


School of Engineering and Science
Department of Electrical and Computer Engineering

Power Quality Management and Classification for Smart
Grid Application using Machine Learning

Chiam Dar Hung

 0000-0001-8455-8658

This thesis is presented for the degree of
Doctor of Philosophy
of
Curtin University

June 2023

Declaration

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature:

Date:

Acknowledgements

First and foremost, I would like to express my sincere gratitude towards my main supervisor Prof. Garenth Lim King Hann for the continuous support and guidance in my Ph.D. study. Thank you for providing us with all the required hardware and research needs. Thank you for your patience in guiding me for more than seven years in my university life. The motivations and mental supports from you are priceless. Next, i would also like to thank my research committee, Dr. Law Kah Haw and A/Prof. Ling Huo Chong in supporting my research with all the useful comments and suggestions.

I would also like to express special gratitude towards my research companion, Dr. Jonathan Phang Then Sien. Thank you for your companionship throughout my studies. Thank you also for all the guidance in the field of machine learning. I couldn't imagine the research journey without your great support. Moreover, I would like to thank all the ECE department staffs in Curtin Malaysia. Thank you for all the support given during my study, especially when carrying out experiments in the lab. Thank you Curtin Malaysia for providing a safe and comfortable place for the research.

Last but not least, I would also like to thank my friends and family for supporting my Ph.D. journey. Thank you for your companionship throughout these years. I am truly thankful to my loving parents, Chiam Tow Jin and Pang Nyuk Ngo for providing me with such a healthy environment for my physical and mental growth, and with the opportunity to further my studies.

Abstract

The advancement of the smart grid is crucial in optimizing renewable energy resource utilization and power flow management with better power quality disturbance (PQD) detection and control. The classification of PQD detection can be divided into three types, i.e. knowledge-based method, model-based method and hybrid method. The knowledge-based method requires expert knowledge to manually extract PQD features using the mathematical form for classification. On the other hand, model-based methods apply a deep learning approach to automatically select features based on data representation to achieve better classification performance. The hybrid method integrates the advantage of knowledge-based signal processing tools in providing semi-processed signals, and combines with model-based methods for automatic feature extraction and selections.

A hybrid method using multi-level signal decomposition (MSD) with wavelet transform (WT) is proposed to improve the poor magnitude sensitivity on high-frequency signals. Our proposed model WT-SFA-LSTM combines MSD with a spatial attention mechanism to achieve a classification accuracy of 93.79%. This performance is better than the state-of-the-art Deep CNN model with 90.56% accuracy. However, further improvement is required to address the 32% larger model size of WT-SFA-LSTM compared to the Deep CNN model. On the other hand, a transformer with a multi-head attention mechanism, and faster computation via parallel processing on sequential input is introduced to replace LSTM. A compact model wavelet-based convolutional transformer (WT-ConvT) is proposed to address the issue of insensitivity to small-magnitude changes. Results show that WT-ConvT achieves a better classification accuracy of 94.11%. An efficient WT-ConvT (EWT-ConvT) is proposed to compress the model size up to 60% reduction via a weight superposition algorithm. EWT-ConvT achieves the

highest classification performance of 94.42% while having a 56% smaller model size of 0.29MB as compared to the state-of-the-art Deep CNN model.

The real-time hardware implementation of the PQD classifier is important, especially in terms of smart grid applications. A real-time PQD classifier is realised using Raspberry Pi 4B microcontroller. Real-time implementation of the PQD classifier is verified with a complete setup in a laboratory environment. Three case studies are performed to verify the system in detecting three classes of PQD. The PQD waveforms are captured and classified using the proposed embedded system. The proposed EWT-ConvT successfully classify a 200ms signal waveform within 75.51ms signifies real-time classification capability. As a result, EWT-ConvT can achieve the highest classification performance of 75% with the lowest computation resources. In conclusion, the proposed EWT-ConvT can perform real-time classification using the low-cost embedded system Raspberry Pi with high classification performance.

Publications

Parts of this thesis and concepts from it have been previously published in the following journal or conference papers.

Journal Papers

1. **Dar Hung Chiam**, King Hann Lim, "Global Attention-based LSTM for Noisy Power Quality Disturbance Classification" IJSCC V14 N1 2023, Inderscience Publishers (accepted). [Content as documented in Chapter 3]
2. **Dar Hung Chiam**, King Hann Lim, and Kah Haw Law. "LSTM power quality disturbance classification with wavelets and attention mechanism." *Electrical Engineering* (2022): 1-8. [Content as documented in Chapter 4]

Conference Papers

1. **Dar Hung Chiam**, King Hann Lim, and Kah Haw Law, "Multi-level Signal Decomposition for Power Quality Disturbance Classification." In *MATEC Web of Conferences*, 2023. (accepted) [Content as documented in Chapter 4]
2. **Dar Hung Chiam**, King Hann Lim, and Kah Haw Law, "Detection of Power Quality Disturbances Using Wavelet-Based Convolutional Transformer Network", 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST) (accepted) [Content as documented in Chapter 5]

3. **Dar Hung Chiam**, and King Hann Lim. "Power Quality Disturbance Classification Using Transformer Network." In International Conference on Cyber Warfare, Security and Space Research, pp. 272-282. Springer, Cham, 2022. [Content as documented in Chapter 5]
4. **Dar Hung Chiam**, King Hann Lim, and Kah Haw Law. "Noise level evaluation on power quality disturbances classification." In 2021 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), pp. 1-6. IEEE, 2021. [Content as documented in Chapter 3]
5. **Dar Hung Chiam**, King Hann Lim, and Kah Haw Law. "PQD Analysis in Smart Grid System Using Attention-Based Neural Network." In New Trends and Applications in Internet of Things (IoT) and Big Data Analytics, pp. 1-14. Springer, Cham, 2022. [Content as documented in Chapter 2 and Chapter 3]

Attribution Statements:

All authors provide equal contribution to the completion of the publication listed above.

Contents

Declaration	ii
Acknowledgements	iii
Abstract	iv
Publications	vii
List of Figures	xii
List of Tables	xiv
List of Acronyms	xvii
Chapter 1: Introduction	1
1.1 Project Overview	1
1.2 Problems Statement	3
1.3 Research Questions	5
1.4 Objectives	5
1.5 Significance and Contributions	6
1.6 Thesis Overview	8
Chapter 2: Literature Review on Real-Time Power Quality Distur-	
bance Classification	11
2.1 Introduction	11
2.2 Power Quality Disturbance Classification	12
2.2.1 Knowledge-based Method	15
2.2.2 Model-based Method	16
2.2.3 Hybrid Method	17

2.2.4	Summary of Power Quality Disturbance Classification	18
2.3	Implementation Studies in Real-time PQD Classification	20
2.3.1	High-Performance Computing	21
2.3.2	Embedded Systems	22
2.3.3	Summary of Implementation of Real-time PQD Classifier	23
2.4	Chapter Summary	23
 Chapter 3: Global Attention Mechanism with Long Short-Term Memory Network		25
3.1	Introduction	25
3.2	Global Attention-based Long Short-Term Memory Network	26
3.2.1	Global Attention Mechanism	27
3.2.2	Long Short-Term Memory	28
3.3	Experiment Setup	30
3.4	Experiment #1 - Single Disturbance Analysis	34
3.4.1	Analysis on Signal Windowing	34
3.4.2	Performance Comparison With Noisy Data	34
3.5	Experiment #2 - Combined-disturbance Analysis	39
3.5.1	Experiment #2 Noiseless Training Analysis	39
3.5.2	Noisy Training Analysis	41
3.6	Chapter Summary	44
 Chapter 4: Multi-resolution Global-attention Long Short-Term Memory Network		46
4.1	Introduction	46
4.2	Multi-level Signal Decomposition	47
4.3	Multi-resolution Attention Models	49
4.3.1	Model #A: Single-period Multi-Resolution Attention LSTM	50
4.3.2	Model #B: Multi-period Multi-resolution Attention LSTM	51
4.4	Experiment Setup	52
4.4.1	Experiment #1 Single-period Multi-Resolution Attention LSTM	53

4.4.2	Experiment #2 Multi-period Multi-resolution Attention LSTM	54
4.5	Experiment #1: Performance Analysis of Single-period Multi-resolution Attention LSTM	56
4.5.1	Overall Classification Accuracy Comparison	56
4.5.2	Individual Class Analysis	57
4.6	Experiment #2: Performance Analysis of Multi-period Multi-resolution Attention LSTM	59
4.6.1	Overall Classification Accuracy Analysis	59
4.6.2	Classification Accuracy Analysis	60
4.6.3	Confusion Matrix Analysis	61
4.6.4	Model Complexity Analysis and Benchmarking	63
4.7	Chapter Summary	64
 Chapter 5: Multi-resolution Convolutional Transformer Network		66
5.1	Introduction	66
5.2	Multi-resolution Attention Using Convolutional Transformer	67
5.2.1	Multi-resolution Signal Decomposition	67
5.2.2	temporal aligned Layer	68
5.2.3	Temporal-spatial Embedding	69
5.2.4	Efficient Superposition Embedding	69
5.2.5	Transformer Network	70
5.2.6	Pooling Multi-head Attention	71
5.2.7	Batch Normalization Layer	72
5.3	Experiment Setup	72
5.4	Experiment #1 - Transformer Encoder Network	74
5.4.1	Single-period PQD Analysis on Transformer Encoder Network	75
5.4.2	Multi-period PQD Analysis on Transformer Encoder Network	79
5.5	Experiment #2 Multi-resolution Attention Using Convolutional Trans- former	80
5.5.1	Classification Performance Comparisons of WT-ConvT Versus Deep CNN	81

5.5.2	Efficient Embedding Multi-resolution Attention Using Convolutional Transformer and Complexity Analysis	84
5.5.3	Analysis on Batch Normalization	87
5.6	Chapter Summary	89
Chapter 6: Real-time Embedded System Implementation with Power Quality Classification		90
6.1	Introduction	90
6.2	Proposed Hardware Implementation	91
6.3	Laboratory Hardware and Experiment Setup	94
6.4	Data Acquisition and Pre-processing	95
6.5	Real-time PQD Evaluation and Performance Analysis	97
6.6	Computation Performance Analysis of Proposed Models on Embedded System	101
6.7	Chapter Summary	103
Chapter 7: Conclusion and Future Work		104
7.1	Conclusion	104
7.2	Future Work	108
References		110

List of Figures

1.1	Generalized PQD classification process.	3
1.2	Proposed attention mechanism.	6
1.3	Multi-resolution attention model.	7
1.4	Generalized PQD classification process.	7
2.1	A framework comparison of three PQDs classification methods.	14
3.1	The proposed global attention-based LSTM network.	26
3.2	Global attention-based LSTM model with signal windowing. . .	27
3.3	Self-Attention mechanism.	28
3.4	LSTM architecture.	29
3.5	Waveform and labelling for single disturbance PQDs. (a) D0: Normal, (b) D1: Sag, (c) D2: Swell, (d) D3: Interrupt, (e) D4: Impulse Transient, (f) D5: Spike, (g) D6: Harmonics, (h) D7: Oscillatory Transient, (i) D8: Notch.	32
3.6	Waveform and labelling for combined-disturbance PQD. (a) P_9 : Sag with Harmonics, (b) P_{10} : Swell with Harmonics, (c) P_{11} : Interrupt with Harmonics, (d) P_{12} : Oscillatory Transient with Harmonics.	32
3.7	PQD windowing and labelling process.	33
3.8	(a) Oscillatory Transient time-series signal and its attention out- put at (b) no-noise level and (c) 20dB level; and (d) Interrupt time-series signal and its attention output at (e) no-noise level and (f) 20dB level.	38
4.1	MSD with 4-levels of decomposition.	48
4.2	Single stage reconstruction.	49
4.3	Multi-resolution Attention LSTM.	50

4.4	Multi-resolution attention LSTM model using (a) Global attention, (b) Band attention.	50
4.5	Multi-resolution Attention LSTM model with attention mechanism.	51
4.6	Confusion matrix at 20dB SNR AWGN test on (a) Model #B1: WT-TFA-LSTM and (b) Model #B2: WT-SFA-LSTM.	62
5.1	Proposed wavelet-based convolutional transformer hybrid model.	67
5.2	Temporal aligned layer.	69
5.3	Efficient temporal aligned layer.	70
5.4	Transformer encoder block.	71
5.5	Confusion matrix at 20dB SNR AWGN test on (a) Deep CNN and (b) WT-ConvT.	83
5.6	Confusion matrix for EWT-ConvT with efficient superposition embedding at 20dB SNR AWGN test.	85
5.7	Confusion matrix for EWT-ConvT without BN layer at 20dB SNR AWGN test.	88
6.1	Proposed hardware implementation (embedded system).	92
6.2	Hardware-software integration.	92
6.3	Proposed efficient wavelet-based convolutional transformer (EWT-ConvT) model.	93
6.4	Lab Hardware setup implementation for the proposed classification model.	94
6.5	Real world lab-simulated PQDs waveforms. (a) Sag, (b) Swell, (c) Interrupt, (d) Phase-aligned Sag, (e) Phase-aligned Swell, (f) Phase-aligned Interrupt.	96

List of Tables

2.1	Comparisons of signal processing techniques used for classification of PQD.	19
2.2	Comparison of Model-based and Hybrid Models Classification Methodologies.	20
2.3	Analysis studies of embedded classifier in dedicated hardware.	22
3.1	Disturbance classes and respective labelling.	30
3.2	PQD Data Generation [82, 96].	31
3.3	Performance of Windowing + LSTM under noiseless conditions.	35
3.4	Noise-level Performance Comparison.	36
3.5	Single-class PQD classification performance comparison between the LSTM network and Global Attention LSTM network.	36
3.6	Confusion Matrix of LSTM on PQD with 20dB AWGN.	37
3.7	Confusion Matrix of LSTM with attention mechanism on PQD with 20dB AWGN.	38
3.8	Classification performance of LSTM model trained with noiseless synthetic PQD data and tested with 20-40dB AWGN and noiseless conditions.	40
3.9	Classification performance of LSTM model trained with noisy synthetic PQD data and tested with 20-40dB AWGN and noiseless conditions.	42
3.10	Unseen noise performance comparison.	43
4.1	Class of power quality disturbances (Experiment #1).	53
4.2	Proposed model under setups with global (GA) and band attention (BA), with spatial (S) and temporal (T) features settings.	54
4.3	Class of power quality disturbances (Experiment #2).	55
4.4	Details of different bench-marking models.	55

4.5	Weighted accuracy of the proposed model under setups with global and band attention, with spatial and temporal features settings and compare with Model #A7 ([134]).	56
4.6	Classification performance of Model #A1 and Model #A2 with additive white Gaussian noise from 20dB-50dB and noiseless conditions.	58
4.7	Classification performance comparisons.	60
4.8	Performance of the proposed WT-SFA-LSTM model tested with 20-50dB AWGN and noiseless conditions.	61
4.9	Model complexity comparisons.	63
5.1	Class of power quality disturbances (Experiment #1).	73
5.2	Class of power quality disturbances (Experiment #2).	74
5.3	Noiseless training: Classification performance comparison between LSTM network and Transformer network.	75
5.4	Noisy training: Classification performance comparison between LSTM network and Transformer network.	77
5.5	Testing classification rate in (%) with three unknown noise performance comparison.	78
5.6	Performance comparison of Deep LSTM with the Transformer Encoder model tested with 20-50dB AWGN and noiseless conditions.	79
5.7	Performance comparison of WT-ConvT with the Deep CNN model tested with 20-50dB AWGN and noiseless conditions.	81
5.8	Performance comparison of WT-ConvT with the EWT-ConvT model tested with 20-50dB AWGN and noiseless conditions.	84
5.9	Model complexity comparisons.	86
5.10	Performance comparison of WT-ConvT and EWT-ConvT without batch normalization layer tested with 20-50dB AWGN and noiseless conditions.	87
6.1	Class of power quality disturbances.	93
6.2	Classification performance on laboratory simulated real-world data.	97
6.3	Confusion analysis on four models.	99

6.4	Model performance on RPi.	101
-----	-----------------------------------	-----

List of Acronyms

ABC	Artificial Bee Colony
ANN	Artificial Neural Network
AWGN	Additive White Gaussian Noise
BA	Band Attention
BN	Batch Normalization
CNN	Convolutional Neural Network
CPU	Central Processing Unit
db4	Daubechies 4
DER	Distributed Energy Resources
DFT	Discrete Fourier Transform
DSP	Digital Signal Processor
DT	Decision Tree
DWT	Discrete Wavelet Transform
EMB	Embedded System
ESS	Energy Storage Systems
FC	Fully Connected
FFT	Fast Fourier Transform
FPGA	Field Programmable Gates Array

GA	Global Attention
HHT	Hilbert Huang Transform
HPF	High Pass Filter
IWT	Inverse Wavelet Transform
LPF	Low Pass Filter
LSTM	Long Short-Term Memory network
MG	Microgrid
MHA	Multi-head Attention
ML	Machine Learning
MM	Mathematical Morphological Operations
MRA	Multi-Resolution Analysis
MSD	Multi-level Signal Decomposition
NN	Neural Network
PMA	Pooling by Multi-head Attention
PNN	Probabilistic Neural Network
PQD	Power Quality Disturbance
RAM	Random-access Memory
RER	Renewable Energy Resource
RL	Resistive-Inductive
RMS	Root Mean Square
RPi	Raspberry Pi Model 4B
SFA	Spatial Feature Attention
SG	Smart Grid

SLT	Slantlet Transform
SNR	Signal-to-noise ratio
SSR	Solid State Relay
ST	Stockwel Transform
STFT	Short-time Fourier Transform
SVM	Support Vector Machine
TD	Time-Domain
TFA	Temporal Feature Attention
WAcc	Weighted Accuracy
WPT	Wavelet Packet Transform
WT	Wavelet Transform

Chapter 1

Introduction

1.1 Project Overview

Technology advancement has increased the demand for clean and affordable energy [1]. A recent record-breaking report on energy-related carbon emissions has raised awareness to focus on cleaner energy production [2]. Distributed energy resources (DER) such as distributed generations, and renewable energy resources (RER) become the new trend due to their lower cost of implementation while having less impact on the environment [3]. The abundant nature of environment-friendly RERs plays a crucial role in solving the increased local demand for electrical energy. However, due to the intermittency characteristics of RERs, energy storage systems (ESS) are introduced to ensure stable supplies of electricity with the stored energy [4]. A good example would be the continuous supply of electricity from solar-charged ESS at night time. These ESS include supercapacitors, large-capacity batteries, flywheels, and thermochemical storage systems. The wide adoption of electric vehicles allows the possibility of applying the vehicle-to-home or vehicle-to-grid concepts. These concepts allow the dual use of electrical vehicles' battery packs as ESS [5]. The use of vehicle battery packs as ESS enables the full use of car battery idling capacity to offload the grid at peak load hours [6]. These advanced concepts can only be realised with advanced control mechanisms with active monitoring and control mechanisms.

Microgrid controller is introduced for better integration of the DERs, RERs, and ESS into a small-scale grid system [7]. Microgrid controller coordinates the operation of the connected DERs, RERs, and ESS with flexible power management and control

system [8]. On a bigger scale, a smart grid controller is used to ensure the stable operation of the power grid by active monitoring and control of the DERs and microgrids connected to it [8]. A smart grid system increases the overall stability of the power systems with centralised monitoring and active management on all the smaller systems connected [9]. The advantage of using DERs evolves centralised energy providers towards a decentralised and open energy market [10]. A new challenge has thus appeared where increased penetration of DER has increased the fault level in the grid [11]. Besides that, the increased number of grid-connected electric vehicles may lead to power quality issues which brings a negative impact on the grid [12]. These issues are usually caused by uncontrolled fast charging of electric vehicles which will eventually lead to decreased power performance of the entire grid [13]. On the other hand, the increased penetration of advanced industrial loads and increased type and amount of advanced electronics such as energy-saving LED lamps, switching capacitor banks, and fast switching relays can also degrade the performance of the power grid with the power quality disturbances (PQD) produced [14].

PQD is defined as a disturbance occurring on the standard rating of voltage, current and, frequency [15]. PQD includes voltage dips, harmonics, and oscillatory transients which reduce the performance of electrical devices and shorten their life expectancy [16]. Prolonged downtime caused by disturbances can lead to a big impact on financial loss. The ability to prevent or restore certain power quality events is thus important for the proper functioning of the entire power system. To carry out proper mitigation actions, power quality issues must be first identified. The detection and classification of the occurrence of PQDs are thus critical in ensuring the stable operation of the power systems. Traditionally, history records of the power systems are logged and analysed manually after the faults occurred. However, this process is complex and requires a long analysis before the full restoration of the power system. Real-time monitoring and management of power quality issues are thus important to carry out instant mitigation steps to reduce the downtime of the power systems [17]. Global monitoring systems are usually implemented via advanced metering infrastructure under the smart grid controllers which allow bidirectional communication control over multiple DERs and microgrids connected to it [18]. The bidirectional information acquired from the metering infrastructure is used to analyse the power quality and network performance of the entire grid. However, these systems are usually complex and less feasible for small-scale monitoring. A smaller-scale dedicated hardware or embedded system is

thus required to allow multiple nodes power quality monitoring with reduced cost and increased feasibility [19].

In this chapter, the importance of PQD classification in power grids is discussed. The challenges in the power systems lead to this research, focusing on power quality monitoring systems by having a PQD detection and classification system. The underlying research problems are covered briefly. Section 1.2 discuss briefly the research problems for this research. The objectives and significance of this research are discussed in Section 1.4 and Section 1.5 respectively. Finally, an overview of the contribution from each chapter is summarised in Section 1.6.

1.2 Problems Statement

Classification of PQD occurring in a power system is crucial to identify sources of abnormality or faults occurring within a power system. A general PQD classification process involves two stages, feature extraction, and disturbance classification, as shown in Fig. 1.1. The advancement of technology especially in the field of machine learning and artificial intelligence enables automation in the PQD classification process. However, the classification techniques using machine learning models have encountered several challenges due to feature extraction methods, model generalization capabilities, and increased complexity. Consequently, these model design issue affects the real-time implementation of the classification model under real-world applications. The research problems are outlined as follows:

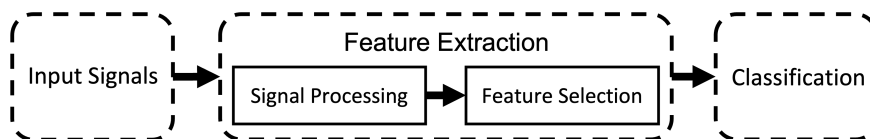


Figure 1.1: Generalized PQD classification process.

1. Classification of PQD relies heavily on the feature set extracted during the feature extraction stage. Traditionally, the use and selection of specific features depend heavily on professional knowledge. Manual statistical feature extraction and selection were normally used to differentiate between PQDs. However, as the complexity of the power grid increases, multiple combinations of PQD could occur

within a single power waveform. The use of handcrafted features becomes impractical to extract distinct features between different PQD classes. The introduction of machine learning to achieve automatic feature extraction and classification is thus proposed by Wang et al. [20] for a more efficient and automated classification process. However, the use of deep neural networks usually required a large number of training samples to achieve better generalization. Most of the literature only focused on improving classification accuracy based on synthetic PQD data without analyzing the generalization capability of the model toward new or unseen conditions in real-world scenarios.

2. Signal processing tools are used to provide better signal representation for distinct feature extraction. Signal processing tools are used for noise suppression to improve the quality of features extracted for better classification performance [21, 22]. Besides that, signal transformations provide alternative insights into the signal by representing the signal in different domain representations. However, these domain feature representations are usually followed by manual statistical feature extraction which set limits to the PQD analysis system in detecting a limited number of PQD based on the professional and preliminary knowledge of the disturbances. An automatic feature selection mechanism is proposed to minimize feature input into a classifier with highly relevant features, but this process increases the complexity of the model with the generation of redundant features [23, 24]. There are very limited studies on the efficient use of signal processing tools with machine learning algorithms to improve the classification performance of PQD.
3. Real-world implementation is the ultimate goal of the research in the field of the classification of PQD. The ability to achieve the highest classification accuracy in detecting more disturbance classes from real-world signals is of top priority in this field of research. Most of the studies in the literature can achieve high classification accuracy using synthetic PQD data and real-world data. The classification process is usually performed using high-power computing platforms such as workstations with high computing power. Most of the studies are performed based on historical data without considering real-time classification performance and feasibility. The fast-growing power grid systems with increasing complexity demand better monitoring approaches with high feasibility and real-time decision-making

capability. However, most studies did not consider real-time analysis, and there is a lack of studies on the PQD classification system using embedded systems.

1.3 Research Questions

To address the current problem statements of PQD classification, several research questions are generated to assist the development of the research. The research questions are listed as follows,

1. The complexity of using hand-crafted features in PQD classification may result in complex neural network model development. How can we generate a generalized feature using a model-based method to detect PQD under a noiseless and noisy environment?
2. How does the input resolution be increased to improve the classification accuracy?
3. How to optimize the neural network model for parallel processing in real-time applications?
4. How to integrate the proposed models into real-time hardware implementation for lab-scale verification?

1.4 Objectives

This research project aims to design and develop an automated PQD detection and classification system using machine learning. The hybrid method with a combination use of signal processing and machine learning is proposed. The objectives of this research can be sub-divided as follows:

1. To perform time-frequency domain analysis using a wavelet-based transformation approach to improve power data resolution for discriminant feature extraction.
2. To design a machine learning model for automated feature extraction and classification, integrating with wavelet-based signal transformation.
3. To verify simulated PQD analysis with real-time hardware implementation in terms of computational performance and feasibility.

1.5 Significance and Contributions

The introduction of machine learning in the power system allows automated PQD detection and classification without human intervention [20]. Signal processing techniques [25] are frequently used in performing time and frequency component analysis due to the increase in data resolution. Besides extracting statistical features, the transformed signals are used for noise removal before PQD classification [26]. In this thesis, a hybrid method using signal processing with machine learning algorithms is proposed. The proposed method mainly includes a signal transformation using multi-level signal decomposition via wavelet transform, followed by classification using the machine learning method. Detailed studies are conducted to demonstrate the performance improvements and limitations of the proposed methods. The scope of work and contribution of this research can be summarised as:

1. A model-based PQD classifier is proposed in Chapter 3 for automated classification. Results show Long Short-Term Memory (LSTM) network has weaker performance on high noise signals. This is due to the characteristics of LSTM in memorizing signal sequences as a feature. Attention-based LSTM is proposed to extract unique characteristics of the PQD signal before LSTM. The proposed method can be summarized as shown in Fig. 1.2. The attention mechanism introduced is proved to have better generalisation on the PQD signals. A more generalised feature extraction is realised which is proven by achieving similar attention score output on signal with and without noise.

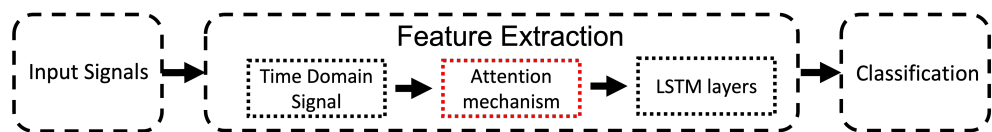


Figure 1.2: Proposed attention mechanism.

2. A hybrid multi-resolution attention model is proposed in Chapter 4 to improve classification performance by having time-frequency domain features. The general diagram of the model is shown in Fig. 1.3. A multi-level signal decomposition feature is introduced to increase input resolution by splitting the frequency components. The split of frequency components allows a clear separation between high and low-frequency signal components. The combined use of multi-level wavelet transform and LSTM allows automatic feature extraction and classification by

removing the manual extraction and selection process for the best statistical features on each of the PQD classes. This is achieved with the proposed novel feature-aligned layer which allows encoding of wavelet coefficients into latent features instead of converting them into statistical features. This process is achieved using a dense perceptron layer and it has the advantage of retaining most of the information from an original signal input. A novel spatial feature extraction between frequency components is proposed in the network. The classification performance of the proposed multi-resolution spatial attention LSTM network exceeded state-of-the-art Deep CNN.



Figure 1.3: Multi-resolution attention model.

3. A novel wavelet-based convolutional transformer (EWT-ConvT) is proposed in Chapter 5 to improve classification and optimize the computation performance of LSTM models. As shown in Fig. 1.4, the convolutional transformer model is an improved model which replaced the LSTM and attention layers. The transformer model has better computation speed with its parallel computation architectural design as compared to sequential LSTM. Besides that, the multi-head attention mechanism within the transformer allows the highlighting of multiple features from the input signal. A novel spatial-temporal embedding layer is proposed to better extract salient features from the MSD coefficients. This embedding layer is achieved by combining the feature-aligned layer with the 1D convolutional layer. Higher-order feature extraction is achieved with the multi-head attention mechanism in the transformer. Besides that, a novel efficient embedding mechanism is achieved using the weight superposition algorithm in the feature-aligned layer. This allows a 61.32% reduction in the proposed model which gives the highest classification performance in this research.

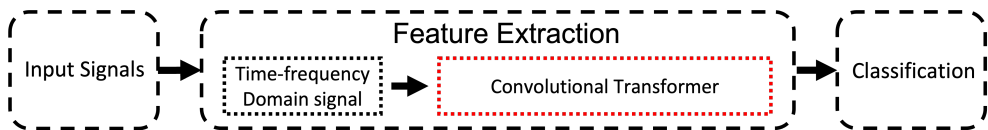


Figure 1.4: Generalized PQD classification process.

-
4. The challenge in achieving real-time PQD classification using an embedded system is addressed in Chapter 6 using Raspberry Pi model 4B (RPi). The proposed system includes a current transformer for grabbing signal waveform, an analog-to-digital converter for digital signal conversion, and RPi embedded system for computation. The proposed model EWT-ConvT achieves the highest classification performance by successfully classifying most of the real-world PQD simulated from laboratory equipment. Besides, the proposed EWT-ConvT can achieve real-time classification requirements on the RPi system. Comparison to literature models on the proposed RPi system shows EWT-ConvT crowns the performance in terms of classification accuracy, computation resources and computation time. The feasibility of implementing real-time PQD classification on an embedded system is verified.

