in Fig. 1. This model is an enhanced version of the wavelet based convolutional transformer network (WT-ConvT) [16]. The effects of adding BN layer are analysed on different depth of the network. The main evaluation technique used in this research are classification performance of the network tested with 20-50dB signal-to-noise (SNR) ratio of additive white Gaussian noise (AWGN).

II. Deep Wavelet-based Convolutional Transformer Network Enchanced with Batch Normalisation Layers

A deeper network architecture is proposed to analyse the needs of multiple layers of abstraction for the classification of PQD. In this study, a deeper temporal-aligned layer with 2 layers of perceptrons is used. The classification performance of the Deep-WT-ConvT and WT-ConvT [16] are compared in terms of classification performance. This study also covers the effect of using BN layers in both models. The components of the proposed model are described in the following subsections.

979-8-3503-0219-6/23/$31.00 ©2023 IEEE

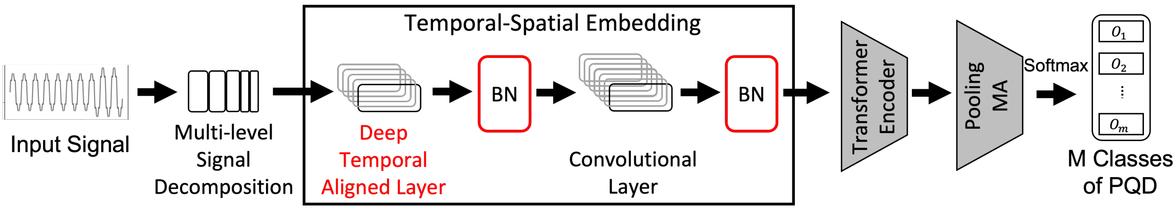

Fig. 1. Proposed deep wavelet-based convolutional transformer network with batch normalization layer (Deep-WT-ConvT).

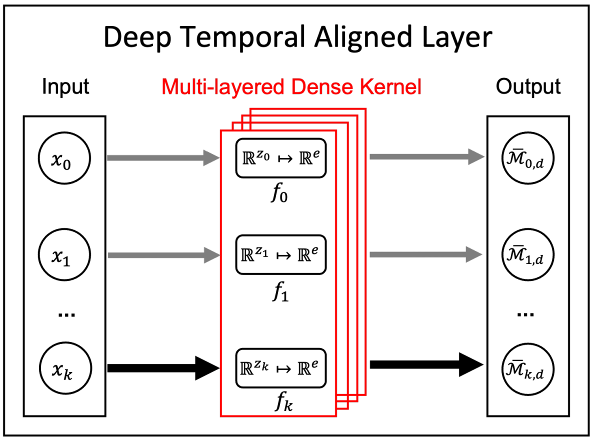

Fig. 2. Deep Temporal aligned layer.

A. Deep Wavelet-based Convolutional Transformer Network

A multi-layered temporal aligned layered is proposed in Deep-WT-ConvT to increase the depth of the WT-ConvT model as shown in Fig. 2. A multi-layered dense kernel is used, i.e. higher abstract level of representation, which allows learning of increased complexity functions [19].

B. Batch Normalization Layer

Deeper networks however can face issues such as over-fitting and internal covariance shift. BN layers are thus introduced to reduce the unwanted internal covariance shift, reduce vanishing or exploding gradients, and regularize the network for better generalization [20]. Two BN layers are proposed as shown in Fig. 1. BN with d-dimensional input, $x = (x^{(1)}, x^{(2)}, ..., x^{(d)})$ can be described as

$$\hat{x}_i^{(k)} = \frac{x_i^{(k)} - \mu_B^{(k)}}{\sqrt{(\sigma_B^{(k)})^2 + \epsilon}}, \tag{1}$$

where m is the total size of the training set, B represent the mini batch size. $k \in [1, d]$, $i \in [1, m]$, $\mu_B^{(k)}$ and $\sigma_B^{(k)}$ are mean and standard deviation respectively. A small constant, ϵ is added mainly for numerical stability. The output of BN layer is depicted as

$$y_i^{(k)} = \gamma^{(k)} \hat{x}_i^{(k)} + \beta^{(k)}, \tag{2}$$

where $\gamma^{(k)}$ and $\beta^{(k)}$ are trainable scaling and shifting parameters respectively.

III. Experiment Setup

In this research, all the models are trained using AMD Ryzen 7 3800X 8-Core Processor with Nvidia P6000 graphic processing unit. Pytorch framework has been used for the experiments. This experiment is carried out by comparing the proposed Deep-WT-ConvT network with WT-ConvT network [16]. The classification performance of the network is validated with 20-50dB SNR AWGN. Classification accuracy is used as the main evaluation matrix. 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,

$$Acc_n = \frac{TP_n}{\sum_{j=0}^{m} S_j}. \tag{3}$$

This experiment has been carried out on the dataset with 16 classes of 10-periods 3200Hz sampled PQDs as listed in Table I. A total of 76.8k samples have been generated using mathematical model [7]. The PQD samples generated include 16 classes of 10-periods PQD. The model used in this experiment is shown in Fig. 1. The experiment started with classification performance analysis on proposed Deep-WT-ConvT, and followed by analysis on BN layers.