1.6 Thesis Overview

This thesis provides a detailed study and development on power quality disturbance classification for real-time smart grid applications. The current stage of this research thesis is outlined in six chapters as follows:

Chapter 2: Literature Review on Real-Time Power Quality Disturbance Classification

The latest literature reviews on PQD classification and real-time classification of PQD are aggregated in this chapter for comprehensive study and comparison. PQD classification can be categorised into three types, i.e. knowledge-based, model-based, and hybrid-based methods. The importance of having a real-time classification of PQD using embedded systems is discussed in this chapter.

Chapter 3: Global Attention Mechanism with Long Short-Term Memory Network

A model-based PQD classifier is implemented in this chapter to detect PQD signals. LSTM is used for feature extraction while the global attention mechanism is proposed to highlight and extract the distinctive features from the input signal before feeding into the LSTM layer for classification. The classification performance of single

disturbance classes and combined-disturbance classes are studied. Different windowing sizes for the pre-processing stage are compared. Additive white Gaussian noises (AWGN) with 20-50 dB signal-to-noise ratio (SNR) are introduced to study the performance of the proposed method under noisy conditions.

Chapter 4: Multi-resolution Global-attention Long Short-Term Memory Network

Multi-level signal decomposition (MSD) is introduced in this chapter to increase the resolution of the input signal from time-domain representation to time-frequency representation. Discrete wavelet transform with Daubechies 4 (db4) wavelet, and 4 levels of decomposition are used. The study is extended by evaluating the performance using 10-period PQD signals to benchmark with literature models. A novel classification model is proposed by embedding the wavelet coefficients into latent features and performing feature extraction and classification via neural networks.

Chapter 5: Multi-resolution Convolutional Transformer Network

The use of transformer architecture is introduced to replace the LSTM model. Further modification of the architecture is performed with the proposed hybrid PQD classification model named EWT-ConvT. The modifications include the introduction of a temporal-spatial embedding layer, efficient superposition embedding and novel architecture based on a transformer multi-head attention mechanism. The studies on the batch normalization layer are also described in brief. The proposed EWT-ConvT model achieves the highest classification performance with low computation resources required compared to the literature models.

Chapter 6: Real-time Embedded System Implementation with Power Quality Classification

A real-world hardware implementation of PQD classification is proposed to tackle the actual scenario of PQD detection. Three scenarios of signal disturbances are simulated using laboratory equipment. The complete real-time implementation of the proposed model is built to study the computation requirements and limitations of the proposed models. The proposed hardware implementation process includes data acquisition using a current transformer, conversion of analog signals into digital signals, data pre-processing, and finally deployment of the proposed classification models on the

embedded system, Raspberry Pi 4B. The proposed EWT-ConvT model shows the best classification and computation performance with the highest classification accuracy and a low requirement on computation resources.

Chapter 7: Conclusion and Future Work

A summary of every chapter is drawn in this chapter to highlight the research contributions and significance of analyzing results. The advantages and limitations of the proposed power quality disturbance classification are pinpointed with future improvements.

Chapter 2

Literature Review on Real-Time Power Quality Disturbance Classification

2.1 Introduction

Distributed energy systems are characterised by local energy systems which include a complete set of energy generation, conversion, transmission, distribution, and consumption [27]. Although distributed energy systems are beneficial from their flexibility, reliability, and cost-effectiveness in terms of deployment coverage, there is a drawback in the management and control of each of the components [27]. The proliferation of distributed energy resources (DERs) such as distributed generation, and renewable energy resources (RERs) increase the complexity of management and control. The integration of multiple DERs to form a microgrid (MG) is necessary due to the intermittency nature of RERs and usually low capacity from distributed generation [28]. MG integrates multiple DERs and energy storage systems. MG functions as centralised monitoring and control in the local small-scale grid [29]. MG can be perceived as a single controllable entity by the main grid with its ability to work off-grid.

Smart grid systems are introduced to upgrade traditional power grid systems, allowing bi-direction monitoring and control of all the connected systems with ease [30]. The main functionality of the smart grid is to integrate DERs which includes MG by providing grid protection, automated energy distribution, and system optimization. MG

on the other hand allows easy integration of local DERs and energy storage systems, especially with RERs, while providing MG protection [31]. One of the main challenges in the integration of smart grids or MG is the maintenance of the power quality [32]. Power quality disturbance (PQD) issues arise in MG typically caused by intermittent supplies of RERs and advanced power electronics control mechanisms [33]. The most common PQD occurring in MG includes voltage sag, swell, and interruption [34]. Mitigation actions must be taken on the power quality issues occurred. Therefore, the ability to detect and identify types of PQD occurred is required.

Constant monitoring of power quality is important for utility providers, especially for industrial or commercial customers with sensitive equipment [35]. Frequent occurrence of PQD in the networks increases the risk of electric shortage and reduces the lifespan of electrical components. The use of advanced metering infrastructure allows continuous monitoring is thus required. Different types of smart meters are introduced for the monitoring of power parameters [17, 36, 37]. Further analysis of power data recordings is required to specify and identify the fault or disturbance. Automatic detection and classification of PQD are thus introduced for real-time identification of faults or disturbances [38, 39]. The detection and classification of PQD can be used as a predictive maintenance tool. Real-time detection and classification of PQD are highly needed for MG or smart grids to carry out instant mitigation steps. In this chapter, the analysis methods for power quality disturbances are discussed. The structure of chapter 2 is as follows: Section 2.2 reviews the general process of power quality disturbance detection and classification methods. The classification techniques are classified into three methods, knowledge-based, model-based, and hybrid methods described in Section 2.2.1, Section 2.2.2, and Section 2.2.3 respectively. Literature of recent development on the real-time classification of PQD is reviewed in Section 2.3. Finally, Section 2.4 summarises this chapter.

2.2 Power Quality Disturbance Classification

Power quality is characterised by steady supplies of voltage with its frequency and waveform stays within the prescribed range. PQD is defined as the disturbance occurred on the standard rating of voltage, current or frequency. Multiple types of PQDs stated in the IEEE standard 1159-1995 [40] include transient (impulsive and oscillatory), short duration variations (interruption, sag and swell), frequency variations, long

duration variations (sustained under voltages and sustained overvoltages) and steady-state variations (harmonics, notch and flicker). These PQDs are categorized into slow disturbance and fast disturbance for simpler evaluation [41]. Fast disturbance refers to transient-like disturbance (transient, spike, and notch) while slow disturbance refers to a disturbance that spread through the entire period of a sinusoidal waveform (sag, swell, interrupt). These PQDs can cause malfunction or breakdown of sensitive end-user equipment. Classification of PQD is important to allow the identification of the root cause for suitable mitigation actions. PQD affects the normal operation of the power systems, causing financial losses due to equipment breakdown and loss of information. Specific types of mitigation action are required to restore the normal operations of the power system [42]. Accurate identification of the PQD helps in reducing the downtime of the power systems. The PQD classification process can be generalised into two stages, feature extraction and disturbance classification.

The feature extraction stage is used to extract distinct features from the input signal. This process may include signal processing and feature selection. Signal pre-processing can be carried out during signal processing and is usually performed to normalise it into the appropriate phase and magnitude range. Pre-processing signal waveform before classification is an important step for accurate processing [43]. The importance of noise suppression is highlighted in [21, 22]. The use of noise suppression with discrete wavelet transform (DWT) in noise reduction shows promising performance over fast Fourier transform (FFT), Stockwell transforms or S-transform (ST) and wavelet packet transform (WPT); while highlighting the need for denoising for low-level harmonics [44]. A noise suppression algorithm is proposed to improve PQD detection using Hilbert transform and slip-singular value decomposition method [45]. Liu et al. [46] proposed shrinkage thresholding over hard/soft thresholding in wavelet denoising. The use of machine learning (ML) methods in signal processing denoising is proposed using Long Short-Term Memory (LSTM) autoencoder [47]. The noise filtering process can improve the classification rate, but it may also risk removing important information from the signal.

Besides, converting signals into different domains using signal transformation can be achieved during the signal processing stage. The signal transformation phase can be categorized into two main groups, frequency-domain analysis, and time-frequency analysis. The methods used for frequency-domain analysis include Fourier transform (FT) [48], Discrete Fourier Transform (DFT), and FFT [49]. FT is limited towards

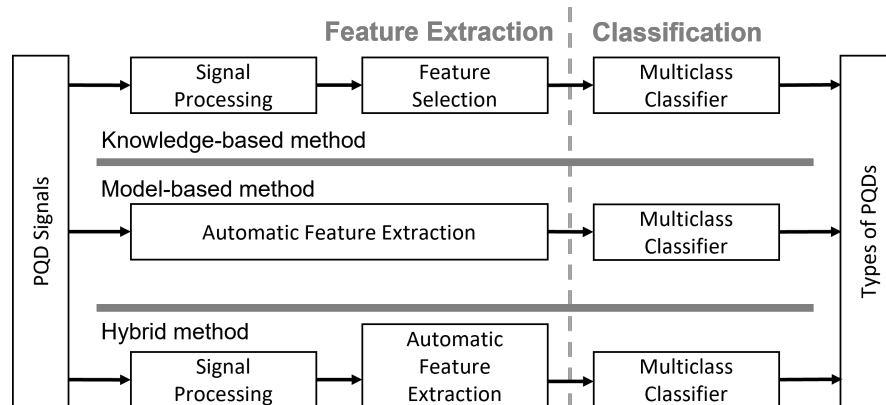


Figure 2.1: A framework comparison of three PQDs classification methods.

analysing stationary signals. FFT is a method similar to DFT but takes less computation time. DFT and FFT are inadequate in analysing time-varying disturbances [50]. Short-time Fourier Transform (STFT) overcome the time-variance issue by introducing time-frequency features [51]. Iterative Hilbert Huang transform (HHT) was introduced in analysing PQD in [52] which solves the close frequencies components. The other time-frequency analysis methods used are Wavelet Transform (WT) [53] and ST [54]. WT includes both time and frequency localization properties which have better analysing power compared to FT. ST evolved from WT by localizing the Gaussian window [54]. The properties of the wavelet transform allow multi-resolution signal decomposition (MSD) which increases the resolution of the input signal into multi-level segments.

Signal transformations are usually followed by statistical feature extraction. Statistical feature extraction can be extracted from two main groups: time-domain statistical features, and frequency-domain statistical features. The statistical features that can be extracted include zero-crossing, amplitude, phase, frequency energy, mode ratio, Energy, Entropy, Standard Deviation, Mean, Kurtosis, Skew, root mean square (RMS) and Range [55]. The selection of specific features extracted normally involves professional knowledge. A recent research has enabled the entire feature extraction process to be done using a machine learning approach [20]. The extracted features are then fed into a classifier for PQD classification. Distinct features extracted from the feature extraction process are used for the PQD classification. Classification of PQDs is studied extensively in [56–58]. In Fig. 2.1, the process of PQD classification can be generalized into three main methods, i.e. knowledge-based method, model-based method, and hybrid methods, which are discussed in-depth in the following subsections.

2.2.1 Knowledge-based Method

The knowledge-based method has two stages in PQD analysis, i.e. feature extraction and feature classification. The knowledge-based method usually involves signal transformation and statistical feature extraction. Literature shows various applications of signal transformations in the field of PQD classification. There is an early example of applying windowed FFT, a version of DFT in detecting power quality problems [49]. Other examples such as DFT [59], STFT [50], and WT [50, 60] are also applied to extract features in the frequency domain. Using FFT and DWT [61] can perform feature vector extraction for better PQD detection and characterization. The use of FT is also explored for steady-state characterization [48]. However, these approaches show inconsistent accuracy in time-varying conditions [48, 49]. The use of STFT using a time-frequency window is introduced with better performance in non-stationary functions [51]. Further studies using WT are performed to solve the fixed resolution of STFT. Zhu et al. [60] proposed wavelet-based fuzzy reasoning in recognizing PQD by extracting energy distribution from MSD decomposition levels. Lopez et al. [62] proposed the combination use of DWT, singular value decomposition and statistical analysis in the classification of PQD. ST is an improved technique from wavelet and is achieved via localizing Gaussian window [54]. Wavelet packet transform with fuzzy k-nearest algorithm is used for optimal feature selection [63]. The embodiment of Slantlet Transform (SLT) into a wavelet basis is proposed for PQD detection which improves the effectiveness in detecting abrupt change [64]. However, having high-computational costs in ST and WPT are not suitable for real-time applications [65, 66]. Iterative HHT with a symbolic aggregate approximation algorithm [52] is proposed to analyze non-stationary signals. However, HHT is limited to narrow-band signals [67].

Signal transformation is usually linked with statistical feature extraction [68]. Statistical features extracted include distribution energy, mean, kurtosis, entropy, RMS, standard deviation, minimum, maximum, and skewness. The use of these statistical features requires professional knowledge. Instead of relying on manual feature selection, optimal feature selection is proposed to automatically select the relevant features from all statistical features generated via machine learning [23, 24, 68]. However, this process generates redundant features and increased the complexity of the network with the optimal feature selection mechanism. Finally, the classification of PQD can be performed from the statistical features extracted. Some of the most common statistical

classification methods include the threshold-based method [69], decision tree (DT) [70–72], k-nearest neighbour [73], pattern recognition using support vector machine (SVM) [74, 75], and expert system [76]. The classification performance of SVM and DT are greatly affected by the training samples used. Expert system has the disadvantage of its high computing cost. The use of knowledge-based methods often requires professional knowledge to maximise classification performance.

2.2.2 Model-based Method

The main issue encountered in the knowledge-based method is the over-reliance of the classifier on the extracted feature. The classification performance of the network is directly associated with the quality of the feature extracted during the feature extraction phase. Proper selection of the hand-crafted statistical features is required from the professional, and the use of the features must be studied extensively before an application. Recently, deep learning (DL) methods have shown marvellous achievement in the field of audio [77], natural language processing [78, 79], and image processing [80]. The use of deep neural networks are therefore implemented and studied in the field of PQD classification. [81–83]. Machine learning and neural networks are used for the feature extraction process, replacing statistical feature extractors. The model-based method is characterised by the use of machine learning models in both feature extraction and disturbance classification processes.

Model-based method uses multiple layers of neural networks to perform both feature extraction and classification process. The use of a neural network as a feature extractor allows automatic feature extraction [84, 85]. Automatic feature extraction and classification without human intervention removed complex manual feature selection stages [86]. The automatic feature extraction process allows the detection of unseen characteristics or conditions that occurred in the signal. The ability to extract features automatically is thus seen as the better approach with minimal information loss. The importance of having a closed-loop feedback system between feature extraction and classification stages is highlighted by Wang et al. [20]. This closed-loop feedback system is seen as an essential element to achieve automatic feature extraction and classification.

The use of the DNN model allows the integration of real-time classification of PQD [81, 87, 88]. Mohan et al. proposed a hybrid combination of convolutional neural network (CNN) and LSTM, CNN-LSTM network for PQD classification [86]. CNN-LSTM network architecture comprises two layers of CNN for spatial information learn-

ing and 1-2 layer of LSTM for temporal characteristic learning. The author achieved the highest accuracy and precision compared to CNN, recurrent neural network, Identity Recurrent Neuro Network, LSTM and Gated Recurrent Units [86]. Balouji et al. proposed an automatic classification model using a 4-layer of LSTM network [87]. PQD classification using Deep LSTM with the number of LSTM layers increasing from 1 to 10 are studied in [86]. Good classification performance can be achieved with the combined use of CNN and LSTM [83, 89, 90]. The Studies of automated feature extraction in classifying voltage dip using deep-CNN is carried out by [91]. The deep convolutional neural network (Deep CNN) proposed in [20] shows magnificent classification performance as compared to ResNet50, LSTM, GRU, and stacked autoencoder. These works of literature prove the capability of model-based methods in achieving high performance in the automatic detection and classification of PQD without human intervention.

2.2.3 Hybrid Method

The combination of knowledge-based techniques and ML techniques leads to the birth of a hybrid method in tackling the shortcomings of knowledge-based or model-based methods. It involves several steps by applying signal transformation and subsequently inserting it into a machine learning model to learn better feature representation in the time or frequency domain. One of the first hybrid methods was presented by Santoso et al. [92, 93], by combining the use of WT and ML for feature extraction and threshold voting for classification. The outputs of WT are squared to form squared wavelet transform coefficients. The extracted features are passed into three parallel artificial neural networks (ANN) layers and simple threshold-ed voting is used as decision-making for the classification. A single-stage multiple power quality event detection using a combination of HHT and PNN is presented in [94]. Another hybrid method of statistical feature extraction with PNN classification was proposed in [95]. Teager-energy operator and mathematical morphology are used to extract statistical features from time-domain PQD signals before passing them into PNN for classification [95]. Optimized ST is used to improve time-frequency resolution using maximum energy concentration with kernel support vector machine functions as a classifier [96]. Compressed Deep Learning (CDL) was proposed by Liu et al. [82] in classifying multiple power quality events. CDL is used to reduce the sampling and computation power. Feature extraction is then carried out on the compressed signal with a sparse autoencoder and deep neural network.

The classification performance of the feed-forward neural network is compared against LSTM [97]. Feature extraction comparisons were performed using STFT, DWT, and ST in finding the best feature extraction methods. The use of WT to extract statistical features and classify using wavelet network and radial basis neural network are presented by Masoum et al. [98] and Kanirajan et al. [99] respectively. The characteristics of WT in transforming signals into multi-level time-frequency representations enable multi-resolution projection of the signals. Khokhar et al. [68] use DWT to extract multi-resolution signal features, and the extracted features are classified using Probabilistic Neural Network (PNN). They also performed optimal feature selection to optimally select the most relevant features for the classification by using Artificial Bee Colony (ABC) optimal feature selection method. The proposed PNN-ABC classification system is proven to have the capability to discriminate single or multiple PQDs with the highest accuracy compared to PNN alone without an optimal selection algorithm [68]. On the other hand, a hybrid classifier using WPT as a feature extractor and ANN as a classifier was proposed in [66]. WPT-ANN is proven to perform better results as compared to FFT-ANN in terms of computation time and accuracy.

2.2.4 Summary of Power Quality Disturbance Classification

PQD can be broadly classified into three categories, knowledge-based, model-based and hybrid methods. Knowledge-based methods are the classical method in detecting and classifying PQDs. This method relies mainly on the use of signal-processing techniques. Usually, statistical features are extracted from the transformed signals. The classification is done by comparing the extracted features via the thresholding method, decision tree, or other mathematical morphological processes. The use of signal processing tools depends heavily on the design of the classifier. The commonly used signal transformation with respective advantages and disadvantages are listed in Table 2.1.

The maturity of machine learning resulted in better approaches in the field of PQD classification. Model-based methods are introduced to allow automatic feature extraction without human intervention. The Feature extraction process is done by simply passing the input signals through multiple layers of neural network. The feedforward and backpropagation mechanism in NN allows feature extraction via learning on the distribution of weights. multiple layers of neural network (NN) in DNN are required for complex feature extraction. Hybrid methods on the other hand are proposed to optimize the computation requirement and computation time. The combined usage of signal

Table 2.1: Comparisons of signal processing techniques used for classification of PQD.

Methods	Advantage	Disadvantage
FT [48, 49, 59]	Low information loss, maintain most of the information on amplitude, harmonics, and phase.	Low resolution in either time or frequency domain.
STFT [50, 51, 97]	Simple implementation, good result for non-varying signals.	Not suitable for time varying signals. Fixed window width, limited time-frequency resolution.
WT [50, 60, 98, 99]	Time and frequency localization. Improved resolution. Improved non-stationary performance.	More sensitive to noise.
WPT [63, 65, 66]	Better control of decomposed signal compared to WT.	More susceptible to noise.
DWT [61, 62, 68, 100]	Reduced filter length, more suitable for real-time implementation	Shift sensitive and lack of phase information.
ST [54, 96]	Frequency-domain analysis, good for time localizing.	Not good for real-time application with false harmonics estimations
HHT [52, 67, 94]	Good feature extraction on highly distorted signals.	Ineffective in constructing frequency spectrum, narrow bandwidth.

processing tools with NN allows automatic feature extraction, feature selections, and classification with minimal human intervention. The combination of statistical methods in feature extraction with NN in classification resulted in increased classification performance. Table 2.2 summarizes the literature model designs of model-based methods and hybrid methods with their respective performances. Besides automatic feature selection, the ability to classify more types of disturbances is a focus of the research trend. The classification accuracy is evaluated by the true positive, TP_n over the total test samples for m classes of PQD denoted in S_j as follows,

$$OverallAccuracy = \frac{\sum_{n=0}^m \frac{TP_n}{\sum_{j=0}^m S_j}}{m}. \quad (2.1)$$

The focus of training methods and efforts lies primarily on ensuring the practicality of application to achieve real-time analysis. The combination of statistical methods with ML methods was researched for performance comparison. Statistical method is

Table 2.2: Comparison of Model-based and Hybrid Models Classification Methodologies.

Model/ Hybrid (M/H)	Classification method	Number of feature selected	Number of PQD	Accuracy (%)
M	DEEP CNN [91]	AUTO	1(voltage dip)	97.72
M	DEEP LSTM [87]	AUTO	1(voltage dip)	93.40
M	DEEP CNN [86]	AUTO	11	98.00
M	HYBRID-DEEP CNN-LSTM [86]	AUTO	11	98.40
M	DEEP LSTM [20, 101]	AUTO	16	99.94
M	DEEP CNN [20]	AUTO	16	99.96
M	DEEP CNN [88]	AUTO	13	99.97
H	FFT+DWT-MLP [61]	4	7	97.10
H	PNN-ABC [68]	9	16	99.87
H	HHT-PNN [94]	15	17	91.60
H	WPT-ANN [66]	8 each	4	99.99
H	CDL [82]	AUTO	31	99.33

used to simplify the NN input which helps in reducing the complexity of the NN. The conventional method PNN-ABC [68] in Table 2.2 achieved high classification accuracy. This is due to the highly optimized feature selected from the feature set using ABC optimization method. In CDL [82], the raw data is compressed statistically before outputting for feature extraction and classification via DNN. Statistical pre-processing is proven in increasing the performance of PQD analysis. Wang et al. [20] highlighted the need for a closed-loop feedback system to link the feature extraction process with the classifier. The research challenge on the design of the classification model thus includes automatic feature extraction, numbers of PQD classes, classification accuracy, the complexity of the network, and real-time performance.

2.3 Implementation Studies in Real-time PQD Classification

Real-time classification of PQD has become an important factor to be considered for implementation feasibility. Massive improvements in technology, especially in the field of processing units allow the exploration of various methods, typically in the field

of machine learning. Real-time classification has a growing interest with the increasing complexity of the power grid. The introduction of smart meters allows big data collection. The benefits of analysing big data are obvious. Real-time PQD analysis allows real-time monitoring and protection of the network by providing instant warning of possible faults occurring on the network. This feature allows automated cut-off or raising awareness to the operator for possible maintenance required. Real-world classification of PQD involves the signal acquisition, analog to digital signal conversion, and finally classification using different types of classifiers. Real-time classification is achieved when the time required to classify a sample is less than the time of the sampling itself.

Real-time automatic PQD classification in the context of smart grids should be made available for better automation control. The implementation of real-time classification of PQD requires the integration of both software and hardware. For instance, an appropriate sensor and analog-to-digital converter are required to convert real-world analog signals into digital format for digital processing. A processor is then required to perform the classification algorithm. The processors are referred to the use of logic gates, digital signal processors (DSP), micro-controller, field programmable gate arrays (FPGA), personal computers, workstations, and cloud computing. The implementation platforms can be categorised into two main groups, i.e. high-performance computing and embedded systems as described in Section 2.3.1 and Section 2.3.2 respectively.

2.3.1 High-Performance Computing

The high-performance computing category includes the use of high computing power such as workstations, personal computers, and cloud computing. The use of massive computing power is an advantage but it's not feasible in terms of scalability and cost-effectiveness. However, most of the studies in the classification of PQD are using PC or workstations. Higher computing power gives the advantage of minimizing the research time. However, these studies need to consider the implementation feasibility with a proper implementation plan to realise their research into real-world applications. The motivation to push the studies of PQD classification to real-world implementation is started by highlighting the computation resources required [20, 102–106]. However, the introduction of higher computation power processors outperforms the computation parameters recorded such as model training time, testing run time, CPU and RAM usage. Besides, information on the generation of synthetic PQD signals using mathematical equations shall be standardised for comparisons. A replicable model design

and validation are important for future comparisons. A good example from [101] study is by replicating the model of Wang et al. [20] on its classification model, and PQD generations using mathematical equations.

2.3.2 Embedded Systems

A limited number of studies can be found on real-time PQD classification on embedded systems (EMB). The application of EMB allows better integration of the real-time PQD classification systems onto the smart grid or microgrid platform. This is especially important with the increasing trend of microgrid applications. A household-based microgrid with a combined use of solar and wind energy has recently become popular due to the maturity of technologies with affordable prices. Table 2.3 shows studies performing real-time PQD classification using EMB. It is worth highlighting that most of the previous studies focusing on knowledge-based methods [107–111]. This is due to the limited processing power from EMB available during the period of studies made. DSPs are mainly used in knowledge-based methods. The use of FPGA can be found using a decision tree as a classifier [109, 110]. The use of a hybrid method on EMB is found to be started by Borges et al. [112] with the use of FFT by converting time domain (TD) input to frequency domain, and the combination use of DT and NN as a classifier. Borges et al. uses μC (TM4C1294NCPDT) and achieved real-time single-period PQD classification with excellent results. Besides, Riberio et al. [113] and

Table 2.3: Analysis studies of embedded classifier in dedicated hardware.

Models ^a	Features ^b	Classification ^b	Dataset ^c	Hardware ^d	PQD	Acc	Real-time analysis	Ref
K	TD, MM	MM, threshold	Real	DSP	10	-	3s 50kHz sampling of data using 121.6ms	[107]
K	Filter, MM	MM, threshold	Real	DSP	-	-	3s 50kHz sampling of data using 121.6ms	[108]
K	TD, DFT	DT	Synth	FPGA	9	99.00	120ms (10 period)	[109]
K	ST, FFT	DT	Synth, Real	FPGA	11	99.27	80ms (10 period)	[110]
K	ST	SVM	Synth	DSP	9	99.50	13.4ms (10 period @ 5 kHz)	[111]
H	TD FFT	DT, DNN	Synth, Real	μC	8	99.30	1ms (1 period @ 1.28 kHz)	[112]
H	Filter, TD	DNN, bayesian	Lab	FPGA	20	97.80	unspecified	[113]
H	HHT	DNN	Synth/lab	FPGA	16	99.81	14ms (6 period @ 3.2kHz)	[114]
M	DNN	DNN	Synth/Lab	FPGA	16	99.40	23.3ms (6 period @ 10kHz)	[115]

^a Methods of classification: K (Knowledge-based), M (Model-based), H (Hybrid).

^b TD (Time-Domain), MM (Mathematical Morphological operations), DT (Decision-tree), SVM (State-Vector-Machine), DNN (Deep Neural Network).

^c Synth (Synthetic), Lab (Laboratory Setup).

^d DSP (Digital Signal Processor), FPGA (Field-Programmable Gate Array), μC (Micro-controller).

Sahani et al. [114] demonstrate good performance with a hybrid real-time classification model on FPGA with more disturbances. Recently, Wang et al. [20] highlighted the importance of having closed-loop feedback systems on the feature extraction stage and classifier. In addition, Sahani et al. [115] proposed a model-based method in the PQD classifier to speed up the real-time processing with less involvement in feature extraction.

2.3.3 Summary of Implementation of Real-time PQD Classifier

The studies on real-time PQD classification are important with the increasing complexity of power systems. Microgrid applications are widely applied due to the proliferation application of renewable energy. The introduction of advanced electronics such as energy-saving LEDs, and advanced charging mechanisms of electric vehicles increase the complexity of the power systems. The applications of edge protection systems are thus essential to meet the demand for the decentralised production of electricity. Most of the previous studies focused on real-time PQD classification systems based on computer systems. However, there is a need to improve the feasibility of the study on PQD classification by reducing the complexity of embedded system deployment. The study on PQD classification should take into consideration the embeddability of the proposed system on the embedded system. Besides, the proposed method shall focus on the ability to replicate the systems onto different platforms for easy comparisons.

2.4 Chapter Summary

The implementation of PQD classification can improve the reliability of the smart grid by monitoring power signals continuously in a real-time manner. It can be categorised into three types, i.e. knowledge-based method, model-based method and hybrid method. The literature review shows that knowledge-based methods rely mostly on signal-processing techniques. Statistical features are usually extracted after signal transformation and then followed by classification using mathematical morphology and decision, or threshold-based methods. Recently, the increased maturity of machine learning algorithms especially the deep learning approach has provided a great impact in the field of PQD classification. The model-based method applies a deep learning approach to automate feature extraction directly from data learning. Automatic feature extraction avoids complex feature extraction processes and the need to generate a large number of statistical features. The hybrid method on the other hand combines

both advantages of the knowledge-based method and the model-based method. Signal processing steps are carried out to guide the learning of the ML models.

The current studies mainly focus on developing a PQD classifier to achieve high accuracy in various disturbances scenarios. However, there is a challenge to achieve a real-time PQD classifier due to limited computing power and algorithm complexity. The computation performance such as the computation resources and run time of the models are the critical parameters to measure the feasibility of the PQD monitoring system. As a result, the implementation using an embedded system can provide a solution with higher scalability and cost-effectiveness in tackling the booming integration of renewable energy resources with the power grid. In conclusion, this research explores different model-based methods and hybrid-based PQD classification models to achieve fast and accurate classifier models on real-world applications.