IV. Classification Performance Analysis on Deep-WT-ConvT

Deep temporal aligned layer is introduced to increase the depth of abstraction to the feature representations. The process is done by having multiple layers of dense kernel during the feature conversion stage. A total of two layers of dense kernels are introduced in Deep-WT-ConvT model, and the results are recorded in Table II. It can be noticed that the classification performance of Deep-WT-ConvT is however degraded especially on high noise 20dB SNR AWGN test which drops from 94.03% on WT-ConvT [16] to 92.95%. The decrease in classification performance can be explained with increased parameters require more training steps for generalization.

The introduction of BN layers help regularize the network with normalised parameters which promotes feature learning for better generalization. The results of Deep-WT-ConvT model trained with added BN layers are tabulated in Table III. Result shows that both models improved across

TABLE I
Class of Power quality disturbances.

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

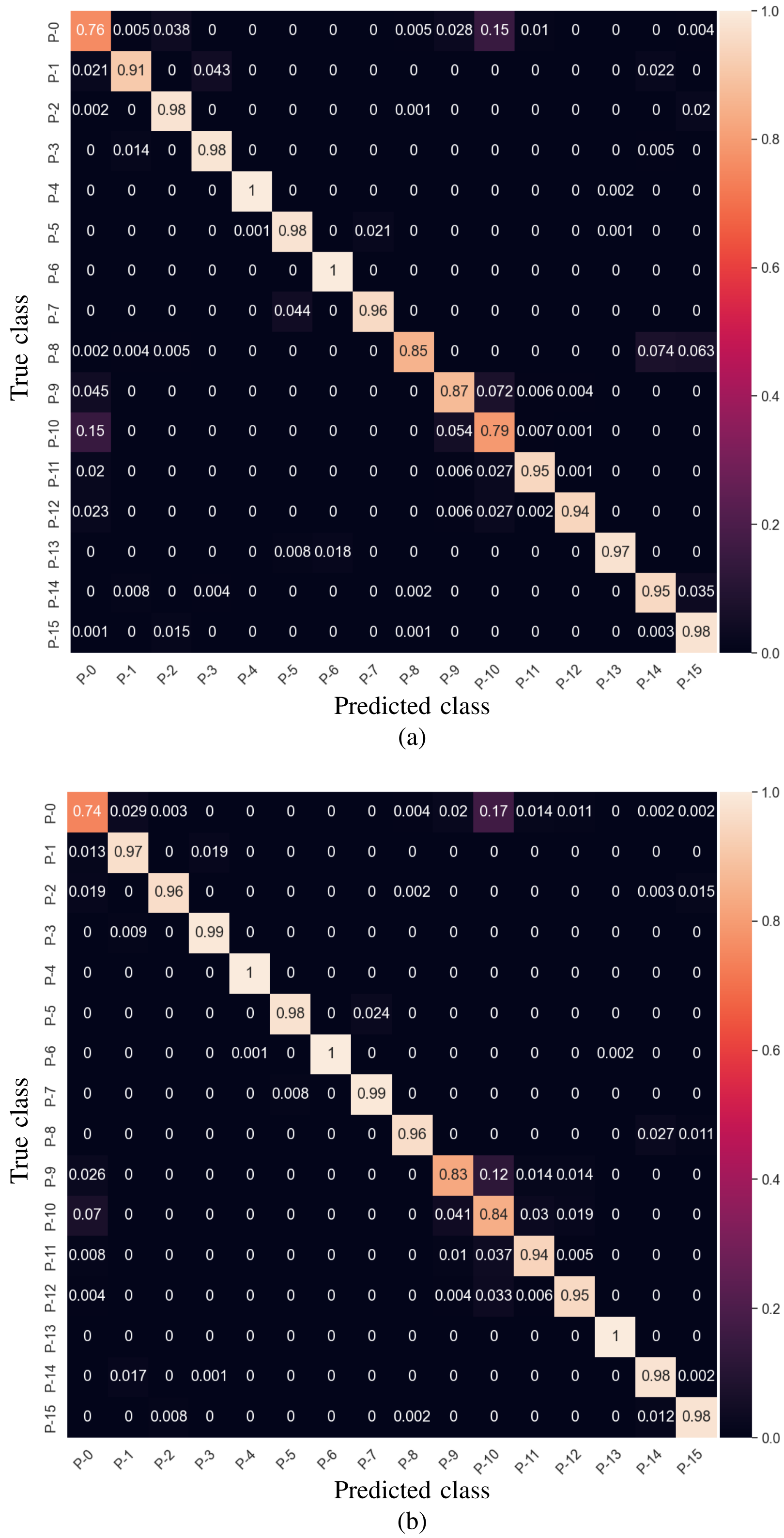

Fig. 3. Confusion matrix for Deep-WT-ConvT at 20dB SNR AWGN test (a) without BN layer and (b) with BN layer.