Chapter 3

Global Attention Mechanism with Long Short-Term Memory Network

3.1 Introduction

Attention mechanism is first introduced in the field of machine translation by Bahdanau et al. [116]. Attention mechanism has grown in popularity among the artificial intelligence community with diverse applications such as natural language processing [117], computer vision [118], and speech recognition [119]. The attention mechanism is inspired by human biological systems, particularly human selective concentration on inputs such as visuals or sounds in perceiving things. For example, humans tend to focus visually only on certain parts of the image while perceiving things, ignoring less important information in the image. Highly relevant information was highlighted using limited processing resources available. Human attention can be categorised into unconscious salient-based attention and consciously focused attention [120]. Salient-based attentions are triggered based on external stimuli such as loud voices and can be related to the max-pooling mechanism. Focused attention on the other hand has a predefined purpose such as focusing on keywords while reading articles. Focused attention is the most applied attention mechanism in machine learning especially on task-specific applications [120]. Generally, the attention mechanism improves the efficiency and accuracy in extracting distinct information for further processing.

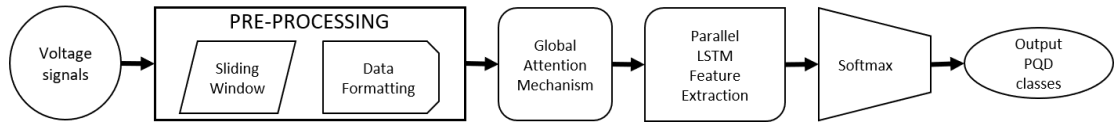


Figure 3.1: The proposed global attention-based LSTM network.

Detection and classification of power quality disturbance (PQD) typically involves two stages, i.e. feature extraction, and classification stage. Deep neural networks are used in [20, 81] for both feature extraction and classification. The automatic feature extraction process eliminates the need for hand-crafted features and avoids complex feature selection processes from the list of statistical features generated. In this chapter, Global attention-based Long Short-Term Memory (LSTM) is proposed for time-series PQD classification. The proposed method consists of a global attention mechanism for highlighting salient features, a layer of LSTM architecture for feature extraction and a fully connected layer with a softmax activation function for classification. A global attention mechanism is added between the input signal and the LSTM layer to highlight the characteristics of input signals. The feature vector output from the attention layer is used by LSTM to extract higher-order representation from the sequence of the signal’s magnitude. The entire automatic detection and classification process is demonstrated in Fig. 3.1. This chapter includes multiple analyses of the proposed global attention-based LSTM network on PQD classification. Section 3.2 includes the theory and design of the proposed model. Section 3.3 describes the experiment setups for Section 3.4 and Section 3.5, and data generation with pre-processing stages to train and evaluate the proposed model. Finally, Section 3.6 summarises the finding in this chapter.

3.2 Global Attention-based Long Short-Term Memory Network

LSTM is used with an attention layer to improve classification accuracy under noisy conditions. This proposed method includes an attention layer between input data and the LSTM layer. The attention layer is used to highlight the abnormalities presented in single-windowed power signal input (global attention) before feeding into the LSTM layer. The important features in the raw signal are first valued in the attention layer and then passed into LSTM for higher dimensional feature extraction. The proposed model is summarised in Fig. 3.2. Pre-processed data were passed into the self-attention layer

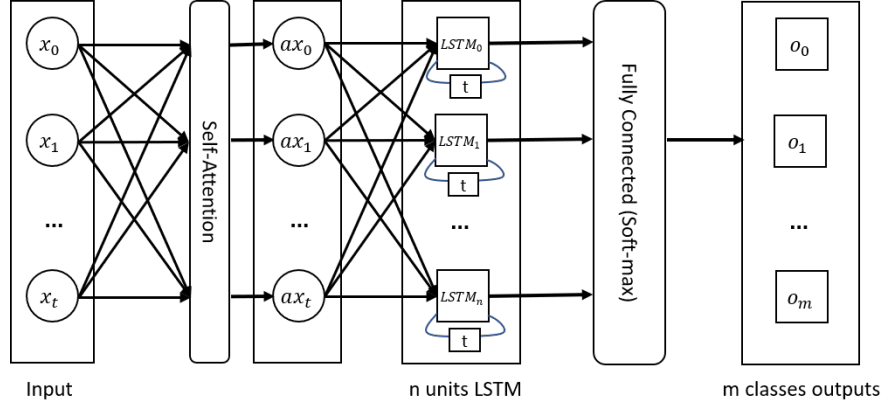


Figure 3.2: Global attention-based LSTM model with signal windowing.

before feeding into LSTM for feature extraction. The term "global attention" is used instead of "self-attention" as the attention is applied over the entire input time-series signals. The details of the attention mechanism are described in the next subsection. A fully connected (FC) layer is used to classify the output of the higher dimensional features from the LSTM layer into different classes of PQDs. The network input vector X retrieves the shifting signal as follows,

$$X = [b_i, \dots, b_{i+T}], \quad (3.1)$$

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 $t - 1$.

3.2.1 Global Attention Mechanism

The self-attention mechanism is proposed to highlight specific features in the signal before feeding into LSTM for the feature extraction process. The attention is applied to all data points present in each input sample, thus forming a global attention mechanism. This global attention mechanism is achieved using a dense layer with softmax activation function as shown in Fig. 3.3. Input signal x_t with T window size or time-steps is first passed into a dense layer to obtain the attention score, y_d as follows,

$$y_d = \sum_{t=0}^T w_{d,t} \cdot x_t, \quad (3.2)$$

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 a

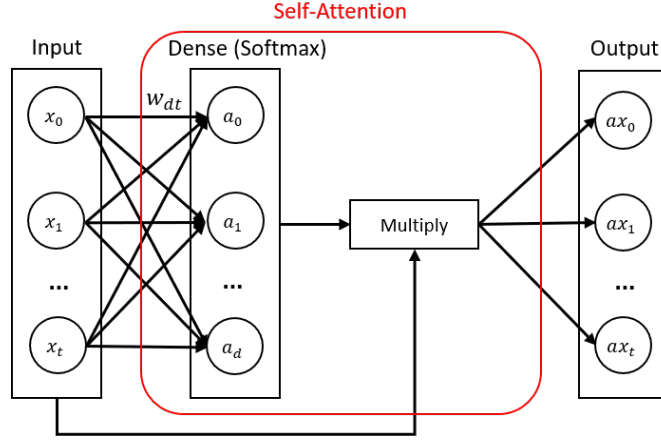


Figure 3.3: Self-Attention mechanism.

range between $(0, 1)$ as follows,

$$a_d = \text{softmax}(y_d) = \frac{e^{y_d}}{\sum_{j=0}^T e^{y_{d_j}}}. \quad (3.3)$$

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

$$ax_t = a_d \odot x_t. \quad (3.4)$$

Attention score represents the important features through a dense layer. The attention weights are multiplied with the input signal which highlights the original signals with trained weight distribution. The highlighted feature vector has higher noise immunity compared to raw signal without attention mechanism.

3.2.2 Long Short-Term Memory

Feature vector output from the attention layer is passed into the LSTM layer for feature extraction. An LSTM architecture [121] can be depicted in Fig. 3.4. Inputs to the LSTM cell consist 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 is processed via multiple gates which control the information flow. There are three "gates" present in the 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, where a value of 1 means to keep all previous memory. The input gate decides which information to retain. The output of the input gate is filtered by the tanh activation function,

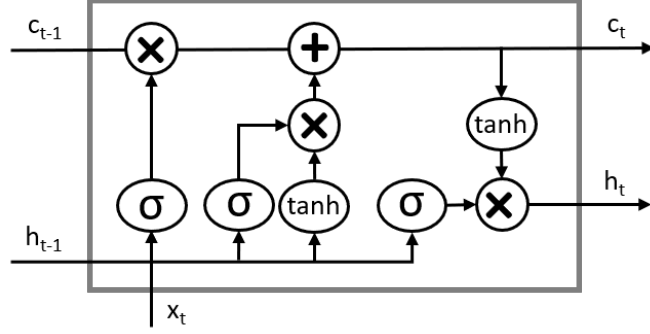


Figure 3.4: LSTM architecture.

producing a new candidate, \tilde{c} for the cell state, while W represents trainable weights and b is the bias.

$$\begin{pmatrix} f_t \\ i_t \\ o_t \\ \tilde{c}_t \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \cdot [h_{t-1}, x_t] + b. \quad (3.5)$$

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.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t. \quad (3.6)$$

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 the 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 the 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,

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \quad (3.7)$$

$$h_t = o_t \odot \tanh(c_t). \quad (3.8)$$

The temporal features of the input are extracted in the LSTM layer. These temporal features representing the specific feature of the PQDs are encoded into higher dimension

representation for the classification process via dense layers. In our experiment, the pre-processed, 1-dimension single-period windowed power waveform which consists of 200 time-steps is used as input. The important characteristics of different classes of PQD are first highlighted in the attention layer. The outputs from the attention layer are fed parallel into 200 units of LSTM. The final hidden state output from LSTM representing extracted features are passed into the FC layer for the classification process.

3.3 Experiment Setup

The classification was implemented using Keras with the Tensorflow backend. The formatted input vector is partitioned into 70% training data, 15% validation data and 15% testing data. The list of PQD signals for analysis is tabulated in Table 3.1. The signals are generated using the mathematical model as listed in Table 3.2, where 10 classes of single-disturbance PQD including normal signal, and 8 classes of combined-disturbance PQD are trained for automatic PQD classification. Each PQD class has 200 samples of the three-period waveform generated with a 10kHz sampling frequency. Point labelling is used where the magnitude difference with more than $\pm 0.005\%$ compared to pure sine wave was labelled as a specific disturbance class. The generated PQDs are displayed in Fig. 3.5 for single-disturbance and Fig. 3.6 for combined-disturbance PQD. The sliding window technique is used to segment the voltage signal as shown in Fig. 3.7. Three different window sizes were used, full-period, half-period, and quarter-period windows, with 200, 100, and 50 data points respectively at 10kHz sampling. A

Table 3.1: Disturbance classes and respective labelling.

Label	Class Description	Label	Class Description
P_0	Normal	P_9	Flicker
P_1	Sag	P_{10}	Sag+Harmonics
P_2	Swell	P_{11}	Swell+Harmonics
P_3	Interrupt	P_{12}	Interrupt+Harmonics
P_4	Impulse Transient	P_{13}	Harmonics+Notch
P_5	Spike	P_{14}	Sag+Transient
P_6	Harmonics	P_{15}	Swell+Transient
P_7	Oscillatory Transient	P_{16}	Sag+Oscillatory Transient
P_8	Notch	P_{17}	Swell+Oscillatory Transient

Table 3.2: PQD Data Generation [82, 96].

Labels	Mathematical Equations and Parameters
P_0	$y(t) = A[1 \pm \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$ Parameters: $\alpha \leq 0.1; T \leq t_2 - t_1 \leq 3T$
P_1	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T$
P_2	$y(t) = A[1 + \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T$
P_3	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$ Parameters: $0.9 < \alpha \leq 1; T \leq t_2 - t_1 \leq 3T$
P_4	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$ Parameters: $0 \leq \alpha \leq 0.414; T/20 \leq t_2 - t_1 \leq T/10$
P_5	$y(t) = \sin(\omega t) + \text{sign}(\sin(\omega t)) \times [\sum_{n=0}^8 K \times [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.02n))]]$ Parameters: $0 \leq t_1, t_2 \leq 0.5T; 0.01T \leq t_2 - t_1 \leq 0.05T; 0.1 \leq K \leq 0.4$
P_6	$y(t) = A[\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i^2 = 1$
P_7	$y(t) = A[\sin(\omega t) + \alpha e^{-\frac{(t-T/2)}{\tau}} \sin \omega_\pi(t-t_1)(u(t_2) - u(t_1))]$ Parameters: $0.1 \leq \alpha \leq 0.8; 0.5T \leq t_2 - t_1 \leq 3T; 8ms \leq \tau \leq 40ms; 300 \leq f_n \leq 900Hz$
P_8	$y(t) = \sin(\omega t) - \text{sign}(\sin(\omega t)) \times [\sum_{n=0}^8 K \times [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.02n))]]$ Parameters: $0 \leq t_1, t_2 \leq 0.5T; 0.01T \leq t_2 - t_1 \leq 0.05T; 0.1 \leq K \leq 0.4$
P_9	$y(t) = [1 + \alpha_f \sin(\beta\omega t)] \sin(\omega t)$ Parameters: $0.1 \leq \alpha_f \leq 0.2; 5Hz \leq \beta \leq 20Hz$
P_{10}	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))][\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i^2 = 1$
P_{11}	$y(t) = A[1 + \alpha(u(t-t_1) - u(t-t_2))][\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i^2 = 1$
P_{12}	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))][\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.9 < \alpha \leq 1; T \leq t_2 - t_1 \leq 3T; 0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i^2 = 1$
P_{13}	$y(t) = A[[\sin(\omega t) - \text{sign}(\sin(\omega t)) \times [\sum_{n=0}^8 K \times [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.02n))]]] + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)]$ Parameters: $0.05 \leq \{\alpha_3, \alpha_5, \alpha_7\} \leq 0.15; \sum \alpha_i^2 = 1; 0 \leq t_1, t_2 \leq 0.5T; 0.01T \leq t_2 - t_1 \leq 0.05T; 0.1 \leq K \leq 0.4$
P_{14}	$y(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))][1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t) \sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0 \leq \alpha \leq 0.414; T/20 \leq t_2 - t_1 \leq T/10$
P_{15}	$y(t) = A[1 + \alpha(u(t-t_1) - u(t-t_2))][1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t) \sin(\omega t)$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0 \leq \alpha \leq 0.414; T/20 \leq t_2 - t_1 \leq T/10$
P_{16}	$y(t) = A[1 - \alpha(u(t_2) - u(t_1))] \sin(\omega t) + \alpha e^{-\frac{(t-T/2)}{\tau}} \sin \omega_\pi(t-t_1)(u(t_2) - u(t_1))$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.1 \leq \alpha \leq 0.8; 0.5T \leq t_2 - t_1 \leq 3T; 8ms \leq \tau \leq 40ms; 300 \leq f_n \leq 900Hz$
P_{17}	$y(t) = A[1 + \alpha(u(t_2) - u(t_1))] \sin(\omega t) + \alpha e^{-\frac{(t-T/2)}{\tau}} \sin \omega_\pi(t-t_1)(u(t_2) - u(t_1))$ Parameters: $0.1 < \alpha \leq 0.9; T \leq t_2 - t_1 \leq 3T; 0.1 \leq \alpha \leq 0.8; 0.5T \leq t_2 - t_1 \leq 3T; 8ms \leq \tau \leq 40ms; 300 \leq f_n \leq 900Hz$

full period window is equivalent to 0.02 seconds of the 50Hz voltage waveform. 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 pieces of information, the magnitude of the signal and a one-hot-encoded window labelling with 9 or 18 categories for single-disturbance PQD classification and combined-disturbance PQD classification respectively. Data formatting is done before the input

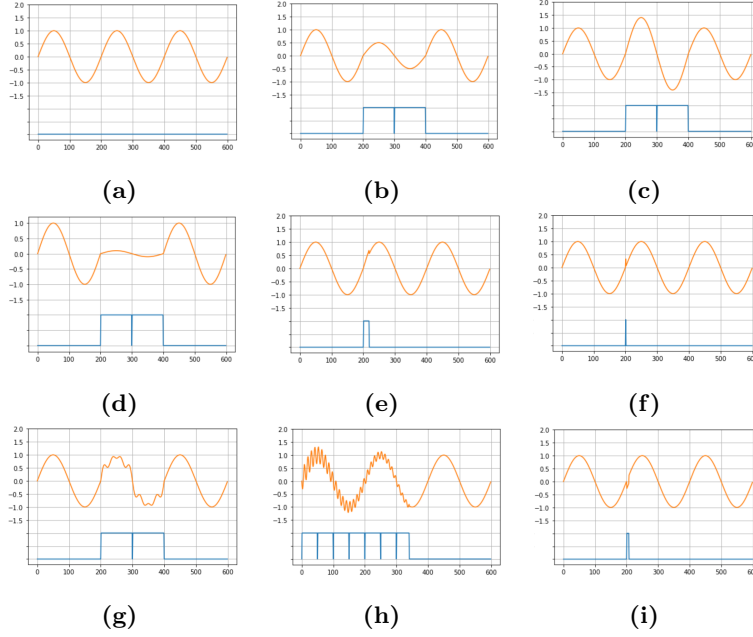


Figure 3.5: Waveform and labelling for single disturbance PQDs. (a) D0: Normal, (b) D1: Sag, (c) D2: Swell, (d) D3: Interrupt, (e) D4: Impulse Transient, (f) D5: Spike, (g) D6: Harmonics, (h) D7: Oscillatory Transient, (i) D8: Notch.

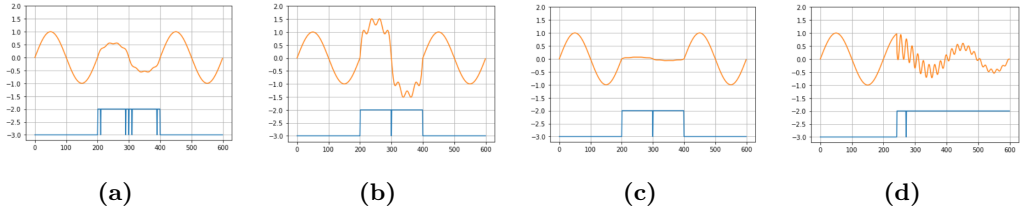


Figure 3.6: Waveform and labelling for combined-disturbance PQD. (a) P_9 : Sag with Harmonics, (b) P_{10} : Swell with Harmonics, (c) P_{11} : Interrupt with Harmonics, (d) P_{12} : Oscillatory Transient with Harmonics.

of windowed data into the LSTM layer by normalising the signal magnitude into ranges of -1 to 1 . Normalization is done by dividing the signal magnitude by the maximum absolute amplitude which can be expressed as,

$$V(t) = \frac{v(t)}{\max_{t \in n} |v(t)|}, \quad (3.9)$$

where n represents the number of windows and t represents window size. The formatted input, $V(t)$ is the normalization of magnitude $v(t)$ over the maximum absolute amplitude present in the entire data sample. The number of hidden units of the LSTM layer is

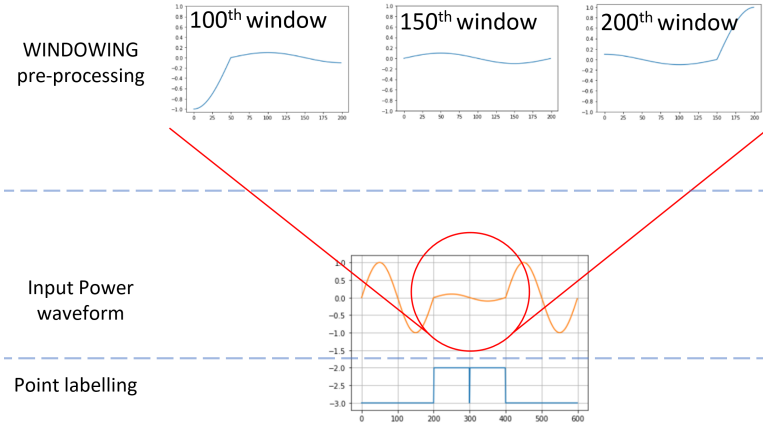


Figure 3.7: PQD windowing and labelling process.

equal to the sizes of windowing. The main evaluation matrix used in these experiments is 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,

$$Acc_n = \frac{TP_n}{\sum_{j=0}^m S_j}. \quad (3.10)$$

Weighted accuracy is used to overcome the imbalanced data sample issue. The weights of each class are calculated by dividing the total number of samples of an individual class by the total number of samples. Weighted accuracy, (WAcc) can be calculated by multiplying the individual class accuracy with its weight, which can be depicted as,

$$WAcc = \sum_{j=0}^m W_j \times Acc_j. \quad (3.11)$$

Two experiments are set up to test the performance of the proposed network in PQD classification. Experiment #1 involves classification performance analysis of the LSTM network and proposed global attention LSTM network (GA-LSTM) on single disturbance PQDs. Three window sizes, i.e. full-period, half-period, and quarter-period of voltage signal waveform having 200, 100, and 50 data points respectively are used to test the single-disturbance classification performance of the LSTM network. The second part of this section compared the classification performance proposed GA-LSTM network against LSTM under noisy conditions. 20-40 dB Signal-to-noise (SNR) levels of additive white Gaussian noise (AWGN) are added to test the network performance. On the other hand, experiment #2 introduces the 8 classes of combined disturbance

into the network for training. The performance of the proposed global attention-based LSTM was discussed. The second part of the experiment focused on noisy condition analysis, where random levels of 20-40 dB SNR AWGN are introduced randomly into the training samples. Three unseen noises were introduced to test the generalization capability of the models.

3.4 Experiment #1 - Single Disturbance Analysis

In this experiment, the single disturbance PQDs classification performance of the proposed GA-LSTM network is analysed and compared to the LSTM network without the attention mechanism. The first part of this experiment analyses the effect of different windowing sizes on the classification performance for the LSTM model. The second part of this section compared the classification performance proposed GA-LSTM network against LSTM under noisy conditions. 20-40dB SNR AWGN are added to test the network performance.

3.4.1 Analysis on Signal Windowing

Three different window sizes were used to study the performance of the network. The experiment is first carried out using a window size of one period, followed by half-period and quarter-period. The training is carried out for 30 epochs and the performance of the network under noiseless conditions is summarized in Table 3.3. The highest WAcc as shown in Table 3.3 is 99.31% for Test B with a half-window size, followed by Test C then Test A. In comparison to Test A, Test B exhibited an overall enhancement in classification accuracy. However, the individual classification accuracy of classes D4 and D5 dropped from Test A to Test B and drops further from Test B to Test C. From the result, a larger window size has better performance in short-period high-frequency disturbance classes such as D4 and D5 and D8. Decreasing window size reduces the accuracy of these classes.

3.4.2 Performance Comparison With Noisy Data

Time Series power signal features can be extracted using LSTM based on the sequential characteristics of the signal. The features for the classification process can be extracted automatically using LSTM. Time series PQD classification using machine learning depends heavily on class labelling for training. When a sliding window is

Table 3.3: Performance of Windowing + LSTM under noiseless conditions.

Test Label	A	B	C	
Window Size	1	1/2	1/4	
No. of LSTM hidden state	200	100	50	
Training Loss	0.0145	0.0145	0.0216	
Training Accuracy(%)	99.51	99.49	99.29	
Validation Loss	0.0273	0.0198	0.0227	
Validation Accuracy(%)	99.21	99.32	99.21	
Testing Categorical classification				
Accuracy (%)	D0 Normal	99.97	99.92	99.84
	D1 Sag	97.84	98.16	96.99
	D2 Swell	97.62	98.01	98.79
	D3 Interrupt	99.68	99.51	99.54
	D4 Impulse Transient	94.21	92.17	87.54
	D5 Spike	98.12	96.13	85.67
	D6 Harmonics	99.94	99.93	99.52
	D7 Oscillatory Transient	99.01	98.84	97.86
	D8 Notch	99.91	99.38	97.95
WAcc(%)	98.95	99.31	99.19	

used, the appropriate window size determines the classification accuracy. In Table 3.4, two sizes of the windows are fed into two LSTM models with and without attention mechanism to evaluate its data representation. As a result, the full-period window has a significantly better performance compared to the half-period window. The proposed model without attention mechanism is drastically affected by the added AWGN in the original signal. Although the proposed GA-LSTM is also affected by the noise, it is still able to maintain the WAcc above 50%. This shows that the attention mechanism has successfully highlighted the important features under noisy conditions. The attention mechanism can perform in highlighting the important features in each disturbance class even under 20dB AWGN.

Table 3.4: Noise-level Performance Comparison.

Noise	Without attention		With attention	
	1/2 period	Full period	1/2 period	Full period
noiseless	99.31	98.95	94.22	95.15
40 dB	59.81	49.62	91.39	92.63
30 dB	30.08	34.36	63.89	77.58
20 dB	15.86	20.52	27.02	52.53

Table 3.5: Single-class PQD classification performance comparison between the LSTM network and Global Attention LSTM network.

Network Class \ SNR	LSTM				Global Attention LSTM			
	20dB	30dB	40dB	noise-less	20dB	30dB	40dB	noise-less
D0	0.00	0.00	0.09	99.98	15.76	55.61	92.35	99.11
D1	7.97	54.14	85.20	97.84	87.75	95.18	96.80	96.89
D2	6.17	42.61	69.96	97.62	65.27	93.27	97.96	98.62
D3	48.06	74.17	90.84	99.69	79.64	93.30	98.53	98.84
D4	0.00	4.96	36.55	94.21	32.57	61.48	71.90	72.84
D5	0.00	4.73	34.98	98.13	41.17	68.91	77.67	79.46
D6	57.10	63.40	77.91	99.94	88.18	98.89	99.78	99.80
D7	99.12	99.81	99.79	99.02	62.85	81.53	88.39	89.25
D8	0.55	10.29	44.31	99.92	58.62	82.05	90.36	91.08
WAcc	20.53	34.37	49.63	98.96	52.53	77.58	92.63	95.15

In Table 3.5, it shows an accuracy matrix of detailed noise performance using the LSTM model with and without attention mechanism in a single full period window. The introduction of a minor 20-40dB SNR AWGN has a substantial impact on the classification performance of the LSTM model. LSTM model can only obtain 20.53% WAcc under 20dB SNR AWGN with the observed information during the training process. It can be noticed that the proposed attention mechanism significantly improved the clas-

Table 3.6: Confusion Matrix of LSTM on PQD with 20dB AWGN.

	D0	D1	D2	D3	D4	D5	D6	D7	D8
D0	0	0	0	0	0	0	0	0	0
D1	14	961	0	0	0	0	7	0	1
D2	0	0	731	0	0	0	0	0	0
D3	0	915	0	5671	0	0	0	0	0
D4	0	0	521	0	0	0	0	0	0
D5	0	0	6	0	0	0	0	0	0
D6	982	735	1402	101	151	96	6920	64	126
D7	33136	9017	9183	6017	5626	6088	5163	7853	5836
D8	145	425	12	12	10	5	21	6	33
Acc(%)	0	7.973	6.166	48.05	0	0	57.10	99.11	0.5547
WAcc (%)				20.53					

sification performance under noisy conditions, especially in 20dB conditions. Table 3.6 shows that the confusion matrix of the LSTM model can classify most of the samples as class D7, Oscillatory Transient. This is due to the high similarities between noisy signals and the nature of Oscillatory Transient. The proposed model without the attention mechanism is thus prone to noise and only perform better under noiseless condition.

When the 20dB SNR is introduced, the attention mechanism successfully highlights the feature of different disturbance classes under noisy conditions, thus improving the classification accuracy to 52.53%, and more than 50% for most of the classes. The global attention mechanism proposed successfully highlighted the input signal. The input signal and output attention vector can be depicted in Fig. 3.8. Fig. 3.8 shows the highlighted feature vector for Oscillatory transient and Interrupt under 20dB SNR and noiseless condition. Specific feature vectors are highlighted for each of the disturbance classes. This shows that the attention layer successfully learns how to highlight the specific feature for each of the disturbance waveforms. From 20dB SNR to no-noise condition, there are only slight changes in the magnitude of the highlighted features. This slight change indicates that the attention mechanism has good immunity against noisy conditions.

Nonetheless, the global attention mechanism still has an existing drawback. Class

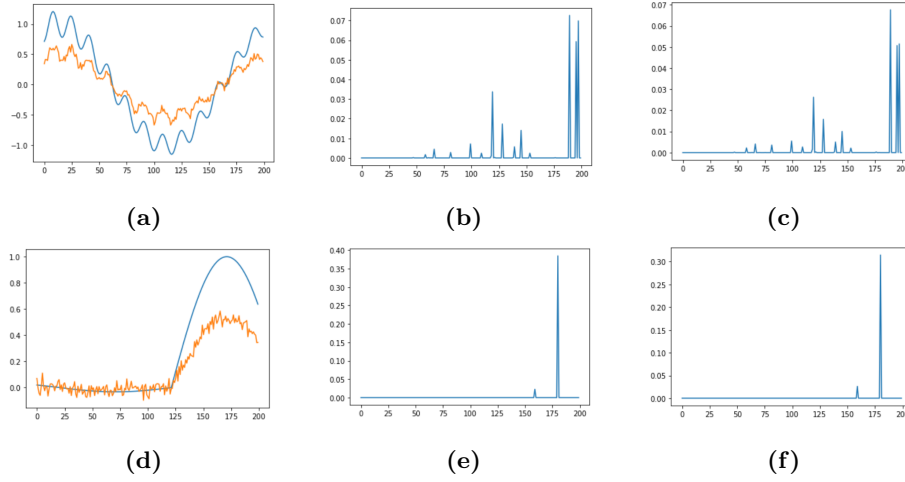


Figure 3.8: (a) Oscillatory Transient time-series signal and its attention output at (b) no-noise level and (c) 20dB level; and (d) Interrupt time-series signal and its attention output at (e) no-noise level and (f) 20dB level.

D0, D4 and D5 obtain a classification accuracy of less than 50%. The classification accuracy of Class D0 or normal class dropped drastically with the increase in SNR.

Table 3.7: Confusion Matrix of LSTM with attention mechanism on PQD with 20dB AWGN.

	D0	D1	D2	D3	D4	D5	D6	D7	D8
D0	5403	167	312	6	347	434	11	190	287
D1	1537	10576	2632	2125	169	111	173	101	156
D2	275	104	7738	2	84	45	135	14	18
D3	102	482	11	9398	23	16	3	44	18
D4	4138	66	256	7	1885	769	212	420	653
D5	5684	103	200	35	1329	2548	254	620	301
D6	1248	82	188	7	280	408	10686	257	195
D7	6302	164	217	110	626	1180	407	4980	853
D8	9588	309	301	111	1044	678	238	1297	3515
Acc(%)	15.76	87.75	65.27	79.63	32.57	41.17	88.18	62.85	58.62
WAcc (%)				52.53					

This drop in accuracy can be visualised from the confusion matrix Table 3.7. Most of the wrongly classified "normal" are related to the D4, D5 D7 and D8 classes. Class D0 is "Normal waveform". When a higher level of noise is added, the waveform is distorted to other similar disturbance classes. This happens especially in those classes with higher disturbance frequency, such as class D4, D5, D7 and D8. Disturbance classes D4 and D5 are very similar in nature, except the duration of a spike is much smaller compared to impulse transient. Class D4 Impulsive Transient, class D5 Spike and class D8 Notch are all having shorter disturbance periods over time. This indicates the imperfect focus of the attention mechanism on "impulse-like" or higher detailed signal.

Since both models are trained with noiseless data, the result shows that the global attention mechanism is more generalised to different noise levels. The noise impact on the attention model is much smaller compared to the LSTM model. The global attention model can achieve better overall classification performance on single disturbance PQD signals polluted with noise. In the next section, combined disturbances are added to the analysis, mimicking real-world scenarios of multiple disturbances occurring within the same timeframe.

3.5 Experiment #2 - Combined-disturbance Analysis

This experiment can be further subdivided into two parts. In the first part of this experiment, combined disturbances are added to the network analysis. The classification performance of the LSTM and GA-LSTM on combined disturbances are discussed. The second part of the experiment is carried out by training the LSTM network and GA-LSTM network using noisy data input. For instance, random levels of 20-50dB SNR AWGN are added randomly into the training samples. Besides that, the analysis proceeds with testing on unseen noises condition, where three unseen noises were introduced. The generalization capability of the model towards different noise conditions is analysed.

3.5.1 Experiment #2 Noiseless Training Analysis

In this experiment, multiple disturbances are added to the training and testing datasets. Similar to Experiment #1, The LSTM model and global attention LSTM models are trained using noiseless datasets and tested with 20-40dB SNR AWGN and noiseless data. Results tabulated in Table 3.8 show overall better classification accuracy

Table 3.8: Classification performance of LSTM model trained with noiseless synthetic PQD data and tested with 20-40dB AWGN and noiseless conditions.

Attention Class \ SNR	Without Attention				With Attention			
	20dB	30dB	40dB	noiseless	20dB	30dB	40dB	noiseless
P_0	0.00	0.00	6.25	99.86	14.78	71.46	97.67	99.31
P_1	0.67	36.04	55.29	96.34	57.23	84.88	91.75	92.50
P_2	4.58	44.18	62.92	98.61	67.65	89.91	96.58	97.67
P_3	9.30	22.72	45.58	99.15	44.91	69.58	83.01	87.20
P_4	0.37	28.67	96.68	98.47	40.90	77.18	85.48	86.66
P_5	0.15	25.69	72.79	98.92	49.58	77.57	83.63	84.39
P_6	0.00	4.86	91.07	99.36	61.75	89.86	94.84	95.63
P_7	99.33	98.74	90.74	96.82	72.23	78.73	79.47	79.37
P_8	0.62	30.74	68.68	97.72	43.06	66.16	72.60	73.91
P_9	0.01	3.65	51.83	99.28	58.21	95.98	99.02	99.17
P_{10}	8.98	57.87	89.87	96.45	79.62	90.99	92.34	92.53
P_{11}	82.04	94.38	97.11	98.51	92.99	97.52	98.14	98.22
P_{12}	57.41	86.48	93.35	94.96	47.26	59.83	65.20	66.82
P_{13}	41.71	97.76	99.09	99.27	67.85	78.29	81.17	81.15
P_{14}	18.23	80.24	98.37	97.45	66.17	81.44	84.80	85.04
P_{15}	25.87	56.59	97.60	98.73	76.32	86.16	88.11	88.21
P_{16}	71.93	78.48	83.47	88.14	54.59	55.30	54.98	55.03
P_{17}	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

on the global attention model compared to the LSTM model. The classification performance of the LSTM model decreases drastically with the increase in noise levels to 20dB SNR. PQD classification using the LSTM model tested under 20dB SNR AWGN gives the WAcc of 18.99%. LSTM model classifiers are biased towards class P_7 -Oscillatory Transient and class P_{11} -Swell+Harmonics, in which both classes have high similarities with the nature of the noisy signal. The LSTM model without the 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 20-40dB SNR AWGN increased significantly to 51.33%, 78.37%, and 87.54% respectively, as shown in Table 3.8. The proposed global attention-based model is affected by the 20dB noise but still maintains the WAcc above 50%. This result proves that the attention mechanism has successfully highlighted the important features under noisy conditions. The added attention mechanism generalised the model by highlighting the signal with an attention vector.

Although the proposed attention mechanism shows better overall results, it is not deniable that it comes with some weaknesses. In Table 3.8, it can be noticed that class P_{16} -Sag+Oscillatory Transient and P_{17} -Swell+Oscillatory Transient encounter some difficulty in classification. It can be noticed that these two classes are very similar in nature, except for only magnitude differences. 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 the overall picture of the input signal, but insensitive to high frequencies and small magnitude changes.

3.5.2 Noisy Training Analysis

In this experiment, the models are trained with a dataset consisting of noiseless and 20-50dB SNR AWGN. The AWGN added into the training samples are learnt as a new type of feature posed on the samples. Temporal features of the targeted signal polluted with noise can still be learnt by the model during the training process. Subsequently, this trained model is tested with added 20-40dB 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 15dB 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.

The classification performance of both models with and without attention mechanism are compared in Table 3.9. Both models are set up using 100 units of hidden

Table 3.9: Classification performance of LSTM model trained with noisy synthetic PQD data and tested with 20-40dB AWGN and noiseless conditions.

Attention Class \ SNR	Without Attention				With Attention			
	20dB	30dB	40dB	noiseless	20dB	30dB	40dB	noiseless
P_0	92.14	98.85	99.64	99.75	55.05	97.46	99.32	99.41
P_1	84.94	94.35	96.77	97.26	71.85	87.06	89.00	88.99
P_2	89.59	95.62	97.13	97.45	82.08	94.60	96.35	96.47
P_3	67.09	83.48	91.16	94.21	37.54	40.24	41.29	41.55
P_4	66.77	88.25	94.04	94.79	52.19	75.05	78.06	78.44
P_5	78.98	97.74	98.47	98.42	51.47	65.20	67.45	67.79
P_6	88.99	98.29	99.08	99.25	75.67	89.60	91.51	91.87
P_7	93.96	96.07	96.22	96.18	74.11	77.48	77.90	77.94
P_8	76.22	97.02	97.80	97.80	47.70	61.80	63.10	63.67
P_9	94.10	99.41	99.82	99.81	87.05	97.77	98.55	98.69
P_{10}	87.56	94.10	95.42	95.59	84.76	90.85	91.58	91.64
P_{11}	96.69	98.01	98.13	98.07	95.04	97.20	97.44	97.47
P_{12}	67.64	82.81	91.55	95.06	67.28	78.64	81.63	82.15
P_{13}	86.82	97.11	98.17	98.16	71.56	77.31	78.06	78.07
P_{14}	66.54	84.47	92.14	93.49	62.60	69.94	70.06	69.94
P_{15}	72.75	89.54	95.02	95.49	76.64	81.51	81.83	81.54
P_{16}	81.17	86.46	86.54	86.46	44.03	44.22	44.43	44.53
P_{17}	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

units and trained for 30 epochs. Both models are showing better classification performance compared to models trained under noiseless conditions. The classification accuracy of the LSTM model without attention improved from the previous 18.99% to 84.87% WAcc under 20dB SNR AWGN. However, the classification performance of The global attention-based LSTM model does not show drastic improvement when trained under noisy conditions, from 51.33% to 66.67%. From Table 3.9, it can be noticed that

Table 3.10: Unseen noise performance comparison.

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

class P_{16} -Sag+Oscillatory Transient and P_{17} -Swell+Oscillatory Transient are still having similar difficulty in classification. By comparing to the previous train with noiseless experiment, it can also be noticed that class P_{12} -Interrupt+Harmonics shows improved performance but reduced performance in class P_3 -Interrupt. Training with added noise changes 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 is insensitive to slight magnitude changes.

Three unknown noises were used to test the trained model. The comparison result is tabulated in Table 3.10. Unknown Noise #A is an AWGN with 15dB SNR. Although the trained model can achieve good accuracy when training with SNR 20-50dB, 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 achieve 14.75% in the model trained without noise and without attention mechanism. Training with a mixture of noiseless and 20-50dB noise without attention only improves the accuracy to 47.86%, which is still far from the expected 84.87%. Comparatively, despite affecting by a higher level of noise, the model with an 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 offsets 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 conditions respectively. The proposed global attention-based LSTM model however achieved stable classification performances of 59.09% and 59.10% for training without and with noise respectively. These classification performances are close to the average classification performance of testing with 20dB SNR AWGN condition. This

result shows that the proposed attention model is still reliable with different variants of noise, as long as the noise level is 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 a uniformly distributed noise with an 11-30 dB SNR range. LSTM model without attention mechanism achieves 20.06% and 53.94% under noiseless training and noisy training respectively. The classification performance of the proposed global attention-based model achieves a consistent performance of 43.06% and 54.09% respectively. Overall, the proposed global attention mechanism has comparatively consistent, and better classification performance for unseen noise conditions.

3.6 Chapter Summary

A global attention-based LSTM model (GA-LSTM) is proposed in this chapter for more generalised PQD classification. A global attention mechanism is used in the presented method to highlight the important features present within the input signal. Temporal features of the highlighted signal are extracted using a single layer of LSTM. Two experiments were carried out to analyse the performance of the proposed GA-LSTM model. Three window sizes, i.e. quarter-period, half-period, and full-period, were performed to study the effect of sequence length in single disturbance PQD. As a result, window size having a full-period waveform with a higher number of sampling points can achieve better performance as compared to others under noisy conditions. With attention, it can obtain better feature selection even if the signal is in a noisy environment. On the other hand, the proposed GA-LSTM network was compared with LSTM using a single full-period PQD waveform. The proposed network is insusceptible to noise with 95.15% and 52.53% of classification rate under noiseless and noisy conditions respectively. As for the LSTM model, it is prone to noise with a significant dropping rate from 98.96% on the noiseless test to 20.52% under 20dB SNR AWGN.

The proposed GA-LSTM is also tested under noiseless and noisy conditions using combined-disturbance PQDs. As a result, GA-LSTM can achieve 66.67% and 51.33% accuracy when the network is trained with and without noise respectively. It shows that GA-LSTM has better noise immunity as compared to the LSTM model with 84.87% under noisy training data and 18.99% under noiseless training data. In addition, three

unseen noises are tested on GA-LSTM and LSTM models. Results show that the proposed GA-LSTM model has more generalised classification performance, where it can achieve averagely better classification accuracy on signals polluted with unknown noise. The presence of noise in training data can significantly affect the performance LSTM model because it requires observed data to be learned by the network. The introduction of the attention mechanism can highlight salient patterns occurred inside a PQD as a feature representation. However, the proposed GA-LSTM might be insensitive to conditions with short-time impulse and small magnitude changes. In the next chapter, multi-resolution global attention LSTM network is introduced to decompose time-domain signal into multiple levels of frequency bands to increase the signal resolution for salient feature extraction.

Chapter 4

Multi-resolution Global-attention Long Short-Term Memory Network

4.1 Introduction

Power quality [122, 123] is always the concern in the integration of high precision and sensitive equipment inside the power utility network. Power quality disturbance (PQD) is defined as a series of disruptions on the magnitude or frequency of the power supply sinusoidal waveform [124]. A PQD occurring in the power grid might cause problems such as reducing the lifespan of electrical devices, causing malfunctioning in sensitive electronic devices, causing unwanted power tripping, and financial losses. PQDs are among the major concerns in improving the functionality and efficiency of a microgrid [125]. PQD detection and classification is thus an important tool for monitoring the power quality of the power systems. The ability to identify the presence of PQDs in a system helps in guiding the energy management operation.

Detection and classification of PQDs can be categorised into three categories, knowledge-based method, model-based method, and hybrid method. The knowledge-based method relies heavily on professional knowledge in finding the most suitable statistical features for the classification process. Model-based method on the other hand involves the use of machine learning to perform feature extraction and classification automatically. Hybrid method extract features using the knowledge-based method,

while classifying the PQDs with the model-based method or machine learning. Hybrid-based PQD classification methods are studied in [66, 68, 82, 126, 127]. Most of these methods extract specific statistical features from the input signal before passing it into a machine learning algorithm for the classification process.

A time-series model-based method, global attention (GA) mechanism is proposed and discussed in Chapter 3, to improve the generalisation capability of the model-based Long Short-term Memory (LSTM) classifier. However, the result shows that the proposed time-series attention-based LSTM method is weaker in classifying high-frequency signals. Wavelet transform (WT) is used extensively in the field of noise suppression due to its properties in detecting discontinuity in the signal [128]. The properties of WT which allows multi-level signal decomposition (MSD) helps in increasing the resolution of the input signal into multi-level segments with different frequency components. In this chapter, WT is proposed to increase the input signal resolution by extracting the signal's features at different frequency bands. Multi-resolution Global-attention LSTM network is proposed to further improve the classification performance of the global attention LSTM (GA-LSTM). This chapter covers the analysis of the proposed multi-resolution global-attention LSTM network. Section 4.2 explains the theory of WT and MSD. Section 4.3 describes the approach used for the proposed methods. Section 4.4 presents the results of the proposed methods. Finally, Section 4.5 summarizes the findings and performance from the experiments.

4.2 Multi-level Signal Decomposition

Wavelet transforms [129] is proven to be efficient in detecting discontinuity in signals. The characteristics of varying window sizes in WT allow it to achieve an optimal time-frequency resolution. The varying window sizes also allow wavelets to detect non-stationary signals, which are posed by most of the PQDs. Wavelet transform can be expressed as [129],

$$f(x) = \sum_{i,j} a_{i,j} \psi_{i,j}(x), \quad (4.1)$$

where $a_{i,j}$ representing DWT expansion coefficients of input $f(x)$, with scaling and shifting parameters, i and j respectively. While $\psi_{i,j}$ represents the wavelet expansion function. The DWT coefficients can be expressed as,

$$a_{i,j} = \int_{-\infty}^{\infty} f(x) \psi_{(i,j)}(x). \quad (4.2)$$

Wavelet basis function can be generated from mother wavelet, $\psi_{i,j}$ by tuning the scaling and shifting parameters, i and j as follows,

$$\psi_{(i,j)}(x) = 2^{-i/2}\psi(2^{-i}x - j), \quad (4.3)$$

MSD can be achieved by providing two-scaled equation, $\psi_{i,j}$ and $\phi_{i,j}$

$$\psi_{i,j}(x) = 2^{-i/2}h(j)\psi(2^{-i}x - j), \quad (4.4)$$

$$\phi_{i,j}(x) = 2^{-i/2}g(j)\phi(2^{-i}x - j), \quad (4.5)$$

where $g(j) = (-1)^j h(1-j)$. $h(j)$ and $g(j)$ can be viewed as high pass and loss pass filter coefficients. MSD is achieved with different I^{th} level decomposition as follows [129],

$$f_i(x) = \sum_{i=1}^I a_{i,j}\psi_{i,j}(x) + \sum_{i=1}^I a_{i,j}\phi_{i,j}(x), \quad (4.6)$$

The function of MSD allows the signal to be band-filtered into multiple levels of frequencies. The original signal is filtered using a high pass filter (HPF) for high-frequency components while a low pass filter (LPF) is used to extract low-frequency components. The band LPF from each level are used as input to the next level decomposition until the desired decomposition level. Fig. 4.1 shows MSD with 4-levels of decomposition. The detailed coefficients, D1 to D4 are outputs from the high pass filter from respective decomposition levels, while A4 is the low pass filter output from the last decomposition level.

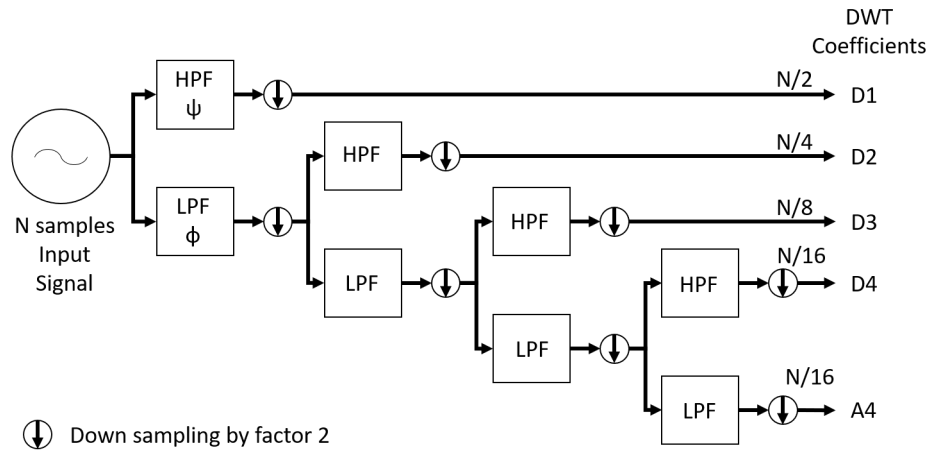


Figure 4.1: MSD with 4-levels of decomposition.

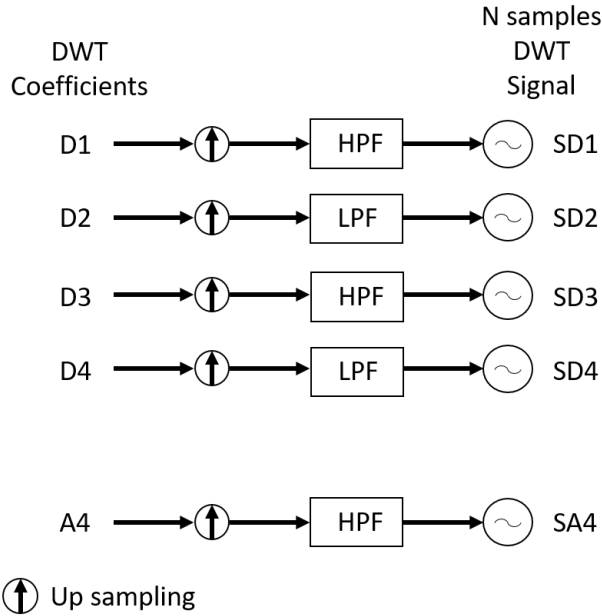


Figure 4.2: Single stage reconstruction.

Single-stage reconstruction performed on each of the detail and approximate coefficients allows retrieval of the original time-series signal composition at respective decomposition levels. The wavelet signal reconstruction is shown in Fig. 4.2, where the detailed coefficient and approximation coefficient of MSD in every decomposition level are reconstructed into 5 signal waveforms at different frequency bands, SD1 to SD4, and SA4.

4.3 Multi-resolution Attention Models

Automatic feature extraction is a crucial component in achieving automatic PQD detection and classification. Global attention is introduced to better generalize the network towards detecting multiple classes of disturbances [130]. Although the GA-LSTM network is proven to have better noise immunity, the network faces difficulty in detecting high-frequency disturbances with small-magnitude changes. Multi-resolution signal decomposition (MSD) is used to transform the time-series input signals into multiple frequency components or time-frequency components [131]. MSD is proposed along with an attention mechanism to achieve better classification performance. In this chapter, two different approaches using either inverse wavelet transform or wavelet coefficients

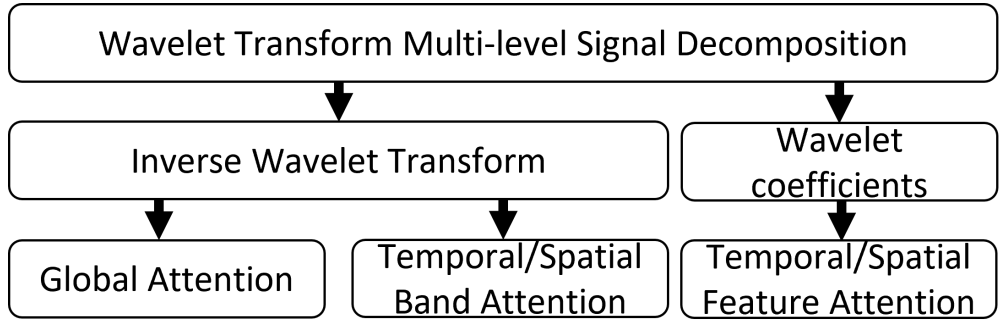


Figure 4.3: Multi-resolution Attention LSTM.

are conducted as shown in Fig. 4.3. The first approach or Model #A uses inverse wavelet transform (IWT) on the wavelet coefficients to reconstruct every frequency band output signal into time-series signals on their respective frequency band. Two types of attention mechanisms are studied, global attention and band attention. The details of the model are explained in Section 4.3.1. The second approach or Model #B remove the need for IWT. The output of MSD is first aligned using a feature align layer, where the unequal-length MSD outputs are converted into equal-length latent representations. Spatial and temporal feature attention are applied to the latent representation to improve the generalisation capability of the network. Model #B is discussed in detail in Section 4.3.2.

4.3.1 Model #A: Single-period Multi-Resolution Attention LSTM

The attention mechanism highlights important characteristics present in the signal. Self-attention is achieved by multiplying attention weight with the input signals. LSTM is good at extracting sequential information from the input data. In [130], it is

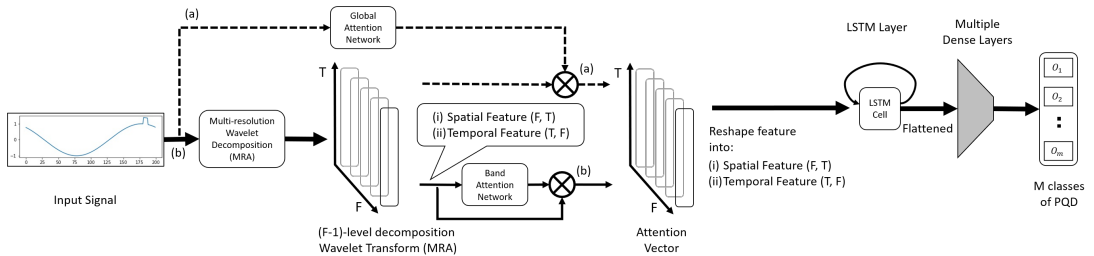


Figure 4.4: Multi-resolution attention LSTM model using (a) Global attention, (b) Band attention.

found that the GA-LSTM network improved the generalization capability compared to the pure LSTM network. However, the decreased performance in noiseless conditions shows that the GA-LSTM network is weak in identifying higher-frequency disturbances.

In this model, multi-resolution attention LSTM PQD classifier is proposed as shown in Fig. 4.4. Daubechies 4 (db4) wavelet filter function is used for the decomposition. Mother wavelet db4 is used as it provides good performance with low computation cost [132]. Besides, db4 provides good time resolution and accurate transient faults analysis. Literature shows that db4 with 4 levels of decomposition is a good starting point for the research model's design [48, 63, 68, 133]. A total of five decomposed wavelet coefficients with unique frequency ranges are yielded for further processing. IWT is applied on the yielded wavelet coefficients to reconstruct the input signal into five time-series signals at different frequency bands.

To achieve the best feature extraction and classification performance, a series of feature manipulation are tested. Two different attention mechanisms are proposed for analysis, global attention (GA) and band attention (BA) as shown in path (a) and path (b) in Fig. 4.4. GA is applied in path (a), where the attention score is harvested from the input signal and then multiplied in element-wise operation for each of the five reconstructed band signals. Path (b) on the other hand applies BA, where self-attention is applied to each of the five reconstructed band signals. The attention feature outputs are then stacked into temporal (T,F) or spatial (F,T) matrices. The matrices are fed into the LSTM layer for higher dimensional feature extraction. The output of the LSTM layer is the representation of the encoded sequence features according to the input pattern.

4.3.2 Model #B: Multi-period Multi-resolution Attention LSTM

As for Model #B, instead of using reconstructed signals at different frequency bands as shown in Model #A, a feature align layer is proposed to embed and align

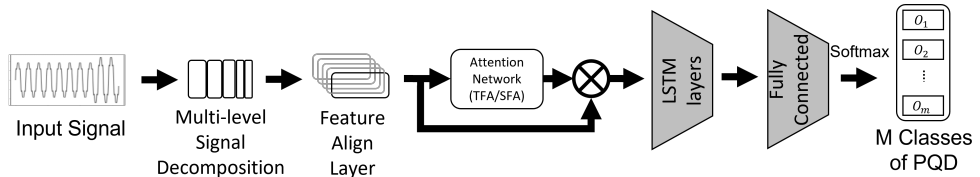


Figure 4.5: Multi-resolution Attention LSTM model with attention mechanism.

the wavelet transformation coefficients using a perceptron layer. An overview of the proposed method is depicted in Fig. 4.5. The output coefficients of MSD are first passed through a feature align layer which embeds different bands into similar dimensions. Feature attention, or the attention mechanism applied on the embedded features is used to improve the generalization capability of the network. Two different attention mechanisms are tested, temporal feature attention (TFA) and Spatial feature attention (SFA). TFA is multiplied element-wise with the temporally aligned features from feature align layer outputs to get the temporal attention vector. While spatial attention vector is acquired by multiplying the spatially aligned features with the SFA attention weights. The highlighted features are then passed into LSTM layers for feature extraction. Finally, two fully connected layers and a softmax activation function are used to classify the features into respective disturbance classes.

4.4 Experiment Setup

All experiments are carried out using AMD Ryzen 7 3800X 8-Core Processor with Nvidia P6000 graphic processing unit. Pytorch framework is used for the experiments. The MSD used in both experiments are based on the db4 wavelet for the wavelet decomposition process. Additive white Gaussian noise (AWGN) is introduced to test the network classification performance. Experiment #1 is trained with noiseless datasets, while Experiment #2 is trained with a random 20-50dB SNR of AWGN. The testing performance on the other hand is tested AWGN with SNR of 20dB, 30dB, 40dB, 50dB, and noiseless conditions. The main evaluation matrix used in the experiments is 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,

$$Acc_n = \frac{TP_n}{\sum_{j=0}^m S_j}. \quad (4.7)$$

Two different models are proposed in this chapter. While both models shared the same concept of multi-resolution attention LSTM, there are slight differences within the experiment's setup. Each model's experiment setup will be discussed in its respective section.

4.4.1 Experiment #1 Single-period Multi-Resolution Attention LSTM

Experiment #1 is set up with the generation of 3-period PQD signals, and the single-period windowing method is used to pre-process the input data into a single-period signal waveform. A total of 200 randomized parameter samples of the three-period waveform are generated for each of the 18 PQD classes listed in Table 4.1. The sampling frequency used is 10kHz. 70% of the total generated samples are used as training data, while the remaining 30% are used as validation samples. The use of the train-to-test ratio of 70:30 follows best practices in machine learning process to prevent overfitting. As this sampling method caused an imbalance in the counts of class, where weight balancing has to be done in the training process, and weighted accuracy is used to accurately projects the classification performance. Weighted accuracy (WAcc) can be calculated by multiplying the individual class accuracy with its weight, which can be depicted as,

$$WAcc = \sum_{j=0}^m W_j \times Acc_j. \quad (4.8)$$

In Table 4.2, a total of six different configurations are analysed to obtain the best classification performance. GA network highlights the important feature presented in the raw input signal, while the BA network highlights the frequency bands output of MSD. The performances of four different BA networks are also compared in the

Table 4.1: Class of power quality disturbances (Experiment #1).

Label	Class Description	Label	Class Description
P_0	Normal	P_9	Flicker
P_1	Sag	P_{10}	Sag+Harmonics
P_2	Swell	P_{11}	Swell+Harmonics
P_3	Interrupt	P_{12}	Interrupt+Harmonics
P_4	Impulse Transient	P_{13}	Harmonics+Notch
P_5	Spike	P_{14}	Sag+Transient
P_6	Harmonics	P_{15}	Swell+Transient
P_7	Oscillatory Transient	P_{16}	Sag+Oscillatory Transient
P_8	Notch	P_{17}	Swell+Oscillatory Transient

Table 4.2: Proposed model under setups with global (GA) and band attention (BA), with spatial (S) and temporal (T) features settings.

Model	Attention	Attention input	LSTM input
#A1	GA	T	S
#A2	GA	T	T
#A3	BA	S	S
#A4	BA	S	T
#A5	BA	T	S
#A6	BA	T	T

experiment. In the proposed method, LSTM is used to extract spatial and temporal features by feeding different input orientations into the LSTM layer. Specifically, the attended MSD frequency bands or attention vector in this context is aligned into (T,F) and (F,T) shapes for temporal sequence feature extraction and spatial sequence feature extraction respectively.

4.4.2 Experiment #2 Multi-period Multi-resolution Attention LSTM

In Experiment #2, 10-period PQD signals are generated and used as input data. The classification and generalization capability of the proposed method is tested with 16 synthetic PQD models which can be benchmarked with [20, 101]. The generated PQDs include normal signal waveform, single disturbance waveform, and multiple disturbance waveform as listed in Table 4.3. A total of 76800 10-period PQD samples are randomly generated, where each disturbance class has 4800 samples. The sampling frequency used is 3200 Hz. As noise is always present during real-world data collection, 20-50dB SNR AWGN are added randomly into the generated training samples. 10-fold cross-validation is used, which consists of 90% training samples and 10% of total samples are used as validation samples. A total of 5 sets of testing data sets are generated for model benchmarking purposes. The 5 sets of the testing data include a set of noiseless samples, and 20dB, 30dB, 40dB, and 50dB SNR AWGN added samples. Each set of testing data consists of 1000 samples per PQD class. This experimental setup is aligned with Machlev et al. [101] and Wang et al. [20] to carry out the benchmarking process

Table 4.3: Class of power quality disturbances (Experiment #2).

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

Table 4.4: Details of different bench-marking models.