all noise levels when BN layers are used. The improvement is especially significant on 20dB SNR AWGN test with improvement from 92.95% to 94.42%. This result shows BN layers are important especially on multi-layered networks.

In Fig. 3, it is noticed that the decrease in classification performance in Deep-WT-ConvT model without BN layers is mainly caused by class P8-Notch, P9-Flicker, and P10-Sag+Harmonics. P8 is having 7.4% wrongly classified samples as P14-Flicker+Sag, and 6.3% classified as P15-Swell+Flicker. This involves confusion across different frequency domain, where class P8, a fast transient disturbance is having confusion with slow disturbance classes, class P14-

TABLE II
DEEP-WT-CONVT.

SNR / Class	20dB	30dB	40dB	50dB	noiseless
P0	76.00	91.50	98.10	98.80	97.56
P1	91.40	95.40	95.90	95.00	95.70
P2	97.70	99.20	99.70	99.50	99.70
P3	98.10	98.30	98.30	98.80	99.30
P4	99.80	100.0	100.0	100.0	100.0
P5	97.70	98.80	98.50	97.70	98.50
P6	100.0	100.0	100.0	100.0	100.0
P7	95.60	96.60	97.30	96.90	96.10
P8	85.20	100.0	100.0	100.0	100.0
P9	87.30	97.30	99.90	99.70	99.80
P10	79.10	88.60	90.40	91.30	93.69
P11	94.60	97.60	98.20	97.20	100.0
P12	94.20	97.00	96.80	97.40	100.0
P13	97.40	100.0	100.0	100.0	100.0
P14	95.10	98.70	99.00	98.50	98.60
P15	98.00	99.50	99.30	99.90	99.80
Acc	92.95	97.41	98.21	98.17	98.67

TABLE III
DEEP-WT-CONVT WITH BN LAYERS.

SNR / Class	20dB	30dB	40dB	50dB	noiseless
P0	74.40	89.90	98.00	98.20	98.59
P1	96.80	98.70	98.20	97.20	98.80
P2	96.10	98.50	99.00	98.80	99.30
P3	99.10	99.50	99.70	99.30	99.80
P4	100.0	100.0	100.0	100.0	100.0
P5	97.60	99.10	98.10	97.90	98.80
P6	99.70	99.90	100.00	99.90	99.90
P7	99.20	99.50	99.80	99.60	99.70
P8	96.20	100.0	100.0	100.0	100.0
P9	82.80	97.20	99.80	100.0	99.90
P10	84.00	90.70	90.50	88.90	91.14
P11	94.00	97.50	98.20	97.20	100.0
P12	95.30	97.00	96.80	97.40	100.0
P13	100.0	100.0	100.0	100.0	100.0
P14	98.00	98.60	98.60	98.70	98.00
P15	97.80	98.30	97.40	97.90	97.70
Acc	94.44	97.78	98.38	98.19	98.86

Sag+Flicker and P15-Swell+Flicker. On the other hand, it is noticed that Class P9 is having 7.2% confusion as P10, and P10 having 15% mutual confusion as class P0-Normal. These confusion are all based on slow disturbance domain. This can be caused by low magnitude disturbance which gives slight magnitude changes between disturbance classes. Added BN layers shows major improvement, where confusion on fast disturbance domain are removed. The main cause of the confusion are from class P9 and P10, were P9 is having 12% confusion as P10, and P10 having 7% confusion as P0, and P0 having 17% confusion as P10. These can be explained with low magnitude difference between these classes.

TABLE IV
MODEL ACCURACY AND COMPLEXITY COMPARISONS.

Model	20dB	30dB	40dB	50dB	Noiseless	Acc_{ave}
Deep LSTM [21]	88.48	96.64	97.83	98.16	98.54	95.93
Deep CNN [15]	90.56	97.69	98.97	99.01	99.57	97.16
WT-ConvT [16]	94.03	97.14	97.98	98.31	98.67	97.226
Deep-WT-ConvT	94.44	97.78	98.38	98.19	98.86	**97.53**

A comparison to literature studies is performed as shown in Table IV. Result shows our proposed Deep-WT-ConvT model achieved highest average classification accuracy of 97.53%. The proposed model also shows highest noise immunity with 94.44% accuracy on high noise 20dB SNR test condition.