Model name	Model details
Deep CNN (DNN) [20]	1D CNN layer: 6; Dense: 3 layers: 256, 128, 16
Deep LSTM (DNN) [101]	3 layers, unit: 32; Dense: 2 layers: 256, 16
WT-LSTM (Hybrid)	MSD: 4 level db4; Dense layer temporal embedding; LSTM: 3 layers, unit:32; Dense: 2 layers: 256, 16
WT-TFA-LSTM (Hybrid)	MSD: 4 level db4; Dense layer temporal embedding; Attention mechanism; LSTM: 3 layers, unit:32; Dense: 2 layers: 256, 16
WT-SFA-LSTM (Hybrid)	MSD: 4 level db4; Dense layer spatial embedding; Attention mechanism; LSTM: 3 layers, unit: 32; Dense: 2 layers: 256, 16

fairly. In this experiment, two types of input arrangement are used to evaluate the proposed multi-resolution attention LSTM model. The features are arranged in either TFA or SFA as described in Section 4.3.2. Bench-marking of the proposed method has also be done with multi-resolution LSTM model without attention mechanism, deep LSTM model[101], and deep convolution neural network (Deep CNN) model[20, 101]. The details of each model’s comparison are summarised in Table 4.4.

4.5 Experiment #1: Performance Analysis of Single-period Multi-resolution Attention LSTM

The classification performance of the proposed multi-resolution attention LSTM model is analysed for evaluation. There are a total of six different setups tested on this model as shown in Table 4.2. In the GA models, the attention mechanism is applied to the input signal directly. The attention output is then multiplied element-wise with the input signal. The output attention vector is shaped into (F, T) for spatial sequence extraction in Model #A1 and (T, F) for temporal sequence extraction in Model #A2. In BA models, the attention mechanism is applied to the MSD output vector, in either (F, T) shape for spatial band attention, or (T, F) shape for temporal band attention. The output attention vector is then passed into the LSTM layer for spatial sequence or temporal sequence feature extraction.

4.5.1 Overall Classification Accuracy Comparison

The classification performance of the proposed method is tabulated in Table 4.5. As for an overall comparison, Model #A1 and Model #A2 using the GA mechanism show better performance with $> 40\%$. The GA layer highlights an important part of the raw input signal before passing the attention vector to the MSD layer. The main difference between Model #A1 and #A2 is in the LSTM layer which takes in MSD output

Table 4.5: Weighted accuracy of the proposed model under setups with global and band attention, with spatial and temporal features settings and compare with Model #A7 ([134]).

Model	20dB	30dB	40dB	50dB	No-noise
#A1	50.71	80.31	90.62	91.40	91.43
#A2	55.31	80.44	89.28	89.91	89.99
#A3	35.44	68.32	90.00	93.93	93.53
#A4	21.21	49.29	73.00	92.32	93.12
#A5	42.42	75.33	86.95	87.83	87.68
#A6	43.54	78.23	89.66	90.83	90.94
#A7[134]	18.99	38.54	62.53	81.07	97.69

with different orientations as explained in the setup section. The LSTM in Model #A1 takes in MRA output which is shaped into a spatial orientation of (F, T), thus producing spatial sequence feature. On the other hand, LSTM in Model #A2 takes in MSD output which is shaped into a temporal orientation of (T, F), thus giving temporal sequence as the output feature. As a result, both models can achieve similar classification accuracy with Model #A2 having slightly better classification on the 20dB AWGN testing data, 55.31% versus 50.71%. However, it is also worth mentioning that the LSTM layer in Model #A1 requires less computation, or performs only $(1/(T/F)) \times$ recurrent steps compared to Model #A2. The recurrent steps in Model #A1 is depending on the spatial dimension, F of the attention vector (F, T), while in Model #A2 recurrent steps occur in the temporal dimension of the attention vector (T, F). The T and F used in this experiment is 200 and 5 respectively, thus Model #A1 has $(1/40) \times$ recurrent steps compared to Model #A2. Time-series LSTM model [134] is also compared as Model #A7 in Table 4.5. The time-series LSTM model without attention mechanism is highly affected by noise, which gives only 18.99% under the 20dB AWGN test.

BA models are models containing attention mechanisms on the output bands of the MSD. Band spatial attention and band temporal attention are applied to the spatial (F,T) and temporal (T,F) orientation of the MSD output. Comparatively, band spatial attention (Model #A3 and #A4) resulted in lower classification accuracy on high noise data, but the classification accuracy on lower noise data is slightly better than all other models. On the other hand, band temporal attention models (Model #A5 and #A6) have better classification accuracy $> 40\%$ on high noise signal with 20dB SNR AWGN. However, the use of GA is proven to be having better classification performance in this multiple bands time-series classification.

4.5.2 Individual Class Analysis

A direct comparison of Model #A1 in Table 4.6a and Model #A2 in Table 4.6b demonstrate distinctive performance in classification, where Model #A1 has weaker classification performance on class P_{12} , P_{16} and P_{17} , while Model #A2 has weaker classification performance on more classes, P_8 , P_{12} , P_{13} , P_{16} , and P_{17} . Both Model #A1 and #A2 perform weaker classification result on class P_{12} , P_{16} , and P_{17} . This shows that these two models have deficiencies in classifying these classes. Referring to the P_{12} -Interrupt+Harmonics confusion matrix, it is noticed that the model is confused between class interrupt. The harmonics present in the P_{12} class has significantly lower magnitude

Table 4.6: Classification performance of Model #A1 and Model #A2 with additive white Gaussian noise from 20dB-50dB and noiseless conditions.

(a) Model #A1.						(b) Model #A2.					
SNR \ Class	20dB	30dB	40dB	50dB	noiseless	SNR \ Class	20dB	30dB	40dB	50dB	noiseless
P_0	10.80	67.49	98.53	99.74	99.75	P_0	12.89	69.35	97.98	99.45	99.50
P_1	46.18	81.14	91.09	91.58	91.40	P_1	59.06	86.85	93.53	93.97	94.18
P_2	58.61	86.04	94.94	96.25	96.57	P_2	66.29	89.70	96.13	96.79	96.97
P_3	49.77	77.88	89.87	92.84	93.10	P_3	58.71	79.57	89.53	91.53	91.71
P_4	30.72	75.49	85.50	86.17	85.94	P_4	54.87	83.91	90.32	90.50	90.66
P_5	33.85	85.06	91.71	92.46	92.39	P_5	50.70	83.51	87.57	87.76	87.90
P_6	65.80	93.19	97.63	98.17	98.08	P_6	77.79	94.35	96.57	96.54	96.52
P_7	78.79	87.29	88.05	87.96	87.96	P_7	78.11	84.34	85.29	85.38	85.38
P_8	52.28	80.00	84.05	84.66	84.58	P_8	61.34	74.53	76.52	76.94	76.86
P_9	57.00	99.04	99.90	99.92	99.91	P_9	63.05	97.06	99.50	99.64	99.66
P_{10}	80.14	91.96	93.54	93.72	93.80	P_{10}	83.58	92.01	92.62	92.67	92.67
P_{11}	92.18	97.73	98.27	98.31	98.31	P_{11}	89.54	97.06	97.72	97.86	97.89
P_{12}	59.61	70.23	75.67	77.06	77.26	P_{12}	59.52	67.46	72.07	72.99	73.27
P_{13}	69.16	85.81	87.20	87.20	87.20	P_{13}	53.21	71.23	74.69	75.56	75.75
P_{14}	73.96	85.04	86.60	86.83	87.19	P_{14}	72.15	83.22	85.25	85.71	85.69
P_{15}	79.97	92.15	94.24	94.63	94.79	P_{15}	79.19	87.98	89.42	89.42	89.34
P_{16}	62.05	62.57	62.92	62.91	62.80	P_{16}	72.86	72.15	72.06	71.90	72.01
P_{17}	44.92	44.61	44.80	45.01	44.89	P_{17}	25.82	25.59	25.38	25.38	25.32
WAcc	50.71	80.31	90.62	91.4	91.43	WAcc	55.31	80.44	89.28	89.91	89.99

changes compared to the magnitude changes of interrupt signal to normal signal. P_{16} -Sag+Oscillatory Transient and P_{17} -Swell+Oscillatory Transient respectively. Referring to the confusion matrix, it is noticed that the model is confused among P_{16} and P_{17} . This can be explained by the nature of an oscillatory transient with varying changes of the magnitude to the original signal. These two findings show that the latent feature learnt by GA models is not sensitive towards small changes of magnitude within similar signals.

Compared to Model #A1, Model #A2 has poor classification performance for class P_8 -Notch and P_{13} -Harmonics+Notch, which are disturbance class notch and harmonics with notch. Referring to the confusion matrix, it is noticed that the weak performance for class P_8 is mainly caused by confusion with class P_4 -Impulsive Tran-

sient. The main difference between these two disturbance classes is that P_8 has a shorter disturbance period. The performance impact of Class P_4 on class P_8 is less in Model #A1 indicating that LSTM spatial sequence feature is better at locating sharp changes in the signal magnitude as compared to temporal sequence features. Referring to the confusion matrix for class P_{13} , it can be noticed that both Model #A1 and Model #A2 are confused between P_{13} with P_6 -Harmonics, with slightly better accuracy on Model #A1. This can be explained similarly with the case of P_8 , where Model #A1 with spatial sequence feature extraction performs better at recognizing sharper gradient change of notch signal compared to harmonics signal.

4.6 Experiment #2: Performance Analysis of Multi-period Multi-resolution Attention LSTM

In this experiment #2, 10-period PQD signals sampled with a 3200kHz sampling rate are used as input into the classification models. A total of 16 classes of PQDs are trained and analysed. Wavelet-based attention LSTM model is proposed. Two model arrangements are analysed, i.e. Model #B1: wavelet-based temporal-aligned feature attention LSTM (WT-TFA-LSTM), and Model #B2: wavelet-based spatial-aligned feature attention LSTM (WT-SFA-LSTM). The feature aligns layer in the WT-TFA-LSTM model aligns the feature into a temporal arrangement, where the attention mechanism is applied across the data sequence. On the other hand, WT-SFA-LSTM has spatial feature attention. Spatial feature attention applies attention over different frequency bands.

4.6.1 Overall Classification Accuracy Analysis

The classification performance of the proposed methods are tabulated in Table 4.7. The performance comparison of Deep LSTM and WT-LSTM shows WT improved the performance across different noise levels. It demonstrates that using WT MSD without attention mechanism may lead to a positive increment in classification performance. There are two types of attention mechanism alignment introduced, temporal feature alignment attention in WT-TFA-LSTM and spatial feature alignment attention in WT-SFA-LSTM. The introduction of the attention mechanism into the WT-LSTM model decreases the classification performance significantly across different

Table 4.7: Classification performance comparisons.

Model	20dB	30dB	40dB	50dB	0dB
Deep CNN [20]	90.56	97.69	98.97	99.01	99.57
Deep LSTM [101]	88.48	96.64	97.83	98.16	98.54
WT-LSTM	89.77	97.40	98.32	98.79	99.21
WT-TFA-LSTM	79.49	92.28	94.48	94.87	95.09
WT-SFA-LSTM	93.79	97.49	98.34	98.51	99.09

noise levels. Self-attention mechanism applied across the temporal feature encoding of the feature align layer. An attention mechanism that is focusing on a single time frame might have caused the information to diminish in other time frames. The introduction of single attention might not sufficient to represent various types of disturbance (fast disturbance and slow disturbance). On the other hand, when spatial feature attention is introduced in WT-SFA-LSTM, The classification shows promising improvement and achieved a classification accuracy of 93.79% on high noise 20dB SNR. Spatial attention applied across different frequency feature bands successfully highlighted the differences of differences disturbances and helps improve the classification performance even under high noise conditions.

4.6.2 Classification Accuracy Analysis

Individual class accuracy of the proposed WT-TFA-LSTM and WT-SFA-LSTM models are tabulated in Table 4.8a and Table 4.8b respectively. WT-TFA-LSTM in Table 4.8a shows significant weaker classification on P2-Swell, P10-Sag+Harmonics, P14-Flicker+Sag , and P15-Flicker+Swell. It can be noticed that all these classes are of slow disturbance class. By employing the temporal attention mechanism, the emphasis is placed on smaller segments of the signal, thereby diminishing its capability to classify slow disturbances that can only be detected across a wider region of the signal. In contrast, it can be noticed that WT-SFA-LSTM performed significantly better over all classes and noise levels. 20dB SNR individual accuracy shows that class P0-Normal, P9-flicker, P10-Sag+Harmonics have individual accuracy less than 90%. This shows that the model is still being affected by noise on the slow disturbance class, but only

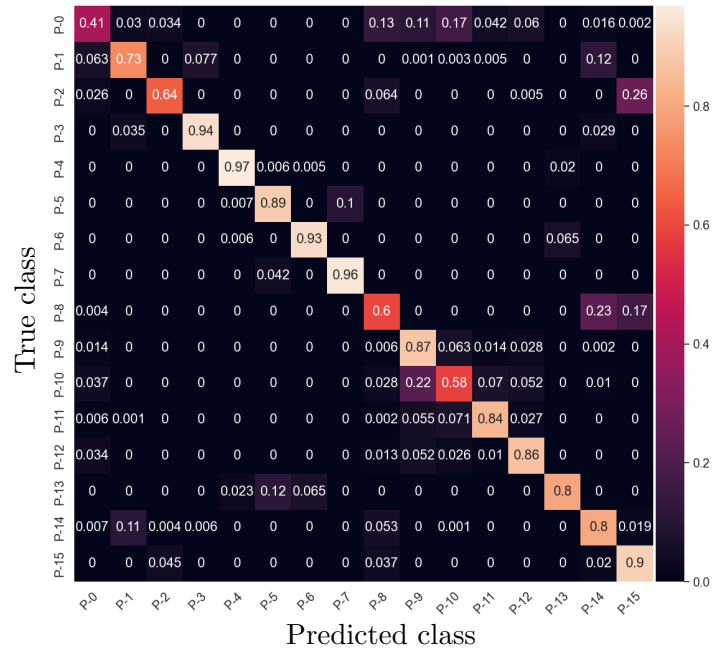
Table 4.8: Performance of the proposed WT-SFA-LSTM model tested with 20-50dB AWGN and noiseless conditions.

(a) Model #B1: WT-TFA-LSTM.						(b) Model #B2: WT-SFA-LSTM.					
SNR \ Class	20dB	30dB	40dB	50dB	noiseless	SNR \ Class	20dB	30dB	40dB	50dB	noiseless
P0	40.80	81.90	91.80	94.40	93.44	P0	69.40	84.80	92.30	96.20	97.47
P1	73.30	90.00	94.00	94.30	94.70	P1	93.80	96.90	97.10	96.00	96.60
P2	64.30	76.40	84.10	87.10	86.50	P2	97.20	98.70	98.80	98.30	98.60
P3	93.60	95.30	96.10	95.50	96.10	P3	99.00	99.40	99.70	99.10	99.70
P4	96.90	100.0	100.0	100.0	100.0	P4	99.60	100.0	100.0	100.0	100.0
P5	89.00	94.10	94.90	95.20	95.00	P5	97.40	98.30	98.00	97.80	98.70
P6	92.90	98.50	98.10	98.30	98.00	P6	99.70	100.0	100.0	100.0	99.90
P7	95.80	97.30	96.90	96.30	96.90	P7	98.10	99.30	99.60	99.20	99.60
P8	59.80	89.20	99.30	100.0	100.0	P8	94.60	100.0	100.0	100.0	100.0
P9	87.30	97.10	99.50	99.60	99.50	P9	86.80	97.30	99.10	99.20	99.30
P10	58.10	89.30	89.40	89.10	91.45	P10	84.70	92.20	96.30	98.30	99.59
P11	83.80	96.90	98.10	97.20	100.0	P11	90.90	97.60	98.20	97.20	100.0
P12	86.50	97.00	96.80	97.40	100.0	P12	93.10	97.00	96.80	97.40	100.0
P13	79.60	99.20	100.0	100.0	100.0	P13	99.80	100.0	100.0	100.0	100.0
P14	80.30	83.40	81.40	80.50	81.10	P14	98.50	98.90	98.60	98.10	97.60
P15	89.80	90.90	91.20	93.00	89.00	P15	98.00	99.40	99.00	99.30	98.60
Acc	79.49	92.28	94.48	94.87	95.09	Acc	93.79	97.49	98.34	98.51	99.09

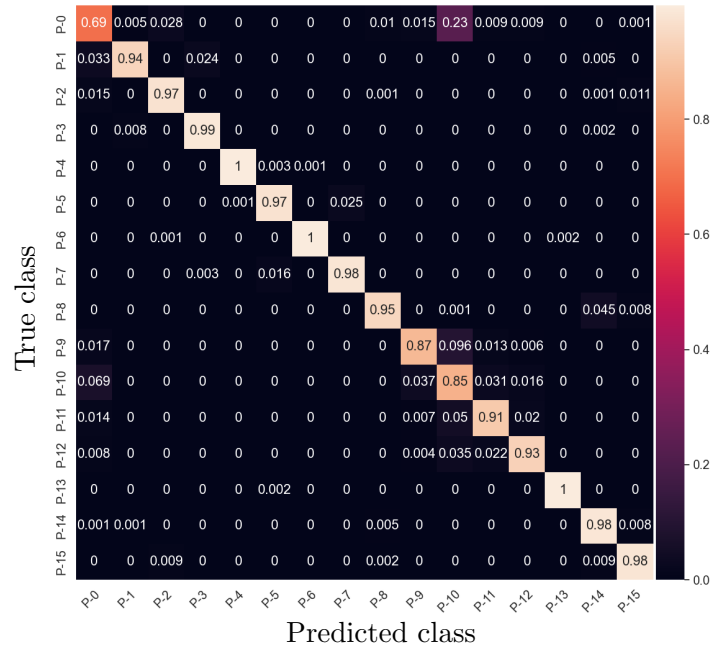
on high noise conditions. High noise conditions may lead to confusion with harmonic disturbance in a classifier model.

4.6.3 Confusion Matrix Analysis

A confusion matrix is used to visualise the in-depth classification performance of the model. Extreme cases of 20dB SNR AWGN are analysed in this section. In Fig. 4.6a, it can be noticed that class P0-Normal, P1-Sag, P2-Swell, P8-Notch and P10-Sag+Harmonics are severely impacted in the confusion matrix with value $< 80\%$. Slow disturbance classes such as P1, P2, and P10 are mostly not detected well due to the temporal attention mechanism which highlights a smaller part of the signals. Due to that P1, P2, and P10 are slow disturbance classes, it shows that WT-TFA-LSTM does not recognize well in fast disturbance class P8-Notch under noisy conditions. In addition, the introduction of SNR AWGN noises may easily be neutralised with the notching



(a)



(b)

Figure 4.6: Confusion matrix at 20dB SNR AWGN test on (a) Model #B1: WT-TFA-LSTM and (b) Model #B2: WT-SFA-LSTM.

effect as its magnitude has decreased with high frequency. Hence, the confusion between P14-Flicker+Sag and P15-Flicker+Swell are salient due to their similarity. The similarity of class P8 with class P14 and P15 can be explained by the neutralised notching effect, while the classifier performs poorly at slow disturbance classes. High noise levels cause confusion in the class of flicker as flicker signal has the characteristic of random oscillation. Other observations in the confusion matrix are that the confusion may occur between classes P1 with P14, P2 with P15, and P10 with P9-Flicker. Class P0-Normal is also confused mostly with class P8-Notch, P9-Flicker, and P10-Sag+harmonics due to the high level of noise disrupting the normal class signal waveform.

4.6.4 Model Complexity Analysis and Benchmarking

WT-SFA-LSTM model tested with 20dB SNR AWGN is shown in Fig. 4.6b. While this model performed very well on lower noise levels (30dB and above), the classification performances are still reduced on 20dB SNR conditions as shown in the confusion matrix. It can be noticed that class P0-Normal, P9-Flicker, and P10 Sag+Harmonic have most confusion (< 90%). It can be noticed that all of these classes are slow disturbance classes, and are easily polluted by noise introduced. There is only a slight difference between the normal class signals and slow disturbance class signals such as sag or swell and flicker. A slightly higher noise introduced could easily disrupt the signal into a different class. From the confusion matrix, class P9 and class P10 have mutual confusion. This can be explained with random High noise introduced altered the magnitude of the signals across the signals. On the other hand, P0 has high confusion

Table 4.9: Model complexity comparisons.

Model	Best 20dB Acc	Time per epoch	No. of parameters	Model size
Deep CNN [20]	90.56	32s	164464	0.657 MB
Deep LSTM [101]	88.48	68s	33936	0.136 MB
WT-LSTM	89.77	34s	206224	0.811 MB
WT-TFA-LSTM	79.49	34s	255120	1.002 MB
WT-SFA-LSTM	93.79	34s	221486	0.871 MB

with class P10 as well. This shows that the high level of noise introduced resembles a harmonics condition.

The complexity and model size for the proposed models are compared to the state-of-the-art Deep CNN, and Deep LSTM models [101] as tabulated in Table 4.9. A similar dataset is used in the training and testing of the models. From the comparison, it can be noticed that the proposed WT-SFA-LSTM model has the highest classification accuracy of 93.79% on the high noise 20dB SNR AWGN test. Compared to the Deep CNN model shows that the WT-SFA-LSTM model performed better with an improved accuracy of a 3% in noisy conditions. This increment performance requires only 32% increment in the model size or 0.214MB which is significantly low. Although the model size for the WT-SFA-LSTM model is much larger compared to that of the Deep LSTM model, the training time required is halved. This shows that the WT-SFA-LSTM model is a better choice for implementation if the computation resources are sufficient. Deep LSTM would be the best choice for computing resource limitation, but the training time is doubled compared to the other methods.

4.7 Chapter Summary

Multi-resolution attention mechanism with LSTM is proposed in this chapter for automatic PQD detection and classification with increasing input resolution. In order to obtain multi-resolution, WT is performed to decompose the signal into five bands, By applying IWT, these five bands are converted back into time-series signals carrying unique frequency information. Subsequently, this frequency information is fed into a global attention mechanism for feature highlighting. As a result, the combination use of a spatial arrangement of the reconstructed signals with GA performs better classification performance with 50.71% as compared to the LSTM model with 18.99% when tested with 20dB SNR AWGN. However, the proposed spatial MSD LSTM might be insensitive to conditions with small-magnitude changes. On the other hand, multi-resolution spatial attention LSTM (WT-SFA-LSTM) is also proposed to classify multi-period PQD signals. Instead of using IWT, the decomposed wavelet coefficients are aligned into the same length feature representation via a single layer of perceptron. Spatial attention features between frequency components are highlighted with a spatial attention mechanism before passing them into LSTM for higher-order feature extraction. Results show

major improvement in the classification accuracy with 93.79% on WT-SFA-LSTM versus 79.49% on WT-TFA-LSTM, 89.77% on WT-LSTM, and 88.48% on deep LSTM. WT-SFA-LSTM achieves major improvements in combined and slow disturbances. The computation complexity analysis shows that the training time required on WT-SFA-LSTM is halved compared to Deep LSTM. However, the model size of WT-SFA-LSTM is 32% higher than state of the art Deep CNN model. The complexity of the model is further optimized in the next chapter with the introduction of a transformer network to replace LSTM.

Chapter 5

Multi-resolution Convolutional Transformer Network

5.1 Introduction

Distributed energy resources [135] such as renewable energy generators and advanced electronic controllers are widely applied in modern power networks to improve power transmission and security. Due to the increase in network complexity, microgrid [136, 137] is formed to improve the synchronization between the monitoring and control of multiple distributed energy resources and loads. However, the monitoring of power quality disturbances (PQD) becomes tedious when there are multiple intermittent resources such as renewable energy sources and the introduction of plugged-in hybrid electric vehicles or energy storage systems in the power grid [138]. The occurrence of single or multiple disturbances at one instance is not easily picked up with human intervention. Some common examples of PQDs include sag, swell, interrupt, harmonics, transients, spike, notch, oscillatory transient, flicker, combined Sag with harmonics, flicker with harmonics, and flicker with swell. Hence, real-time PQD identification and classification are required to mitigate the PQ issues.

Deep learning is applied in the field of PQD detection and classification to achieve automatic feature extraction and classification [20, 139]. Long Short-term Memory (LSTM) network is studied extensively due to its ability in extracting information from sequence-structured PQD signals [140]. Recently, the transformer network is introduced as a superior architecture in the field of natural language processing [78] and automatic

speech recognition [141]. Besides excellent performance in extracting sequential information or long-term dependencies, the transformer network allows parallel processing, which reduced the training time required significantly as compared to LSTM. Multi-head attention mechanism (MHA) of the transformer network allows the highlighting of multiple features from the series of inputs.

A novel multi-resolution attention network using a convolutional transformer is proposed in this chapter to improve the PQD classification. A transformer network is used in extracting salient sequential features, replacing the sequentially trained LSTM network. The performance of the transformer network and LSTM are compared in terms of their accuracy and computational cost. The combination use of multi-level signal decomposition (MSD) with a transformer network is studied for classification performance comparison. This chapter is organised as follows: Section 5.2 discusses the background knowledge of transformer networks, as well as the new mechanism proposed in this chapter. Section 5.3 layout the setups for the experiments carried out in Section 5.4 and Section 5.5. Finally, Section 5.6 summarises the finding from this chapter.

5.2 Multi-resolution Attention Using Convolutional Transformer

A hybrid model composing of multi-level signal decomposition with a layer of convolutional kernel and transformer mechanism is proposed in Fig. 5.1. Details of the components in the proposed architecture are explained in the following subsections.

5.2.1 Multi-resolution Signal Decomposition

Wavelet transform (WT) is used to transform the time-series signal into time-frequency coefficients. MSD via discrete wavelet transform is used by performing wavelet

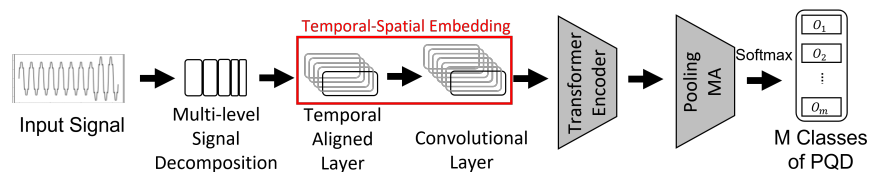


Figure 5.1: Proposed wavelet-based convolutional transformer hybrid model.

decomposition multiple times along the lower frequency components. MSD increases signal resolution into multiple time-frequency components representation, which allows better detection of different types of PQDs. MSD allows the use of a single type of mother wavelet in detecting disturbances occurring at different frequency ranges. As proven in Chapter 4, the application of MSD can improve classification performance. Daubechies with four filter coefficients or db4 mother wavelet are used for the decomposition as db4 is good at detecting short and fast transient [142]. A higher decomposition level of MSD, i.e., four decomposition levels allows the extraction of signal frequency components into five different representations with unique frequency ranges. The use of four levels is suggested by literature as it provides better classification performance [63, 68]. Besides, four levels of MSD allow decomposition into five coefficients with frequency ranges, 1600-3200Hz, 800-1599Hz, 400-799Hz, 200-399Hz, and 0-199Hz. This separation is sufficient to differentiate between high-frequency disturbances and low-frequency disturbances. Approximate coefficients represent the smoothed signal, while detailed coefficients represent the detailed signal containing sharper edges or sharper magnitude transition. As a result, fast and short transient disturbances are detected as lower-level decomposition, i.e., level 1; whereas long and slow transient disturbances can be detected at the level 4 approximate coefficient. The output of MSD can be noted as \mathcal{M} and the I^{th} level can then be mapped as:

$$\mathcal{M} = [cA_I, cD_I, cD_{I-1}, \dots, cD_1] \quad (5.1)$$

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

5.2.2 temporal aligned Layer

A temporal aligned layer is introduced to align the different length output coefficients of MSD. A functional block diagram of the temporal aligned layer is shown in Fig. 5.2. A 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 for the total number of coefficients output $K=4+1=5$. The temporal aligned MSD output, $\hat{\mathcal{M}}$ is obtained as follows,

$$\hat{\mathcal{M}} = [f_k(\mathcal{M}_k)]_{k=0}^K \quad (5.2)$$

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

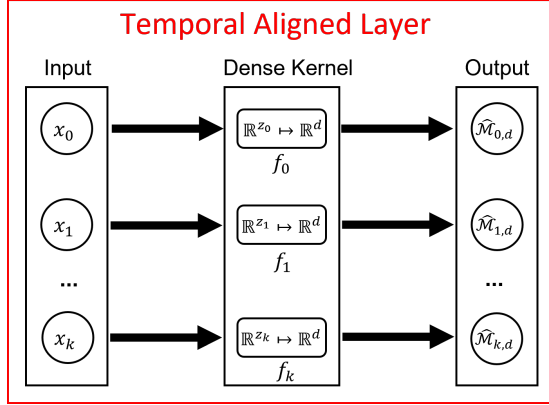


Figure 5.2: Temporal aligned layer.

5.2.3 Temporal-spatial Embedding

The temporal-Spatial embedding layer is introduced by further extracting the spatial relationship of the temporal-aligned MSD output using a layer of the 1D convolutional kernel. A convolution layer with kernel size = 5, stride = 1, input dimension $K = 5$, and output dimension = 16 is applied over the temporal aligned MSD output. Spatial-temporal information of the K frequency bands is thus extracted. These embedded temporal-spatial features are then fed into the transformer network for classification.

5.2.4 Efficient Superposition Embedding

Instead of applying a single perceptron layer over each of the frequency bands as described in the temporal aligned layer, efficient superposition embedding is used via an efficient embedding layer. Efficient embedding is achieved with an additional layer of shared perceptron layer as shown in Fig. 5.3. This shared perceptron layer significantly reduced the number of parameters required while giving the same output dimension with added spatial information linking between frequency bands. The first perceptron layer can be noted as $[f_k]_{k=0}^K : \mathbb{R}^Z \mapsto \mathbb{R}^{K \times e}$, where the single perceptron layer is a mapping function $f_k : \mathbb{R}^{z_k} \mapsto \mathbb{R}^e$, and K represent the total number of output coefficients from MSD such that $K = I + 1 = 5$. Note that this is similar to the temporal aligned layer except using a smaller value embedding parameter, $e=64$ in this experiment. The second perceptron layer is shared among the K outputs from the first embedding layer such that $[s_k]_{k=0}^K : \mathbb{R}^{K \times e} \mapsto \mathbb{R}^{K \times d}$. The mapping function for each of

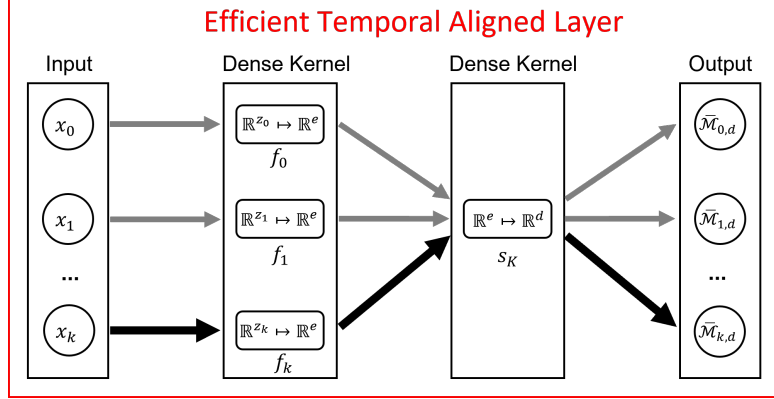


Figure 5.3: Efficient temporal aligned layer.

the output embedding can be described as $s_k : \mathbb{R}^e \mapsto \mathbb{R}^d$. The efficient aligned MSD output, $\bar{\mathcal{M}}$ is obtained as follows,

$$\bar{\mathcal{M}} = [s_k(\hat{\mathcal{M}})]_{k=0}^K = [s_k(f_k(\mathcal{M}_k))]_{k=0}^K \quad (5.3)$$

where $\bar{\mathcal{M}} \in \mathbb{R}^{K \times d}$ and fixed embedding output $d = 256$ is used. Similar to Temporal-spatial embedding as described in Section 5.2.3, the efficient aligned MSD output ($\bar{\mathcal{M}}$) is then processed via 1D convolutional kernel to extract higher order of spatial-temporal features.

5.2.5 Transformer Network

Transformer network was introduced by Vaswani et al. [143] which saves computation time by allowing parallel computation. The transformer network follows an encoder-decoder framework and is based solely on MHA. The input features are first normalised, then projected into three vectors, query Q , key K , and value V [144]. W_q , W_k , W_v representing the trainable parameters as depicted in Eq. (5.4).

$$Q, K, V = IW_q, IW_k, IW_v. \quad (5.4)$$

A single-head attention mechanism can be represented as,

$$Attn(Q, K, V) = Softmax\left(\frac{A}{\sqrt{d_k}}\right)V, \quad (5.5)$$

where $\frac{1}{\sqrt{d_k}}$ 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 .

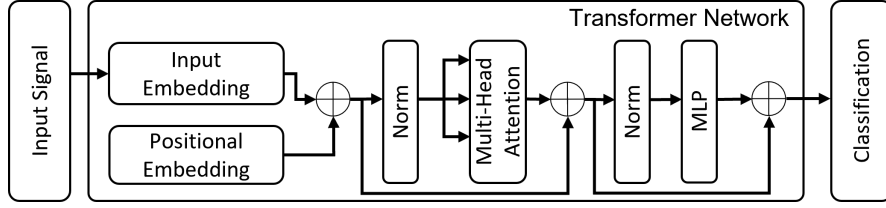


Figure 5.4: Transformer encoder block.

MHA [144] can be achieved by multiple sets of trainable matrices, $W_q^{(i)}$, $W_k^{(i)}$, $W_v^{(i)}$ as,

$$Q^{(i)}, K^{(i)}, V^{(i)} = IW_q^{(i)}, IW_k^{(i)}, IW_v^{(i)}, \quad (5.6)$$

$$Head^{(i)} = Attn(Q^{(i)}, K^{(i)}, V^{(i)}), \quad (5.7)$$

$$MultiHead(Q, K, V) = Concat(Head^{(1)}, \dots, Head^{(n)})W_O, \quad (5.8)$$

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. 5.4, the transformer encoder takes in input signals and embeds them with positional embedding. Data normalisation is done before passing into the MHA. Output data from the MHA is normalised and processed before outputting as an encoded latent feature.

5.2.6 Pooling Multi-head Attention

Pooling multi-head attention (PMA) mechanism [145] is introduced as an alternative for averaging or max-pooling in aggregating feature vectors. MHA is applied on a trainable 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,

$$y_k(Z) = MultiHead(S, Z, Z), \quad (5.9)$$

The output of PMA is the predicted result of each class based on their time-series pattern. To confine the M classes in probability, softmax is required as follows,

$$y = \frac{e^{y_k}}{\sum_{j=0}^M e^{y_{k_j}}}. \quad (5.10)$$

5.2.7 Batch Normalization Layer

A batch normalization layer is introduced to reduce the unwanted internal covariance shift, reduce vanishing or exploding gradients and regularize the network for better generalization [146]. For a d -dimensional input, $x = (x^{(1)}, x^{(2)}, \dots, x^{(d)})$, batch normalization can be described as

$$\hat{x}_i^{(k)} = \frac{x_i^{(k)} - \mu_B^{(k)}}{\sqrt{(\sigma_B^{(k)})^2 + \epsilon}}, \quad (5.11)$$

where m is the total size of the training set, and B represents 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 the batch normalization layer is depicted as

$$y_i^{(k)} = \gamma^{(k)} \hat{x}_i^{(k)} + \beta^{(k)}, \quad (5.12)$$

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

5.3 Experiment Setup

All simulation experiments are carried out using AMD Ryzen 7 3800X 8-Core Processor with Nvidia P6000 graphic processing unit. Pytorch framework is used for the experiments. Similar settings are used in Chapter 4, the MSD in both experiments are developed based on Daubechies 4 (db4) wavelet for the wavelet decomposition process. Additive white Gaussian noise (AWGN) is introduced to test the network classification performance. The main evaluation matrix used in the experiments is 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,

$$Acc_n = \frac{TP_n}{\sum_{j=0}^m S_j}. \quad (5.13)$$

Two main experiments are carried out focusing on the performance of the transformer network. Experiment #1 is carried out mainly to analyse the performance of the pure transformer network as compared to the LSTM network. In this experiment, a single-layer transformer encoder is modelled. The transformer model is depicted as in Fig. 5.4. The first part of this experiment is carried out using 18 classes of 10kHz

Table 5.1: Class of power quality disturbances (Experiment #1).

Label	Class Description	Label	Class Description
P_0	Normal	P_9	Flicker
P_1	Sag	P_{10}	Sag+Harmonics
P_2	Swell	P_{11}	Swell+Harmonics
P_3	Interrupt	P_{12}	Interrupt+Harmonics
P_4	Impulse Transient	P_{13}	Harmonics+Notch
P_5	Spike	P_{14}	Sag+Transient
P_6	Harmonics	P_{15}	Swell+Transient
P_7	Oscillatory Transient	P_{16}	Sag+Oscillatory Transient
P_8	Notch	P_{17}	Swell+Oscillatory Transient

sampled a single-period PQDs as shown in Table 5.1. In this experiment, the input dimension of the transformer encoder is the length of single-period waveform, which is 200. Four attention heads with 64 dimensions are used. The attention output is added to the input signal before passing to the feedforward layer with an internal dimension of 128. the input and output dimensions of the feedforward layer are the same as the input signal dimension of 200. The output of the feedforward layer is added and normalised as the final transformer output. Finally, transformer encoder outputs are passed to a fully connected layer which classifies the latent features into the total number of PQD classes. The results of training with and without noise are analysed. To benchmark the model with the available state-of-the-art, 16-classes PQDs dataset as shown in Table 5.2 were used and analysed in the second part of this experiment. The setting of this second part of the experiment is similar to the first part except for changing the input dimension to 640 for a 10-period 3200Hz sampled input signal.

The main proposed model is demonstrated in Experiment #2. This experiment is carried out entirely on the dataset with 16 classes of 10-periods 3200Hz sampled PQDs. The model used in this experiment is shown in Fig. 5.1. The first part of Experiment #2 starts with extracting multi-resolution output from the input signal via MSD. A total of five different frequency bands coefficients is harvested and passed through the temporal-spatial embedding layer as mentioned in Section 5.2.3. As similar settings are used for the the MSD process, the output from MSD process is always a feature matrix with a dimension of 5×256 . The 1D convolutional kernel in the embedding layer is set with input dimension 5, kernel size of 5, stride = 1, and output dimension = 16, giving

Table 5.2: Class of power quality disturbances (Experiment #2).

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

the output feature with dimension 16×252 . This feature output is then passed into the transformer encoder layer. Spatial feature attention is used. The input dimension of the transformer encoder is set spatially with size 16. Eight attention heads with a dimension size of two are used. The input and output of the feedforward layer are the same as the input dimension of 16, with an internal dimension of 256. Pooling with Multi-head attention or PMA is used. The outputs from the transformer encoder are pooled from the dimension of 252×16 to 1×16 via PMA. The outputs of PMA are then used for the classification process. The second part of Experiment #2 attempts to reduce the computation resources by using Efficient superposition embedding. The settings are similar to Experiment #2 first part except using superposition embedding instead of a temporal-spatial embedding mechanism.

5.4 Experiment #1 - Transformer Encoder Network

This experiment is carried out based on the transformer encoder network. AWGN is added to the simulated real-world noisy condition. The result of training with and without noise are performed for comparison. SNR AWGN with 20-50dB are randomly added into the training samples when training with noise. The model is then tested with 20dB, 30dB, 40dB, 50dB, and noiseless testing samples. The first part of this experiment is carried out using a single-period, 18-PQD classes dataset. While the

second part in this Experiment #1 is tested with 10-periods, 16-PQD classes dataset, which aligns with literature [20, 101].

5.4.1 Single-period PQD Analysis on Transformer Encoder Network

In this part of the experiment, noiseless data samples are used in training both LSTM and transformer networks. The trained models are then tested with noiseless, and 20dB, 30dB, 40dB AWGN, and the results are tabulated in Table 5.3. The proposed transformer network attained an accuracy of 55.56% which is better than that of the LSTM model with only 13.26% on the 20dB SNR test. The transformer network proves better generalization capability against noisy conditions by attaining $> 80\%$

Table 5.3: Noiseless training: Classification performance comparison between LSTM network and Transformer network.

Model SNR Class	LSTM				Transformer			
	20dB	30dB	40dB	noiseless	20dB	30dB	40dB	noiseless
P_0	0.00	0.00	1.33	99.83	9.38	76.62	99.74	99.95
P_1	0.05	20.31	59.54	98.01	59.17	87.51	92.29	92.47
P_2	0.05	11.27	54.57	98.46	76.60	94.20	96.61	96.81
P_3	12.54	44.64	69.23	97.85	53.66	66.68	69.78	70.37
P_4	0.00	5.50	70.32	97.33	40.28	80.44	86.90	87.58
P_5	0.00	6.01	56.03	97.78	62.74	85.29	87.99	87.86
P_6	0.20	31.51	84.21	99.41	79.93	97.99	99.41	99.59
P_7	98.30	98.13	95.87	94.95	84.97	90.70	91.38	91.25
P_8	0.39	18.28	53.50	96.83	48.41	75.67	78.55	78.45
P_9	0.00	0.76	42.16	99.86	68.92	98.51	99.88	99.88
P_{10}	3.70	60.84	91.68	96.20	80.91	88.84	89.75	90.00
P_{11}	52.78	77.89	95.25	97.93	94.83	97.47	97.67	97.80
P_{12}	26.74	60.65	84.23	90.70	63.91	72.83	74.44	74.35
P_{13}	14.51	83.61	97.60	97.92	59.86	58.35	58.24	57.91
P_{14}	13.09	57.18	92.89	94.34	65.71	74.34	74.70	74.51
P_{15}	6.70	32.59	86.33	95.25	72.26	76.80	77.00	76.84
P_{16}	69.36	73.09	80.28	81.25	80.72	80.65	80.74	80.91
P_{17}	48.56	55.55	59.68	72.29	15.31	15.09	15.02	14.93
WAcc	13.26	31.84	58.36	96.42	55.56	81.39	87.71	87.84

weighted accuracy (WAcc) on $> 30dB$ SNR noise introduced. The WAcc of the LSTM model on the other hand shows a significant impact on the high noise condition introduced. It can also be noticed that the classification is biased towards class P_7 -oscillatory transient, which has oscillating characteristics similar to that of high noise conditions. The transformer network has an overall better classification performance as compared to the LSTM network in this case. However, weaknesses are still present in some of the PQD classes. These classes includes fast disturbance classes such as P_8 -Notch, P_{13} -Harmonics+Notch, P_{14} Sag+Transient, P_{15} -Swell+Transient. Besides, the poor-performing disturbance classes affected include slow disturbance classes such as P_3 -Interrupt, P_{12} Interrupt+Harmonics, P_{17} -Swell+Oscillatory Transient. For the fast disturbance classes, most of the listed less performing classes have $> 70\%$ WAcc except class P_{13} . The analysis of the confusion matrix shows that class P_{13} is confused with the single disturbance class P_6 -Harmonics. By looking at the miss-classified data, it can be found that smaller magnitude notches or transients confuse the classification network. As for the slow disturbance classes, class P_3 and class P_{17} show poor classification performance in confusion matrix analysis. Class P_3 is confused with class P_{12} -Interrupt+Harmonics. Class P_{12} has both Interrupt (larger magnitude changes) and Harmonics (smaller magnitude changes). When noise is introduced, the slight changes in the magnitude of the interrupt signal cause P_3 to become similar to P_{12} . As for class P_{17} , it can be noticed that the confusion happens mostly between class P_{17} and P_{16} -Sag+Oscillatory Transient. These two classes are distinct from each other purely based on magnitude difference. A small difference between the two classes is causing confusion for the transformer network. This shows that the transformer classifier is limited to finding distinct features from magnitude differences. Slight magnitude changes imposed gives a negative effect on the classification performance. Comparatively, LSTM is highly sensitive to noisy data and the performance of the network is badly impacted by the noise introduced.

This experiment continues with training the model using data samples randomly added with 20-50dB SNR AWGN and the results are tabulated in Table 5.4. Drastic improvements are shown on both LSTM and transformer models. The noisy conditions introduced during the training phase exposed the models to the characteristics of the noise. This allows the machine learning model to learn the noise-polluted features, thus increasing the classification performance. The drastic improvement in the LSTM model on "seen noises" shows that the model is capable of learning extra features added to the

Table 5.4: Noisy training: Classification performance comparison between LSTM network and Transformer network.

Model Class	LSTM				Transformer			
	20dB	30dB	40dB	noiseless	20dB	30dB	40dB	noiseless
P_0	96.83	99.61	99.85	99.94	88.42	99.74	99.94	99.95
P_1	85.92	94.32	96.42	96.87	78.44	90.77	91.74	91.77
P_2	87.82	94.80	96.81	96.99	87.10	93.91	94.50	94.58
P_3	70.96	85.55	90.66	91.89	53.96	58.69	58.76	58.90
P_4	49.54	81.74	90.06	91.28	46.72	68.48	71.31	71.70
P_5	66.05	95.24	97.33	97.46	44.82	52.31	53.41	53.88
P_6	92.83	98.70	99.18	99.27	91.78	98.48	98.77	98.77
P_7	91.47	92.10	91.73	91.85	86.55	89.38	89.42	89.53
P_8	72.52	94.76	97.02	96.91	38.45	49.83	52.09	52.48
P_9	91.59	97.60	98.09	98.19	93.15	99.34	99.87	99.93
P_{10}	87.63	93.95	95.09	95.24	84.16	87.81	88.12	88.17
P_{11}	96.30	97.19	97.34	97.39	95.91	97.41	97.61	97.62
P_{12}	59.16	77.59	87.62	89.64	65.17	69.33	68.70	68.36
P_{13}	81.53	93.79	96.01	96.25	49.81	46.82	46.49	46.49
P_{14}	57.29	76.68	84.85	86.33	56.62	56.78	57.21	57.13
P_{15}	68.39	86.05	90.72	91.47	70.66	72.57	71.94	72.02
P_{16}	73.93	74.88	75.01	75.07	77.04	77.27	77.12	77.11
P_{17}	42.86	48.92	49.80	49.95	19.96	20.38	20.06	20.10
WAcc	82.62	91.52	93.68	94.07	76.06	82.90	83.27	83.31

training datasets. While drastic improvement can be seen from transformer network classification on 20dB SNR AWGN, the classification performance on 30dB onwards is showing decreasing in classification performance. The poor-performing classes from the previous subsection are still similar (training without noise conditions). This shows that the transformer network can learn more generalised features even without manually feeding in noisy conditions that might occur during the training phase. However, noisy data fed into the training phase does improve the classification performance on the 20dB test. Besides the previously discussed fast disturbances classes (P_8 , P_{13} , P_{14} , P_{15}) as discussed in the previous subsection, the performance of P_4 -Impulse transient, and P_5 -Spike also showing weaker classification. The confusion matrix of different noise levels

Table 5.5: Testing classification rate in (%) with three unknown noise performance comparison.

Model \ Noise	Train Without noise			Train With noise		
	A	B	C	A	B	C
LSTM	10.36	46.85	16.28	52.42	48.34	56.69
Transformer	40.84	60.27	41.64	56.55	53.17	57.96

on these two classes shows that there is more confusion with class P_0 , or signal without disturbance. When noises are added to the training samples, the boundary between the normal signal and the disturbance signal are blurred by the noise introduced. Training with noise closes the gap between disturbance and non-disturbance waveforms due to the overlapping of noises on the magnitude of the disturbance signals. While the transformer network is proved to be insensitive towards signals with similar magnitude, these results strengthen the proof of the statement. Meanwhile, the slow disturbance class ($P_3, P_{12}, P_{16}, P_{17}$) problems still exist as discussed in the previous subsection. The transformer encoder network is limited to finding distinct features from magnitude differences that persist during training with noise conditions.

The experiment is then continued by testing both models with unknown noises. Three unknown noises which are not previously exposed to the models are introduced into the test samples. Noise #A is a 15 dB SNR AWGN, Noise #B is a 20-25 dB SNR of positive-uniformly distributed random noise, and Noise #C is 15-30 dB SNR uniformly distributed random noise. The performance of the model trained with and without noise are tabulated in Table 5.5. Training without noise results in low classification accuracy for the LSTM model. On the other hand, the transformer network is still able to achieve $> 40\%$ accuracy for noise #A, #B and #C. When 20-50 dB AWGN are introduced during the training phase, the classification performance of both models increases. However, it can be noticed that LSTM failed to achieve a similar performance level of 82.62% even on noise #B and noise#C having similar noise levels. This shows that the LSTM model only works best for "seen conditions". Transformer on the other hand shows more generalised performance on both trained with and without noise tests. However, the classification performance is still limited due to the transformer not being able to classify signals with changes in magnitude.

5.4.2 Multi-period PQD Analysis on Transformer Encoder Network

This part of the experiment analyses the transformer model trained and tested using 16-classes PQD dataset as shown in Table 5.2. 10-period 16 classes PQD signals sampled at 3200Hz are used. The use of remodelled models and datasets from literature [20, 101] allows benchmarking of the proposed model. The LSTM model and transformer encoder model are trained using the 16-classes PQDs randomly added with 20-50 dB SNR AWGN. The classification performance comparisons of LSTM and Transformer model are tabulated in Table 5.6.

As compared to the previous result in Section 5.4.1, the performance of model classification can be improved using 10-period input data. Although there are differences between the classes of PQDs inserted into the training mechanism, having more data generally increases the information content per input for processing. As a result, the average classification accuracy has increased to $> 85\%$ for all the condition tests.

Table 5.6: Performance comparison of Deep LSTM with the Transformer Encoder model tested with 20-50dB AWGN and noiseless conditions.

(a) Deep LSTM.						(b) Transformer Encoder.					
SNR \ Class	20dB	30dB	40dB	50dB	noiseless	SNR \ Class	20dB	30dB	40dB	50dB	noiseless
P0	85.40	99.80	99.60	99.70	99.63	P0	71.30	89.50	94.90	97.90	98.03
P1	93.30	97.00	96.60	96.80	96.80	P1	93.70	96.20	96.40	94.70	95.60
P2	97.20	97.10	98.40	98.60	97.90	P2	95.90	98.00	98.50	98.60	98.90
P3	98.20	99.70	99.90	99.90	100.0	P3	99.00	99.50	99.30	99.10	99.70
P4	99.90	100.0	100.0	100.0	100.0	P4	99.90	100.0	100.0	100.0	100.0
P5	98.90	99.50	99.30	99.30	99.90	P5	96.60	97.20	97.00	96.20	97.90
P6	98.60	99.60	99.30	99.70	99.50	P6	99.90	100.0	100.0	100.0	100.0
P7	97.80	99.80	99.10	99.50	99.80	P7	98.50	99.30	98.60	98.70	98.60
P8	66.70	95.90	100.0	100.0	100.0	P8	100.0	100.0	100.0	100.0	100.0
P9	94.50	99.20	99.90	99.90	99.70	P9	83.00	92.70	95.40	94.20	94.30
P10	56.50	74.80	85.80	88.40	90.53	P10	62.40	93.40	96.40	97.40	99.08
P11	83.30	96.30	98.00	97.00	99.49	P11	80.20	96.30	98.10	97.20	100.0
P12	74.00	93.00	95.80	96.50	98.87	P12	83.80	95.50	96.80	97.40	100.0
P13	89.50	100.0	100.0	100.0	100.0	P13	100.0	100.0	100.0	100.0	100.0
P14	94.10	97.50	96.60	97.00	96.20	P14	98.90	98.90	99.50	99.30	98.70
P15	87.80	97.00	96.90	98.20	98.10	P15	99.80	99.40	99.30	99.30	99.20
Acc	88.48	96.64	97.83	98.16	98.54	Acc	91.43	97.24	98.14	98.13	98.74

By looking into the 16-classes dataset, most of the multiple disturbance classes (P10 to P15) are the combination of slow disturbance classes. In addition, the single disturbance classes can still be tested with the single disturbance classes. Hence, the model performance on slow and fast disturbances can be evaluated using 10-period input data.

Results in Table 5.6 show that the transformer encoder has better overall classification performance compared to the LSTM model. The classification performance of the Deep LSTM network is notably weak on class P10-Sag+Harmonics regardless of the level of noise introduced. The confusion matrix shows that the confusion has occurred with class P0-Normal. Harmonics in class P10 and P11-Swell+Harmonics might contribute to the average magnitude of the signal, which brings the P10 or P11 average magnitude close to the average magnitude of a normal signal. This shows that the LSTM model struggles to identify the characteristics of harmonics, especially within P10. When high noise of 20dB SNR AWGN was used, fast disturbance class P8-Notch was also affected. This can be caused by the additive effect of AWGN noise introduced which neutralised small notching characteristics. Besides, multiple disturbance classes such as P12-Interrupt+Harmonics, P13-Flicker+Harmonics, P14-Flicker+Sag and P15-Flicker+Swell are also having lower classification accuracy on high noise 20dB. Although there is more information available in each 10-period data sample, the effect of noise on the overall magnitude of the entire signal is also stronger and thus causing the misclassification.

5.5 Experiment #2 Multi-resolution Attention Using Convolutional Transformer

In this experiment, multi-resolution analysis is performed. MSD is introduced using multiple levels of the discrete wavelet transform. Several feature processing efforts are made to the MSD features before feeding into the transformer encoder for higher-order feature extraction and classification. The first part of this experiment analyses the performance of the proposed wavelet-based convolutional transformer network as shown in Fig. 5.1. The detailed classification performance analysis of the proposed model is performed using a confusion matrix. The analysis is done by comparing it to the Deep CNN model [20]. The second part of this experiment continues with the proposed efficient embedding mechanism which significantly reduces the model size of

the proposed method. The model complexity analysis is computed for comparison between the proposed models with the state-of-the-art methods.

5.5.1 Classification Performance Comparisons of WT-ConvT Versus Deep CNN

The proposed wavelet-based convolutional transformer (WT-ConvT) is compared with Deep CNN [20] to demonstrate the strength of the model in classifying PQDs under 20-50dB SNR AWGN and noiseless conditions. It is noticed in Table 5.7 that WT-ConvT attained higher classification accuracy of 94.11% tested with 20dB AWGN. Deep CNN on the other hand only acquire classification accuracy of 90.56%. It can also be noticed in Table 5.7a that WT-ConvT perform slightly poorly for P0-Normal, P9-Flicker, and P10-Sag+Harmonics. All these classes are from slow disturbance classes, which occurred at a lower frequency. Comparatively, it can be noticed that the Deep

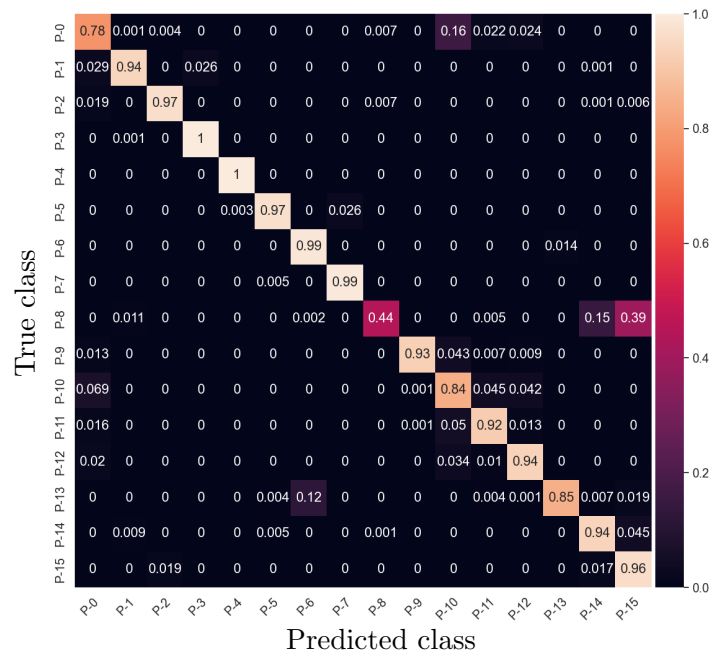
Table 5.7: Performance comparison of WT-ConvT with the Deep CNN model tested with 20-50dB AWGN and noiseless conditions.

(a) WT-ConvT.						(b) Deep CNN [20].					
SNR \ Class	20dB	30dB	40dB	50dB	noiseless	SNR \ Class	20dB	30dB	40dB	50dB	noiseless
P0	67.10	85.00	97.60	99.30	99.44	P0	86.00	96.30	99.30	100.0	100.0
P1	94.80	97.30	96.80	96.40	97.20	P1	93.80	98.30	98.40	97.80	98.10
P2	95.30	98.10	99.20	98.80	98.80	P2	97.80	97.60	98.70	98.90	98.80
P3	99.60	99.80	99.80	100.0	99.80	P3	100.0	99.80	99.80	100.0	100.0
P4	100.0	100.0	100.0	100.0	100.0	P4	100.0	100.0	100.0	100.0	100.0
P5	97.90	98.70	98.60	98.60	98.80	P5	95.80	98.70	98.70	99.10	99.20
P6	99.80	100.0	99.90	100.0	99.90	P6	99.40	99.90	100.0	100.0	100.0
P7	97.40	98.90	99.10	99.00	98.80	P7	100.0	100.0	100.0	99.90	100.0
P8	91.60	100.0	100.0	100.0	100.0	P8	39.80	89.10	100.0	100.0	100.0
P9	86.00	97.90	99.80	99.70	99.70	P9	95.50	100.0	100.0	100.0	100.0
P10	91.60	92.30	92.70	90.30	92.87	P10	82.10	91.80	96.40	97.00	99.59
P11	94.70	97.60	98.20	97.20	100.0	P11	88.70	97.50	98.20	97.20	100.0
P12	94.90	97.00	96.80	97.40	100.0	P12	91.00	96.40	96.50	96.60	99.59
P13	99.60	100.0	100.0	100.0	100.0	P13	88.50	100.0	100.0	100.0	100.0
P14	98.00	98.90	98.70	98.50	98.60	P14	92.90	98.40	99.20	98.40	98.90
P15	97.50	98.00	98.10	98.30	97.70	P15	97.70	99.20	98.30	99.20	98.90
Acc	94.11	97.47	98.46	98.34	98.86	Acc	90.56	97.69	98.97	99.01	99.57

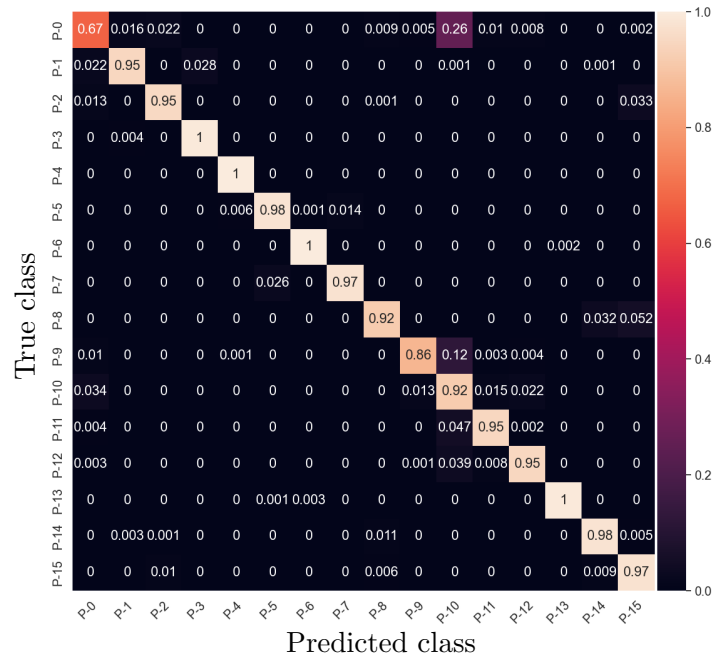
CNN model contains more classes that acquire less than 90% accuracy. These classes include four slow disturbance classes (P0, P10, P11, P13) and one fast disturbance class P8. As a result, the proposed WT-ConvT has better classification performance, especially in identifying fast disturbance.