V. CONCLUSION

Deeper network architectures are having advantages in extracting non-linear characteristics and thus allowing extraction of distinct features from the input signals. In this paper, Deep-WT-ConvT is proposed by using multiple layers of perceptron in the temporal-aligned layer. Deep temporal aligned layer is proposed as deeper network allows better feature discrimination and suppress less relevant features. However, the classification performance of Deep-WT-ConvT shows slight reduction compared to WT-ConvT, which is from 94.03% to 92.95% when tested under high noise 20dB SNR AWGN condition. This drop in classification performance may indicate the need of longer training period with the increased size of network. This issue has been resolved with the introduction of batch normalisation layers within the network. This modification results in the best model with highest average classification accuracy of 97.53%. The classification accuracy on 20dB SNR AWGN test also shows drastic improvement from 92.95% to 94.44%. The introduction of BN normalised the inputs and stabilised the distribution of activation values throughout training process. Besides from requiring longer training time, deeper networks also poses increased number of parameters. In the future work, a deeper network architecture , and a more efficient network connection can be studied.

REFERENCES

[1] J. E. Caicedo, D. Agudelo-Martínez, E. Rivas-Trujillo, and J. Meyer, “A systematic review of real-time detection and classification of power quality disturbances,” *Protection and Control of Modern Power Systems*, vol. 8, no. 1, pp. 1–37, 2023.

[2] M. Biswal and P. K. Dash, “Measurement and classification of simultaneous power signal patterns with an s-transform variant and fuzzy decision tree,” *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 1819–1827, 2012.

[3] S. De and S. Debnath, “Real-time cross-correlation-based technique for detection and classification of power quality disturbances,” *IET Generation, Transmission & Distribution*, vol. 12, no. 3, pp. 688–695, 2018.

[4] M. A. Rodriguez-Guerrero, A. Y. Jaen-Cuellar, R. D. Carranza-Lopez-Padilla, R. A. Osornio-Rios, G. Herrera-Ruiz, and R. d. J. Romero-Troncoso, “Hybrid approach based on ga and pso for parameter estimation of a full power quality disturbance parameterized model,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 1016–1028, 2017.

[5] X. Shi, H. Yang, Z. Xu, X. Zhang, and M. R. Farahani, “An independent component analysis classification for complex power quality disturbances with sparse auto encoder features,” *IEEE Access*, vol. 7, pp. 20 961–20 966, 2019.

[6] K. Thirumala, S. Pal, T. Jain, and A. C. Umarikar, “A classification method for multiple power quality disturbances using ewt based adaptive filtering and multiclass svm,” *Neurocomputing*, vol. 334, pp. 265–274, 2019.

[7] R. Mahla, B. Khan, O. P. Mahela, and A. Singh, “Recognition of complex and multiple power quality disturbances using wavelet packet-based fast kurtogram and ruled decision tree algorithm,” *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 12, no. 05, p. 2150032, 2021.

[8] S. Z. Motlagh and A. A. Foroud, “Power quality disturbances recognition using adaptive chirp mode pursuit and grasshopper optimized support vector machines,” *Measurement*, vol. 168, p. 108461, 2021.

[9] Y. Liu, T. Jin, M. A. Mohamed, and Q. Wang, “A novel three-step classification approach based on time-dependent spectral features for complex power quality disturbances,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.

[10] E. A. Nagata, D. D. Ferreira, C. A. Duque, and A. S. Cequeira, "Voltage sag and swell detection and segmentation based on independent component analysis," *Electric Power Systems Research*, vol. 155, pp. 274–280, 2018.

[11] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 209–220, 2020.

[12] I. Topaloglu, "Deep learning based a new approach for power quality disturbances classification in power transmission system," *Journal of Electrical Engineering & Technology*, vol. 18, no. 1, pp. 77–88, 2023.

[13] 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 *2018 Chinese Automation Congress (CAC)*. IEEE, 2018, pp. 1764–1768.

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

[15] 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, pp. 1126–1140, 2019.

[16] D. H. Chiam, K. H. Lim, and K. H. Law, "Detection of power quality disturbances using wavelet-based convolutional transformer network," in *2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*. IEEE, 2022, pp. 150–154.

[17] Y. Gao, Y. Li, Y. Zhu, C. Wu, and D. Gu, "Power quality disturbance classification under noisy conditions using adaptive wavelet threshold and dbn-elm hybrid model," *Electric Power Systems Research*, vol. 204, p. 107682, 2022.

[18] A. da Costa Pinho and E. G. A. Garcia, "Wavelet spectral analysis and attribute ranking applied to automatic classification of power quality disturbances," *Electric Power Systems Research*, vol. 206, p. 107827, 2022.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[21] 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, p. 107152, 2021.