A confusion matrix is a visualization summary of prediction results on a classification problem. Fig. 5.5a shows the confusion matrix of the Deep CNN method tested on 20dB SNR AWGN. From the matrix, class P0-normal is confused with class P10-sag+harmonics. This effect can be explained with a higher level of additive noise which increases the overall signal power and covers the sag effect, while high noise also added to harmonics effects. The deep CNN method also performed badly in classifying P8-Notch when the signal is under 20dB SNR condition. Class P8-Notch, a fast transient disturbance is confused with class P14-flicker+sag, and class P15-Flicker+swell which are both slow disturbance classes. While the small notching effect can be easily covered by the additive noise, this notching effect is also contributing to the changes in the average magnitude of the signal, thus producing confusion with class P14 and P15. As for analysis, a notch is categorised as a fast transient disturbance, while flicker with sag or swell are slow disturbances that 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 can be across different frequency signals.

WT-ConvT model uses MSD to decompose the 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. 5.5b shows that the WT-ConvT model suffers less impact from P8-Notch due to wavelet component could help in splitting the low and high-frequency noise to improve the generalization of feature extraction. The confusion matrix shows that the proposed WT-ConvT has higher confusion on two slow disturbance classes. Similar to Deep CNN, Class P0-normal is confused with Class P10-sag+harmonics. There is also a slight confusion that happened on class P9-flicker, where class P9 can be classified wrongly as class P10. This confusion can be caused by uneven noise added to the flicker signal, where the slightly lower magnitude of the flicker signal is classified wrongly as sag class, and the noise introduced is identified as harmonics. Both of these results show that the proposed WT-ConvT can still be improved in its performance in detecting a slow disturbance. However, the proposed WT-ConvT performed better on fast disturbance classification as compared to the Deep CNN method.



(a)



(b)

Figure 5.5: Confusion matrix at 20dB SNR AWGN test on (a) Deep CNN and (b) WT-ConvT.

5.5.2 Efficient Embedding Multi-resolution Attention Using Convolutional Transformer and Complexity Analysis

A weight superposition mechanism is introduced to enhance the performance of the proposed WT-ConvT network while reducing the model size. Instead of embedding each of the DWT bands individually, a weight superposition is introduced. The weight-sharing mechanism of this method reduces the model size significantly while maintaining the performance of the network. In this experiment, the performance comparison of the proposed efficient wavelet-based convolutional transformer (EWT-ConvT) with WT-ConvT is carried out to analyse the advantage of using efficient embedding.

The classification performance of the proposed EWT-ConvT is compared with WT-ConvT as shown in Table 5.8. It can be noticed that both WT-ConvT and EWT-ConvT models have similar overall performance. The classification performance on P0-Normal, P9-Flicker, and P10-Sag+Harmonics are poorer. Both models are more

Table 5.8: Performance comparison of WT-ConvT with the EWT-ConvT model tested with 20-50dB AWGN and noiseless conditions.

(a) WT-ConvT.						(b) EWT-ConvT.					
SNR \ Class	20dB	30dB	40dB	50dB	noiseless	SNR \ Class	20dB	30dB	40dB	50dB	noiseless
P0	67.10	85.00	97.60	99.30	99.44	P0	68.40	82.80	93.80	98.10	97.84
P1	94.80	97.30	96.80	96.40	97.20	P1	94.50	96.70	96.60	96.10	97.10
P2	95.30	98.10	99.20	98.80	98.80	P2	98.30	99.10	99.10	99.30	99.50
P3	99.60	99.80	99.80	100.0	99.80	P3	99.10	99.20	99.70	99.10	99.50
P4	100.0	100.0	100.0	100.0	100.0	P4	100.0	100.0	100.0	100.0	100.0
P5	97.90	98.70	98.60	98.60	98.80	P5	96.20	97.70	97.50	96.50	98.10
P6	99.80	100.0	99.90	100.0	99.90	P6	100.0	100.0	100.0	100.0	99.90
P7	97.40	98.90	99.10	99.00	98.80	P7	99.70	99.20	99.60	99.60	99.40
P8	91.60	100.0	100.0	100.0	100.0	P8	97.20	100.0	100.0	100.0	100.0
P9	86.00	97.90	99.80	99.70	99.70	P9	82.20	97.20	99.60	99.60	99.30
P10	91.60	92.30	92.70	90.30	92.87	P10	90.20	93.40	93.40	92.80	95.62
P11	94.70	97.60	98.20	97.20	100.0	P11	93.80	97.60	98.20	97.20	100.0
P12	94.90	97.00	96.80	97.40	100.0	P12	94.60	97.00	96.80	97.40	100.0
P13	99.60	100.0	100.0	100.0	100.0	P13	100.0	100.0	100.0	100.0	100.0
P14	98.00	98.90	98.70	98.50	98.60	P14	99.20	99.50	99.50	99.40	99.20
P15	97.50	98.00	98.10	98.30	97.70	P15	97.30	96.70	95.90	97.50	96.80
Acc	94.11	97.47	98.46	98.34	98.86	Acc	94.42	97.26	98.11	98.29	98.89

sensitive to disturbance signals which cause lower classification accuracy on high noise class P0. This comparison also shows that both WT-ConvT and EWT-ConvT are having similar performance, with the EWT-ConvT model having far fewer parameters.

The confusion matrix of the EWT-ConvT on 20dB SNR AWGN is shown in Fig. 5.6. Similar to the confusion matrix of WT-ConvT, there are a few slow disturbances that have lower classification performance. The confusion on class P0-Normal being classified as class P10-Sag+Harmonics is caused by a higher level of additive noise changing the overall signal power and adding to the effects of the harmonics. Similarly, the confusion of class P9-Flicker as class P10 is also presents in this EWT-ConvT model. This confusion can be caused by uneven noise added to the flicker signal, where the slightly lower magnitude of the flicker signal is classified wrongly as sag class, and the noise introduced is identified as harmonics. It can be noticed that classes P10 is mainly confused with class P0 and P9, where all of the classes are slow disturbance classes. The effect of noise levels shows that the proposed model can still be improved in detecting slow disturbance classes.

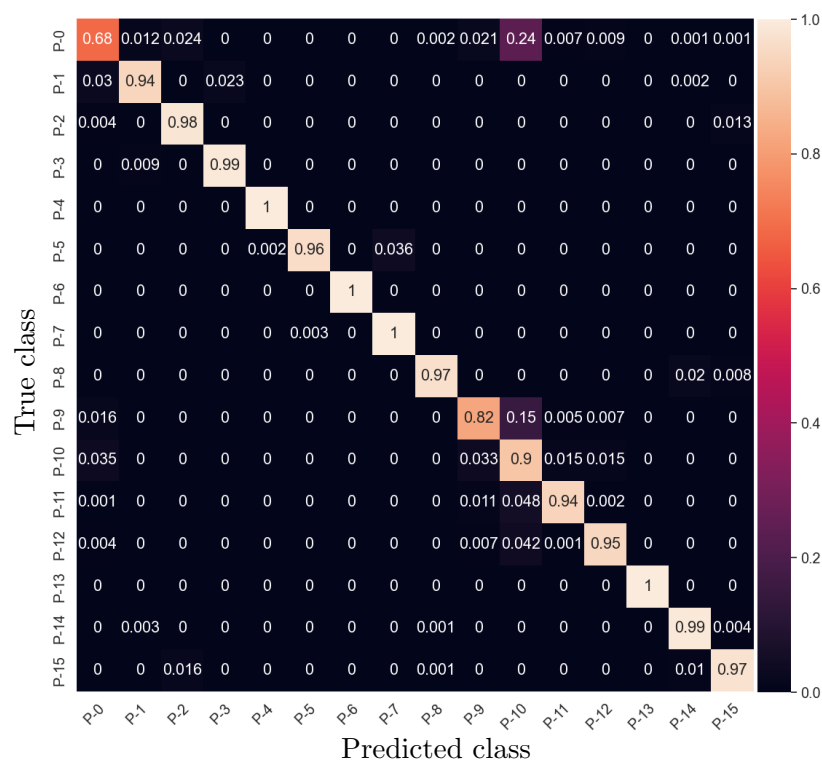


Figure 5.6: Confusion matrix for EWT-ConvT with efficient superposition embedding at 20dB SNR AWGN test.

Table 5.9: Model complexity comparisons.

Model	20dB	30dB	40dB	50dB	noiseless	Time per epoch	No. of parameters	Model size
Deep LSTM [101]	88.48	96.64	97.83	98.16	98.54	68s	33936	0.136 MB
Deep CNN [20]	90.56	97.69	98.97	99.01	99.57	32s	164464	0.657 MB
WT-SFA-LSTM	93.79	97.49	98.34	98.51	99.09	34s	221486	0.871 MB
WT-ConvT	94.11	97.47	98.46	98.34	98.86	32s	182970	0.728 MB
EWT-ConvT	94.42	97.26	98.11	98.29	98.89	32s	70778	0.290 MB

The complexity of the proposed EWT-ConvT, WT-ConvT, WT-SFA-LSTM, Deep CNN[20], and Deep LSTM [101] are compared as demonstrated in Table 5.9. The best 20dB accuracy, training time per epoch of the training on 76.8k samples, the number of parameters and model size are recorded and compared. The deep LSTM method achieved a good classification accuracy of 88.48% on high noise conditions with the lowest number of parameters. However, the time required to train an epoch of samples is doubled compared to the other models. Deep CNN shows better performance of 90.56% classification accuracy on 20dB SNR condition with only 32 seconds training time per epoch and with an adequate model size of 0.657 MegaBytes. Our proposed WT-SFA-LSTM model achieved a better classification accuracy of 93.79%. However, the training time per epoch, and the number of parameters used or model size are at the highest of 0.871 Megabytes. To improve the model complexity and enable faster training, WT-ConvT is proposed. This model achieved the highest classification accuracy of 94.11% and with fewer parameters compared to the WT-SFA-LSTM method. However, the model size of WT-ConvT is slightly larger than that of the Deep CNN method. EWT-ConvT is thus proposed. An efficient weight superposition mechanism is introduced to compress the model size of WT-ConvT into 60% fewer parameters. The model size of 0.290 Megabytes is positioned at 43.03% of the Deep CNN method and 38.68% of the WT-ConvT model. While having significantly reduced parameters, the classification performance of EWT-ConvT is slightly better with 94.42%. Most of the characteristics of WT-ConvT are kept with a slight improvement on the EWT-ConvT model.

5.5.3 Analysis on Batch Normalization

The studies of batch normalization (BN) layer are carried out on both WT-ConvT and EWT-ConvT models. The performance of both of the models without the BN layer is tabulated in Table 5.10. Similar classification performance is observed in WT-ConvT without the BN layer as shown in Table 5.10a. This can be due to only low-level representation embedding being used before the transformer network in the WT-ConvT model. This is not the case in the EWT-ConvT model. The efficient superposition embedding mechanism in the EWT-ConvT model involves sharing of trainable weights. This weight superposition allows the network to use the same set of parameters to learn representation from each of the input bands. Besides the benefits of using fewer parameters, these shared weights also give a generalised representation of all the input bands. However, as this shared weight involves more forward passes during the training process, vanishing or exploding gradients are more prone to happen. The results in

Table 5.10: Performance comparison of WT-ConvT and EWT-ConvT without batch normalization layer tested with 20-50dB AWGN and noiseless conditions.

(a) WT-ConvT without batch normalization.

SNR \ Class	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

(b) EWT-ConvT without batch normalization.

SNR \ Class	20dB	30dB	40dB	50dB	noiseless
P0	68.10	90.10	98.00	99.50	99.72
P1	94.10	97.50	97.70	95.50	97.10
P2	97.50	97.90	98.90	98.70	98.40
P3	97.50	97.40	97.20	96.30	98.10
P4	100.0	100.0	100.0	100.0	100.0
P5	97.40	98.70	98.10	97.70	98.20
P6	99.70	100.0	99.90	100.0	100.0
P7	96.80	98.30	98.40	97.40	97.10
P8	86.50	100.0	100.0	100.0	100.0
P9	86.60	98.30	99.90	99.90	99.80
P10	80.20	89.00	92.10	92.20	94.09
P11	91.50	97.50	98.20	97.20	100.0
P12	94.80	96.90	96.80	97.40	100.0
P13	97.00	100.0	100.0	100.0	100.0
P14	97.60	98.90	99.10	99.20	98.90
P15	94.80	97.50	97.10	98.40	98.30
Acc	92.51	97.38	98.21	98.09	98.74

Table 5.10b show lower classification accuracy on high noise 20dB SNR AWGN test, with only 92.51% without BN layer as compared to 94.42% with BN layer as shown in previous section Table 5.8b.

In addition, the decrease in classification performance in the EWT-ConvT model without the BN is mainly caused by class P8-Notch and P10-Sag+Harmonics as shown in Fig. 5.7 the. The decrease in classification accuracy on class P8 suggests that the model without BN layer is facing difficulty in detecting fast transients in high noise conditions. Where class P8, a fast transient disturbance is confused with slow disturbance classes, class P14-Sag+Flicker and P15-Swell+Flicker. On the other hand, the decreased accuracy in class P10 shows decreased performance in classifying slow disturbance classes. Without a BN layer, a weight distribution issue may happen, which causes a decrease in performance. The BN layer is thus proved to be important for the EWT-ConvT model which has weight sharing mechanism.

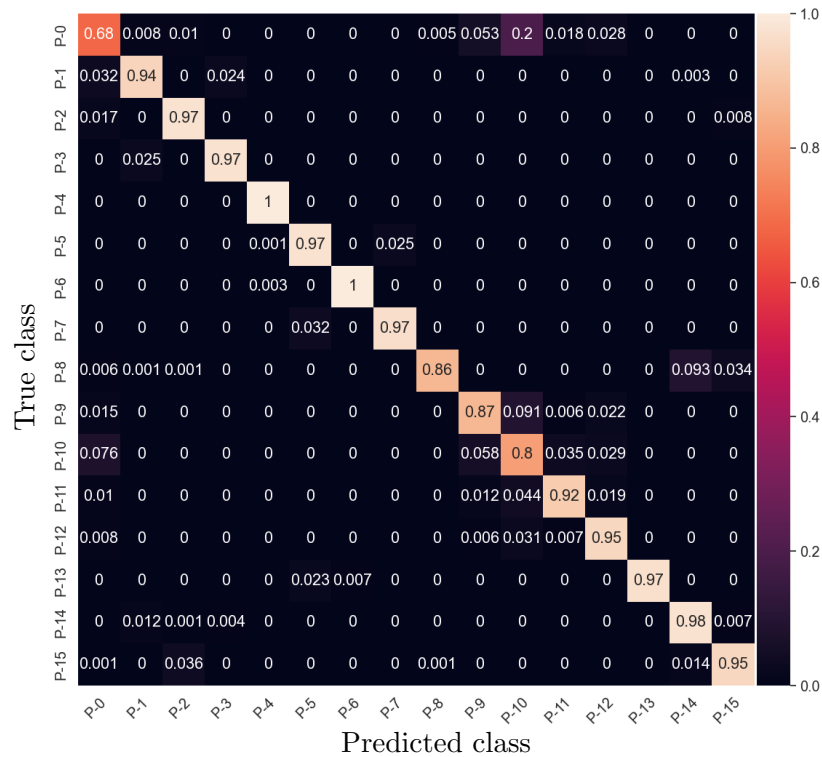


Figure 5.7: Confusion matrix for EWT-ConvT without BN layer at 20dB SNR AWGN test.

5.6 Chapter Summary

A transformer network is proposed in this chapter to improve the training mechanism of deep learning networks with parallel processing capability. It contains multi-head attention in a transformer allowing multiple feature embedding for better sequential information extraction. As a result, the transformer encoder network can achieve 76.06% and 55.56% when the model is trained with and without noise respectively using a single-period data sample. This result shows better noise immunity compared to the LSTM model with 82.62% and 13.26% when tested with 20dB SNR AWGN. The transformer model also shows similar generalization characteristics when compared to GA-LSTM. These models are then tested using a benchmarking dataset with a multi-period waveform. Results show that both the transformer and LSTM model achieve better performance with longer sequence input. Transformer leads the performance with 91.43% versus 88.48%, showing higher generalization capability. To further increase the resolution of the input signal, MSD is applied to decompose the signal into multiple bands. Hence, a hybrid model of the WT-ConvT network is proposed to improve the performance of PQD classification. In the WT-ConvT network, MSD coefficients are fed into a temporal-spatial embedding layer, which consists of a temporal aligned layer and a layer of the 1D convolutional kernel. It is used to align coefficients in equal lengths for feature embedding. Subsequently, the embedded features are passed to the transformer encoder to extract higher-level latent representation. Finally, these higher-level latent representations can be aggregated using the MHA pooling mechanism for PQD classification. As a result, WT-ConvT can highlight the salient difference between fast transient and slow disturbance with 94.11% as compared to the benchmark Deep CNN with 90.56%. Efficient embedding mechanisms are introduced to better optimize the model using the weight superposition algorithm. The optimized EWT-ConvT achieves a similar classification performance of 94.42% with a 61.31% reduction in model size when compared to WT-ConvT. EWT-ConvT with 0.290 MB model size and the highest classification accuracy stands out as the best-performing model when compared to WT-ConvT, WT-SFA-LSTM, Deep CNN, and Deep LSTM models. The importance of having batch normalization in models involving complex weight sharing is also highlighted to prevent vanishing or exploding gradient on weights, and regularise the network for better generalization.

Chapter 6

Real-time Embedded System Implementation with Power Quality Classification

6.1 Introduction

Energy monitoring and control are of paramount importance in ensuring the quality of power supplies. Proper management and control efforts become more important with the blooming integration of microgrids, especially with renewable energy resources (RER), which are environmentally friendly with additional benefits of low operating cost [147]. The intermittency nature of RERs can bring a negative impact on the grid with improper handling [148]. Besides from RERs, a microgrid also consists of other distributed energy resources, multiple types of large loads, and large storage systems which include electric vehicles [149]. The increased amount and type of loads onto the power systems increased the complexity of the power systems, which brings a negative impact on the power quality [150]. This poses serious challenges towards the recent advanced technologies which use sensitive electronic components that require pure sinusoidal power supplies for proper operations. The increased integration of RERs, and grid-tied storage systems such as electric vehicles and fuel cell technologies demand synchronised multiple point monitoring systems [151]. A scalable, real-time monitoring system is thus required. Due to their feasibility and cost-effectiveness, embedded systems have emerged as the optimal option for scaled deployment [151]. For instance,

smart meters data are used for power quality disturbances (PQD) classification in [112]. The use of field-programmable gate arrays (FPGA) is widely studied in [152, 153].

Real-time classification of PQD is important in the field of power management and control. This is due to the increased complexity of the power grid with an increasing number of advanced power electronics connected [154]. Real-time monitoring of the power systems helps in improving the intermittency issue presence in renewable energy microgrids by allowing energy scheduling [155]. Real-time classification of PQD is required for immediate mitigation action [156]. For instance, the control of MG islanding is required when faults occurred. Real-time classification of PQD is one of the critical features required to monitor the reliability of the entire power system. The real-time monitoring and control of a smart grid controller is also critical to ensure continuous and reliable supplies of electricity across the entire grid. The ability to identify PQD occurring in the power systems allows preventive maintenance to be taken before a major breakdown of the critical systems. The ability to detect and classify types of PQD in real time is thus an important feature to ensure the reliability of the entire power system.

A real-world PQD classification is proposed and examined in this chapter using the setup of the Raspberry Pi embedded system. In Section 6.2, The overview of the proposed hardware-software integration is described. Laboratory experiments and the setup is discussed in Section 6.3. In Section 6.4, The data acquisition and pre-processing of the data before feeding into the classification model are described. Classification performance of the proposed models and literature models are compared in Section 6.5. The hardware performance is discussed in Section 6.6. Finally, a conclusion of hardware development for PQD classification is drawn in Section 6.7.

6.2 Proposed Hardware Implementation

The proposed embedded systems for real-world implementation of the proposed machine learning model is depicted in Fig. 6.1. The main components included in this setup include a microcontroller, Raspberry Pi 4B (RPi), Waveshare AD-HAT ADS1263 Analog-to-Digital converter (ADC), and a current transformer. The proposed embedded system can be applied in a junction box or before the specific load. The operating voltage of the subject being measured as shown in Fig. 6.1 is a single-phase load with

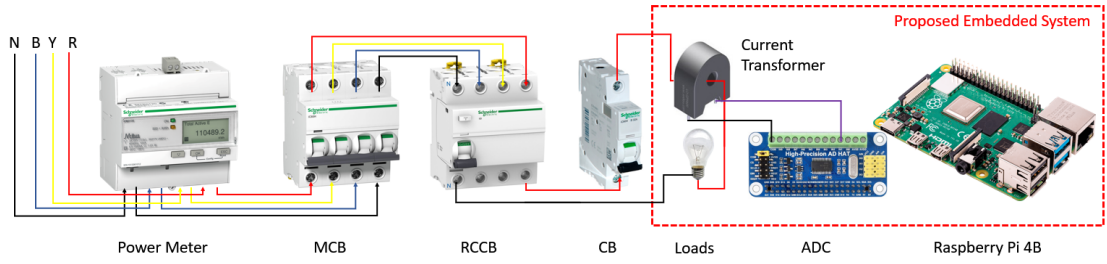


Figure 6.1: Proposed hardware implementation (embedded system).

240VAC. The voltage or current transformer used should step down the voltage or current to $\pm 2.5\text{VDC}$ before inputting into RPi with an operating range of 5VDC. Figure 6.2 demonstrates the hardware-software linkage for PQD detection and classification. The implementation of real-time detection and classification of PQDs mainly involves three steps, data acquisition, signal pre-processing, and model inferencing with a classification model. The real-world current or signal waveform is acquired using a current transformer or voltage transformer. The analogous signal is sampled via ADC which converts analog signals to digital signals. This conversion is essential for digital signal processing. The sampled data is pre-processed before passing it into the model for classification. The pre-processing process includes down-sampling the sampled signal into 3.2kHz, magnitude normalization to -1 to $+1$ range. This process is done to ensure fixed data size and range are input for the classification model. Data shifting is also included in the proposed model, where the phase of the signal collected is always started at zero.

The proposed efficient wavelet-based convolutional transformer (EWT-ConvT) is taken as the classification model due to its smaller size of the model and faster inferencing. EWT-ConvT takes in the processed data from the signal pre-processing

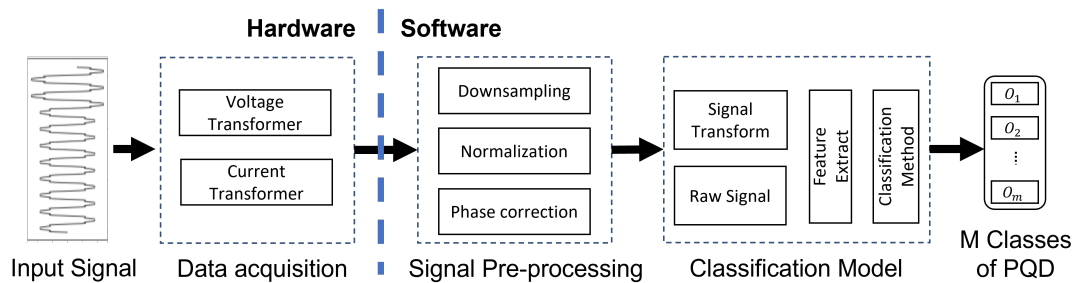


Figure 6.2: Hardware-software integration.

Table 6.1: 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

step. EWT-ConvT model is depicted as shown in Fig. 6.3. The signal is first transformed into a wavelet domain using 4-level multi-level signal decomposition. A temporal aligned layer and efficient superposition embedding mechanism is used to embed the wavelet coefficients into a temporal embedding matrix. 1D convolution is then applied over the temporal embedding matrix, giving spatial characteristics among the extracted bands. The embedded temporal-spatial feature output is then fed into the transformer network for the classification process via a multi-head attention mechanism. This proposed network is trained using AMD Ryzen 7 3800X 8-core Processor with Nvidia P6000 graphic processing unit. In Table 6.1, a total of 16 classes of PQD are simulated using mathematical equations [82, 96] for the training process.

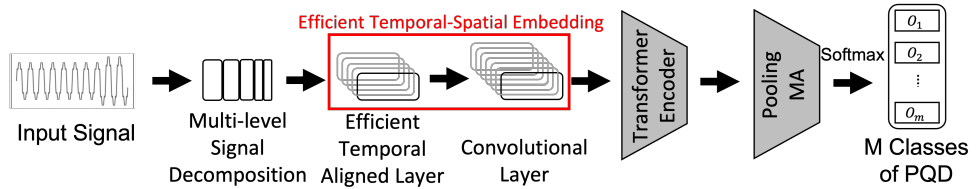


Figure 6.3: Proposed efficient wavelet-based convolutional transformer (EWT-ConvT) model.

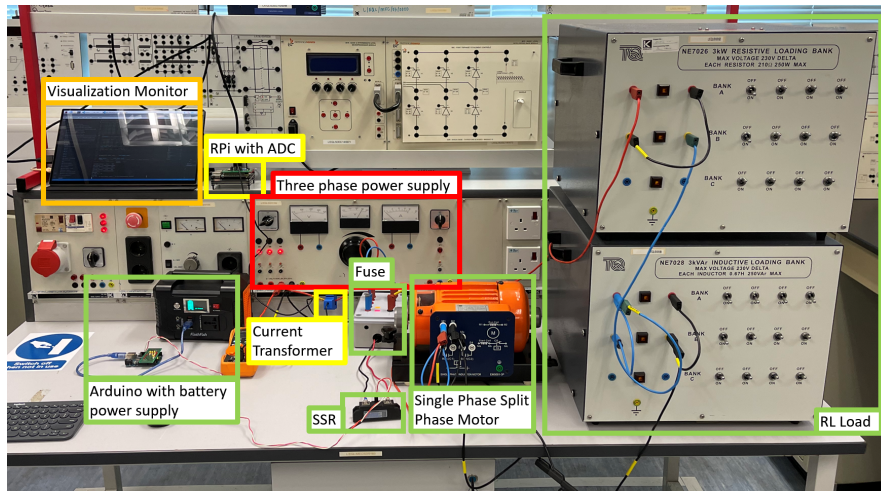


Figure 6.4: Lab Hardware setup implementation for the proposed classification model.

6.3 Laboratory Hardware and Experiment Setup

The validation of the proposed hardware setup and ML models are conducted using the laboratory setups as shown in Fig. 6.4. Two types of loads are used, 300 Watt single-phase split phase induction motor, and a resistive inductive (RL) load with 221.2 Ohm resistance and 116.2 milli Henry Inductance connected in series. The proposed real-time system setup consists of Raspberry Pi 4B, ADC, and a current transformer. The components required to carry out this laboratory experiment include single phase induction motor, resistor and inductor banks, an independent power supply unit, a fuse box for protection, a solid state relay (SSR) for fast switching, and a battery-powered Arduino micro-controller for relay mode switching control. Power disturbances are simulated by switching the mentioned loads using a solid-state relay. A total of three case studies are investigated under real-world lab simulations. The details of the case studies are listed as follows.

- (a) **(Case #A):** Voltage sag is among the most problematic PQD other than improper grounding [157]. The simulation of a short circuit fault that causes a voltage sag event is simulated in this event by using an RL load and a split phase induction motor. A sag condition is simulated by turning off the RL load for 80 milliseconds within 200 milliseconds time frame, which is equivalent to 4 out of 10 periods of sag for a 50Hz 10-period signal sample.
- (b) **(Case #B):** Disconnection of large load or large capacitor banks causes voltage

swell [158]. Disturbance class swell is simulated using a non-linear RL load and a split-phase induction motor. Different from a Sag condition, only the induction motor is turned on at all times. A swell condition is simulated by switching on the RL load for 80 milliseconds, which is equivalent to 4 out of 10 periods of swell for a 50Hz 10-period signal sample.

- (c) **(Case #C):** Short interruptions are caused by triggering of protection devices after a fault event [159]. The simulation on an interrupt signal is done by switching only the RL load. RL load is switched off for 80 milliseconds to generate the interrupt condition. This is equivalent to 4 out of 10 periods of interrupt signal for a 50Hz 10-period signal sample.

The analysis of the real-world hardware-software integration is carried out in a laboratory environment. The hardware is set up as demonstrated in Fig. 6.4. Three types of disturbances are simulated using the lab equipment, P1-Sag, P2-Swell, and P3-Interrupt. Data collection and model inferencing are performed using RPi. Two main analyses are performed. In Section 6.5, the classification performance of the proposed EWT-ConvT is analysed. Several comparisons are conducted to evaluate the models such as the proposed WT-SFA-LSTM, WT-ConvT, EWT-ConvT, Deep LSTM [101] and Deep CNN [20] in terms of accuracy and computational performance using RPi embedded system.

6.4 Data Acquisition and Pre-processing

The simulation of the disturbance signal is collected using a current transformer. The real-world analogous current waveform is first scaled down and converted into a voltage signal using an appropriate burden resistor. As RPi only takes in positive voltage, a biasing voltage divider is used. The current sensor output is fed into the middle of the biasing voltage divider before feeding into ADC. Signal sampling using ADS1263 ADC can be up to 38kHz, however, high sampling requires more computation resources and may result in data loss when the computation resources are insufficient. Besides, the sampling rate of ADS1263 ADC is fluctuating between 5-20% from the defined sampling rate. For example, the sampling rate set at 4800Hz fluctuates between 4000-4800Hz. To solve the mentioned issue, a sufficiently low sampling rate of 7200Hz

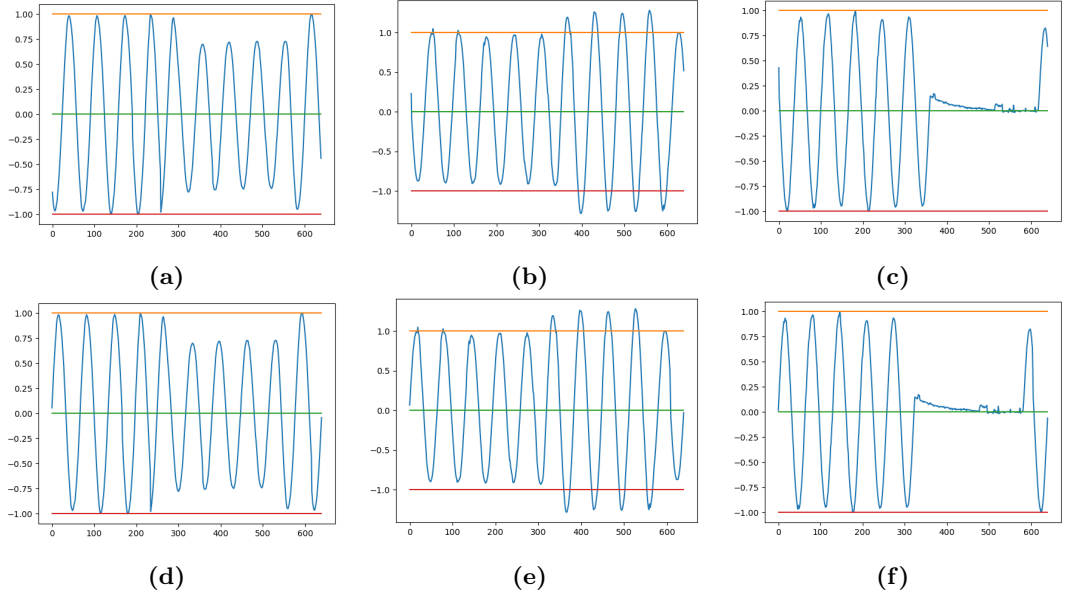


Figure 6.5: Real world lab-simulated PQDs waveforms. (a) Sag, (b) Swell, (c) Interrupt, (d) Phase-aligned Sag, (e) Phase-aligned Swell, (f) Phase-aligned Interrupt.

is used, thus reducing computation requirements. The sampled signal is then down-sampled into exact 3200Hz by re-sampling the collected data using the interpolation technique [160]. Each collected data sample consists of 10 periods and is sampled at 3200Hz which is equivalent to 640 data points.

Shift variance is one of the issues faced by wavelet transform and neural network [100, 161]. Shift variance is examined in this experiment. A simple data augmentation or phase alignment is performed by aligning zero starting of the signal. This data augmentation is done by snipping data before first rising zero crossing and appending the snippet to the back of the signal. This data augmentation is applied to align the signal with zero starting while retaining common features to avoid shift variance issues. The collected disturbances data are displayed in Fig. 6.5a, Fig. 6.5b, and Fig. 6.5c without any phase alignment. This situation causes the inconsistency of signal initialization. On the other hand, the respective aligned signals are displayed in Fig. 6.5d, Fig. 6.5e, and Fig. 6.5f to standardize the common rising edge at zero crossing to get rid of phase shifting issue.

6.5 Real-time PQD Evaluation and Performance Analysis

A total of 900 samples (300 samples for each type) of disturbances are collected in this lab. The classification performance of the proposed EWT-ConvT and other models are tabulated in Table 6.2. The effect of shift variance is examined in this experiment. The result in Table 6.2 shows that the Deep CNN model scored higher average classification accuracy of 52.67% without a phase-aligned mechanism. However, the Deep CNN model performed badly for swell and interrupt classes, which is 49.67% and 9% respectively although it has a high classification rate for sag case with 99.33% classification on class P1-Sag, It can also be noted that the classification performance on class P3-Interrupt is extremely low due to the dissimilarity of interrupt signal generated in lab condition as compared to the signal generated using mathematical equations. The interrupt produced in lab condition has an inductance effect from the RL load. The inductor resists the change of current, forming the decreasing slope of current when the power is switched off. This result shows that the Deep CNN model is not generalised enough when there are some changes and differences in the signal waveform compared to the training condition. When the phase-aligned mechanism is introduced, the Deep CNN model shows an insignificant change in classification performance, with 98.33%, 44.00%, and 9.33% on P1, P2, and P3 respectively. This shows that the Deep CNN model is shift-invariant. Apart from the Deep CNN model, other models contain shift-variant characteristics, and it can be solved by applying a phase-aligned mechanism to correct the signal synchronization issue. While achieving shift-invariant has its advantage of minimising pre-processing step and preventing data augmentation, higher-order

Table 6.2: Classification performance on laboratory simulated real-world data.

Models	Without Phase-Aligned				With Phase-Aligned			
	P1	P2	P3	Ave	P1	P2	P3	Ave
Deep CNN	99.33	49.67	9.00	52.67	98.33	44.00	9.33	50.56
Deep LSTM	30.33	21.00	48.00	33.11	98.33	65.67	52.33	72.11
WT-SFA-LSTM	34.00	6.00	42.33	27.44	78.00	41.67	77.33	65.67
WT-ConvT	34.67	7.67	45.00	29.11	91.33	46.67	83.33	73.78
EWT-ConvT	56.67	8.33	62.33	42.44	99.00	48.33	77.67	75.00

feature representation might be required for significant feature extraction.

After introducing a phase-aligned mechanism for the input pre-processing, the performance of all the models are comparable to Deep CNN model as shown in Table 6.2. This shows that the models possess a shift-variant characteristic. A phase-aligned mechanism is required at the signal pre-processing stage when using these models. Among the models, the proposed EWT-ConvT scored the highest average accuracy of 75.00%. This is followed by the proposed WT-ConvT model, with 73.78%. The proposed WT-ConvT and EWT-ConvT models are thus verified in terms of performance with real-world data. It can be noticed that the Deep LSTM model acquired high accuracy of 72.11%. Comparing WT-ConvT models with the Deep LSTM model shows that the Deep LSTM model performs better on class P2-Sag, however, the classification performance on class P3 is comparatively weaker. Wavelet transform and spatial feature attention from the WT-SFA-LSTM model improved the performance on class P3. However, the WT-SFA-LSTM model does not perform well in real-time scenarios to detect P1 and P2 signals.

The confusion analysis of the models is tabulated in Table 6.3. Deep CNN model in Table 6.3a shows a similar pattern with and without a phase-aligned mechanism. Class P1 achieved high classification accuracy with over 99.33% accuracy. The wrongly classified cases are confused with class P14-Flicker+Sag. This can be due to the sampling which produced a slight flicker effect on the sampled signal. The classification accuracy on class P2-Swell is however having higher confusion of 168/300 with class P-15 Flicker+Swell. There are two possible reasons for the less performance. The first reason can be caused by less quality signal sampled which shows a slight flicker effect as shown in Fig. 6.5e. The second reason for that can be caused by the model being too sensitive towards the slight flicker effect. This effect can be solved with a better sampling technique which is discussed in the later section. The classification performance of P3-Interrupt is however seriously impacted and has major confusion of 173/300 on P1, 95/300 on P14 and even 3/300 on class P0-Normal. The interrupt signal used for training does not have a similar micro-pattern, causing confusion for the Deep CNN model. This modified waveform can be formed with combined characteristics of different electronic components. This confusion shows that the Deep CNN model is not generalised enough to identify a slightly modified signal pattern that it has not seen before during the training phase.

Both LSTM models in Table 6.3b and Table 6.3c produce high confusion on signal without zero-aligned phase shift pre-processing. This shows that the models

Table 6.3: Confusion analysis on four models.

(a) Deep CNN.

Phase Aligned	Without			With		
Classes	P1	P2	P3	P1	P2	P3
P0	0	0	3	0	0	3
P1	298	0	177	295	0	173
P2	0	149	0	0	132	0
P3	0	0	27	0	0	28
P4	0	0	0	0	0	0
P5	0	0	0	0	0	0
P6	0	0	0	0	0	0
P7	0	0	0	0	0	1
P8	0	0	0	0	0	0
P9	0	0	0	0	0	0
P10	0	0	0	0	0	0
P11	0	0	0	0	0	0
P12	0	0	0	0	0	0
P13	0	0	0	0	0	0
P14	2	0	93	5	0	95
P15	0	151	0	0	168	0
Acc (%)	99.33	49.67	9.00	98.33	44.00	9.33
Ave Acc (%)	52.67			50.56		

(b) Deep LSTM.

Phase Aligned	Without			With		
Classes	P1	P2	P3	P1	P2	P3
P0	0	0	0	0	0	1
P1	91	0	38	295	0	129
P2	0	63	0	0	197	0
P3	0	0	144	0	0	157
P4	0	0	0	0	0	0
P5	64	0	37	1	0	4
P6	0	56	0	0	2	0
P7	0	0	2	0	0	0
P8	0	0	0	0	0	0
P9	0	0	0	0	0	0
P10	0	0	0	0	0	0
P11	0	0	0	0	0	0
P12	0	0	0	0	0	0
P13	1	10	0	0	1	0
P14	144	106	79	4	3	9
P15	0	65	0	0	97	0
Acc (%)	30.33	21.00	48.00	98.33	65.67	52.33
Ave Acc (%)	33.11			72.11		

(c) WT-SFA-LSTM.

Phase Aligned	Without			With		
Classes	P1	P2	P3	P1	P2	P3
P0	0	0	0	0	0	0
P1	102	61	38	234	43	22
P2	1	18	3	0	125	0
P3	30	37	127	3	6	232
P4	0	0	0	0	0	0
P5	0	0	0	0	0	0
P6	0	0	0	0	0	0
P7	2	0	12	0	0	1
P8	2	5	7	0	6	7
P9	0	0	0	0	0	0
P10	0	0	0	0	0	0
P11	0	0	0	0	0	0
P12	0	0	0	0	0	0
P13	0	0	0	0	0	0
P14	158	176	112	63	92	38
P15	5	3	1	0	28	0
Acc (%)	34.00	6.00	42.33	78.00	41.67	77.33
Ave Acc (%)	27.44			65.67		

(d) EWT-ConvT.

Phase Aligned	Without			With		
Classes	P1	P2	P3	P1	P2	P3
P0	0	0	0	0	0	0
P1	170	99	108	297	19	61
P2	0	25	5	0	145	6
P3	130	149	187	3	5	233
P4	0	0	0	0	0	0
P5	0	0	0	0	0	0
P6	0	0	0	0	0	0
P7	0	2	0	0	1	0
P8	0	0	0	0	0	0
P9	0	0	0	0	0	0
P10	0	0	0	0	0	0
P11	0	0	0	0	0	0
P12	0	0	0	0	0	0
P13	0	0	0	0	0	0
P14	0	18	0	0	107	0
P15	0	7	0	0	23	0
Acc (%)	56.67	8.33	62.33	99.00	48.33	77.67
Ave Acc (%)	42.44			75.00		

are impacted by shift-variant characteristics. Shift variants can be solved by using phase-aligned data augmentation. Besides, the phase-shift problem can also be solved by giving shifted signals during the training phase. Both models showed significant improvement with phase-aligned pre-processing introduced. The deep LSTM model shows good classification performance on class P1 with 98.33% accuracy. Similar to the Deep CNN model, there is confusion between class P2 with class P15. However, Deep LSTM is showing less confusion of only 97/300 confusion in class P15. This shows that the Deep LSTM model is better at identifying small magnitude differences in the flicker signal. The classification on class P3 is also better compared to the Deep CNN method with 52.33% accuracy. Major confusion of 129/300 occurred with P1-Sag. This can be due to unseen noise conditions as described before, but Deep LSTM is showing more generalization capability compared to the DCNN method. WT-SFA-LSTM method on the other hand showing weaker performance compared to Deep LSTM. Confusion occurred in class P1 with class P14, where 63/300 signals are misclassified as class P14. Besides that, Class P2 are also having 92/300 signals misclassified as P14, 43/300 as P1 and 28/300 as P15. This shows that WT-SFA-LSTM is extremely sensitive to smaller magnitude change which causes major confusion to class P14 with the flicker effect. The classification performance in class P3 is however better with high 77.33% accuracy. The high sensitivity on magnitude changes allowed this model to detect interrupt classes with ease.

The classification performance of the EWT-ConvT model is tabulated in Table 6.3d. Similar to LSTM models, EWT-ConvT is also a shift-variant model that requires the aid of pre-process phase alignment. EWT-ConvT model attained the highest average classification accuracy of 75.00% among all the models compared. While having a good classification of 99.00% on class P1, EWT-ConvT is also having the highest classification accuracy of 77.67% on class P3. The classification performance of class P2 is however having high confusion of 107/300 with class P14. This confusion again shows that the models have high sensitivity towards the flicker effect caused by the low-quality signal collected. A better sampling technique should be applied to overcome this issue.

6.6 Computation Performance Analysis of Proposed Models on Embedded System

With the increasingly powerful embedded system being introduced, a real-time classification based on the embedded system can be executed. The advantage of deploying real-time classification models using embedded systems includes comparatively low cost versus dedicated PC, feasibility, and scalability. RPi 4B is used in the proposed system as it allows real-time inferencing with its powerful Quad-core Cortex-A72 (ARM v8 64-bit SoC 1.5GHz). Multithreading is used to allow parallel processing between data grabbing, model inferencing, and output data saving. Inferencing using proposed models and literature models are performed and the computation performances are recorded in Table 6.4. In this experiment, the current waveform of the test subject is sampled into RPi with the aid of a current transformer and ADC. For every 0.2 seconds (equivalent to 10 periods of a 50Hz signal), the data is pre-processed and fed into the classification model. Real-time inferencing can be achieved if the proposed model can finish the classification process using less time as compared to the input sample. For instance, the proposed classification model should complete its classification process before the next sample is ready, which is <200ms with our setup.

In Table 6.4, it can be noticed that the Deep CNN model with parallel processing architecture used 85.12ms to classify a single data input. The parallel computation of Deep CNN allows fast inference and can achieve real-time classification. The computation resources required for the model is 0.66 Megabytes, which is considerably small. However, the average classification performance in the laboratory test is only 50.56%. The deep LSTM model on the other hand uses an average of 262.06ms to classify a

Table 6.4: Model performance on RPi.

Models	Inference Time (ms)	Model Size (MB)	Best Acc (%)
Deep CNN	85.12	0.66	50.56
Deep LSTM	262.06	0.14	72.11
WT-SFA-LSTM	13.00	0.87	65.67
WT-ConvT	86.82	0.73	73.78
EWT-ConvT	75.51	0.29	75.00

single input sample. Although the Deep LSTM model can achieve a high classification accuracy of 72.11% with a small 0.14 MB resource requirement, it failed to meet the requirement of <200ms for real-time classification. The sequential processing of LSTM slows down the inferencing process. WT-SFA-LSTM is proposed to improve the input resolution and increase computation efficiency. WT-SFA-LSTM is the fastest model which takes only 13.00 ms to classify a single input. This model however comes with a slightly larger model size of 0.87 MB, and with lower average classification accuracy of 65.67%. WT-ConvT is proposed with increased resolution by wavelet transform, and the advantage of parallel processing via multihead attention mechanism within transformer architecture. Results show that WT-ConvT can perform real-time classification with 86.82ms per sample classification. The average classification accuracy of 73.78% also outperforms other methods. The model size of WT-ConvT is however slightly larger compared to the Deep CNN method. EWT-ConvT is proposed to optimize the predecessor WT-ConvT in terms of model size and performance. The weight superposition introduced reduces the number of parameters required and thus increases the model efficiency. EWT-ConvT can achieve real-time classification with only 75.51ms per sample classification. The model size of EWT-ConvT is also significantly reduced from its predecessor, with only 0.29 MB. The classification performance of EWT-ConvT tops the other model with 75.00% accuracy.

There are some hardware limitations found during the experiment. The current transformer has a clip-on non-intrusive design. This design allows easy installation of it onto any wire of the instrument. However, the user must ensure proper clipping of the instrument such that both sides of the split iron core of the sensor are in tight contact. Loose contact between the iron core results in high noise waveform which does not favor the inferencing requirement. Besides that, The communication between ADS1263 ADC and RPi uses Serial Peripheral Interface and c programming language. ADS1263 can sample up to 38 kHz signal when the sampling script is run. However, when the inferencing model is run in parallel, data loss might occur. The data loss that occurred might affect the sample collected, which changes its shape to become a non-sinusoidal waveform. This might cause complete failure of any of the classification models applied. The solution for that is to lower the sampling rate to either 4800Hz or 7200Hz which significantly reduces the computation load. Running the sampling using another separate microcontroller and sending the complete data sample instead

of live feed to the microcontroller running inferencing is also a way to avoid this problem besides getting a more powerful microcontroller such as Nvidia Jetson.

6.7 Chapter Summary

Hardware-software integration using the embedded system Raspberry Pi 4B (RPi) is implemented to perform real-world PQD classification. A total of three PQD scenarios are simulated under laboratory conditions. These include a voltage sag event caused by short-circuit faults, a swell event triggered by the disconnection of large loads, and an interruption caused by a fault event that is recovered with protection devices. The PQD signal is captured using a non-intrusive current transformer sensor, followed by analog-to-digital conversion using ADS1263 ADC. The digital signals received in RPi are pre-processed before passing into the proposed EWT-ConvT model for classification. Signal downsampling, signal normalization and phase correction are performed in the pre-processing stage. EWT-ConvT achieves the highest average classification accuracy of 75.00% on the lab-simulated data. Besides, the EWT-ConvT model is proven to be able to perform real-time PQD classification with 75.51 ms inference time on a 200ms sample. EWT-ConvT also stands out as the smallest model to achieve real-time classification with a model size of 0.29 MB. Analysis of classification results shows a negative impact from the sampling process that caused flickering characteristics. This issue can be improved by proper handling of the clip-on non-intrusive current transformer. The computation power of RPi is also degraded due to the high-speed sampling process. For future work, signal sampling optimisation should be performed. A dedicated DSP is also suggested for the sampling process. Alternatively, a more powerful microcontroller such as Nvidia Jetson which comes with CUDA cores can also be considered for better performance.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The advancement of smart grid systems is essential to optimize resource usage as the growth of power networks is exponentially increased. The complexity of the power grid increased drastically with the blooming of renewable energy generations and the mature adoption of electric vehicles. To maintain the stable functionality of the power system, an active monitoring and control scheme is highly needed to carry out instant mitigation action on any occurring faults. This process is usually performed with the use of advanced metering infrastructure which collects and monitors the real-time power quality. The identification of PQD signifies the fault that occurred on the systems. Most of the studies focus on PQD classification in simulations and on historical data. However, due to the advancement in integrated circuits especially with the increased computation performance, complex operations can now be performed on smaller computation units. In addition, the increasing maturity in the field of machine learning enables the implementation of classification models with better efficiency and smaller model size. An increasing number of monitoring nodes improve the pinpoint of the exact fault location which allows better management and control of the power system. Hence, a real-time implementation of PQD classification with a fast detection rate is required in localizing the disturbance in power signal quality for a reliable power network.

The study on the classification of PQD can be categorised into three main methods, knowledge-based method, model-based method, and hybrid method. The

knowledge-based system is used traditionally which requires the involvement of professional knowledge for manual feature extraction. Model-based method automates the feature selection process but requires large training set to achieve generalisation. The hybrid model combines the advantages of knowledge-based and model-based methods by having more control over the feature extraction process with the use of signal transformation tools. However, the two-step process increases the complexity of the model design. Real-time classification of PQD is crucial to carry out instant mitigation actions to reduce downtime of the power system. However, Limited studies are found deploying the PQD classifier into dedicated hardware, especially onto a dedicated embedded system. Hence, this research work focused on designing an automatic PQD classification model that can be implemented into an embedded system for real-time implementation.

The complexity of using hand-crafted features in the classification of PQDs involves the complex development of the model. To automate the classification process, Model-based classification models are introduced. However, the model-based classification model requires a huge amount of training data to achieve better generalization. The global attention mechanism is thus proposed in Chapter 3 to improve the LSTM model with a more generalised PQD classification. Global attention is used to extract significant features representing the input signal. The use of attention is proved to have similar features extracted when tested with noisy and noiseless conditions. The classification performance is compared between the global attention Long Short-term memory network (GA-LSTM) and LSTM model by testing both networks with additive white Gaussian noise with different signal-to-noise ratios. Results show GA-LSTM achieve better performance with 66.67% and 51.33% when the network is trained with and without noise respectively. Although LSTM without attention mechanism scored higher accuracy of 84.87% when trained with noise, however, the low accuracy of 18.99% when the network is trained without noise signifies the network only performed on seen condition. When both networks are tested using noises that were not previously seen, GA-LSTM gives better classification performance on both trains with and without noise experiments. This result shows that the attention mechanism helps improve the generalization capability of the LSTM classifier. However, it is noted that the GA-LSTM model is insensitive to short-time impulse and small-magnitude changes.

Time-series input is limited to single-dimensional analysis. The use of signal transformation enables signal analysis in different domains. Multi-resolution signal decomposition (MSD) using wavelet transform is thus introduced to increase the signal

resolution into multiple frequency bands. A hybrid model using MSD and LSTM is introduced in Chapter 4 to improve the classification performance of the LSTM model. Four levels of decomposition are performed using Daubechies 4 wavelet to increase the signal resolution into 5 frequency ranges. Discrete wavelet transform (DWT) is performed on the low pass output of each decomposition level. Time-series analysis is performed using a single-period signal waveform. Global attention and band attention mechanisms are studied, and the results suggest the use of global attention for time-series signals is better compared to band-attention mechanisms with better noise performance. The introduction of MSD and global attention mechanism on LSTM improves the classification performance of the LSTM model from 18.99% to 50.71% when tested with high noise 20dB SNR AWGN. On the other hand, multi-resolution spatial attention LSTM (WT-SFA-LSTM) is proposed to classify multi-period PQD signals. Instead of using inverse wavelet transform, the decomposed wavelet coefficients are aligned into same-length feature representation before passing into LSTM for higher-order feature extraction and classification. Results show major improvement of the proposed WT-SFA-LSTM with 93.79% when benchmarking deep LSTM and Deep CNN models with 88.48% and 90.56% accuracy respectively. A total of two experiments are set up in this chapter. Results show better performance with the use of wavelet coefficients instead of performing an inverse wavelet transform on the signal back into a time-series signal. This is achieved with the introduction of the feature aligns layer. Besides that, the WT-SFA-LSTM model is proved to have lower computation complexity with the halved training time requirement as compared to the Deep LSTM model. However, the model size of the WT-SFA-LSTM model is comparatively larger than the state-of-the-art Deep CNN model.

Transformer architecture is proposed in Chapter 5 to improve the LSTM model with parallel processing capability. Instead of using a global attention mechanism, the multi-head attention within the transformer network allows multiple features embedding with better extraction of the sequential features. As a result, the transformer encoder network shows better classification performance of 76.06% and 55.56% when the model is trained with and without noise respectively on single-period data. This result shows better noise immunity as compared to the LSTM model with 82.62% and 13.62% when tested with 20dB SNR AWGN. The transformer encoder network shows similar generalization characteristics when compared to GA-LSTM as discussed previously. The benchmarking result with the literature model using multi-period signal input shows

that the transformer encoder network has a better classification performance of 91.43% as compared to 88.48% on the Deep LSTM model. Further optimization of the model is performed by introducing MSD to increase input resolution. A Hybrid model of a wavelet-based convolutional transformer network (WT-ConvT) is thus proposed to improve classification performance. The temporal-spatial embedding layer is proposed to extract temporal and spatial features from the MSD coefficients. This embedding layer is achieved with a temporal aligned layer for feature alignment and a layer of the 1D convolution kernel. Subsequently, the extracted features are passed into the transformer encoder layer to extract higher-level latent representation. Finally, the latent representations are aggregated by a multi-head attention pooling mechanism for PQD classification. Results show significant improvement when tested with 20dB high noise AWGN, with 94.11% as compared to the Deep CNN model with 90.56%. WT-ConvT is further improved by optimizing the temporal aligned layer with an efficient embedding layer with a weight superposition algorithm. The efficient embedding WT-ConvT model (EWT-ConvT) reduces 61.31% of the model size while having similar classification performance with 94.42% classification accuracy when tested with 20dB SNR AWGN. The optimization performed in this chapter contributed to our best-performing classification model, EWT-ConvT with the highest classification accuracy of 94.42% with the lowest model size of 0.29 MB when compared with literature models.

Real-time classification of PQD is important in reducing downtime of power systems. A real-world implementation of the proposed automatic PQD classification is proposed in Chapter 6 using an embedded system. This proposed system consists of Raspberry Pi 4B as a microcontroller, an analog-to-digital converter, and a current transformer for data acquisition. A total of three case scenarios are simulated under lab conditions. These include the short-circuit faults causing voltage sag, disconnection of large loads which causes voltage swell, and short interruption with protection devices which causes an interruption. The result shows that our proposed EWT-ConvT achieves the highest classification performance with an average accuracy of 75%. EWT-ConvT achieves real-time classification by taking 75.51ms to classify a 200ms input sample. the proposed model has a great result with the smallest model size of 0.29MB. There is also a limitation found on the RPi use, where a high sampling rate causes data loss on the signal sampled. The problem can be minimized with the use of a lower sampling rate.

7.2 Future Work

For future work, the time location of the original input signal is augmented with the signal pre-processing phase-aligned mechanism used in Chapter 6. A better alternative to DWT, dual-tree complex wavelet transform (DTCWT) is proposed by Kingsbury et al. to achieve shift invariant [162]. DTCWT provides better directional selectivity with a limited increase in computation. Besides, multi-level DTCWT can achieve shift invariance by removing the downsampling in the MSD process. Besides changing the MSD DWT to MSD DTCWT, further research can be performed by feeding shifted training samples into the proposed EWT-ConvT. The integration of the CNN layer before the transformer is one of the current trends to increase the performance because of the advantage of the Deep CNN model [20] having shift invariant properties.

On the other hand, the proposed model can be further improved by using better optimization techniques. The new two-phase optimization strategy such as Stochastic diagonal Approximate Greatest Descent (SDAGD) can be used to improve the training process with faster convergence and prevent vanishing gradient problems [163]. SDAGD adopted the concept of a control system with two stages of control [164]. The search iteration is defined at the boundary of the local search region when a minimum point is absent, and the Newton method is used to find the optimum solution when the minimum point is present in the search region [165].

The hardware implementation of the proposed automated classification of PQD can still be improved in many ways. A direct improvement that can be found in our proposed method is the use of a dedicated waveform sampling controller. Real-time signal grabbing requires a lot of computation resources for the IO. A low-level IO programming can be used to optimise the computation resources used for the sampling process. Besides, the use of a dedicated digital signal processor (DSP) can free up the computation resources on the main processor for the machine learning algorithm. Recently, more powerful low-cost microprocessors are introduced. These include a thumb size Espressive (ESP) microcontroller and RockPi micro-controllers which support the implementation of AI algorithms. Research can be done with real-time implementations on these platforms. Other than that, FPGA can be used to design the dedicated processor required for the sampling as well as ML processing. Mass production on application-specific integrated circuits (ASIC) consisting of DSP and ML processors will surely reduce the implementation cost and complexity.

This research can be extended with more real-world case studies. This can be done by using a laboratory PQD simulator such as Megger Power Quality Simulator for transient, harmonics, and phase-shift issues. In addition, the harvest of real-world PQD data can lead to a better training mechanism in the machine learning model. The power quality disturbance detection and classification algorithm can be used in high-power industries. The aluminium smelting process is an energy-intensive industry demand for liable power supplies [166]. Besides ensuring stable power supplies via constant power monitoring on the supplies side, automatic power quality disturbance detection and classification can be further applied to the direct current network for the smelting and refining process. The constant monitoring of the individual anode allows monitoring of the cell conditions of the smelter [167, 168].

References

- [1] J. Twidell, *Renewable energy resources*. Routledge, 2021.
- [2] G. Energy, “Co2 status report,” *IEA (International Energy Agency): Paris, France*, 2019.
- [3] V. Nikam and V. Kalkhambkar, “A review on control strategies for microgrids with distributed energy resources, energy storage systems, and electric vehicles,” *International Transactions on Electrical Energy Systems*, vol. 31, no. 1, p. e12607, 2021.
- [4] K. M. Tan, T. S. Babu, V. K. Ramachandaramurthy, P. Kasinathan, S. G. Solanki, and S. K. Raveendran, “Empowering smart grid: A comprehensive review of energy storage technology and application with renewable energy integration,” *Journal of Energy Storage*, vol. 39, p. 102591, 2021.
- [5] H. Ekhteraei Toosi, A. Merabet, and A. Swingler, “Dual-layer power scheduling strategy for ev-ess-controllable load in bi-directional dynamic markets for low-cost implementation,” *International Transactions on Electrical Energy Systems*, vol. 31, no. 1, p. e12681, 2021.
- [6] S. B. Slama, “Design and implementation of home energy management system using vehicle to home (h2v) approach,” *Journal of Cleaner Production*, vol. 312, p. 127792, 2021.
- [7] R. H. Lasseter and P. Paigi, “Microgrid: A conceptual solution,” in *2004 IEEE 35th annual power electronics specialists conference (IEEE Cat. No. 04CH37551)*, vol. 6. IEEE, 2004, pp. 4285–4290.

-
- [8] X. Lu, S. Xia, G. Sun, J. Hu, W. Zou, Q. Zhou, M. Shahidehpour, and K. W. Chan, "Hierarchical distributed control approach for multiple on-site ders coordinated operation in microgrid," *International Journal of Electrical Power & Energy Systems*, vol. 129, p. 106864, 2021.
- [9] L. Zhang, N. Gari, and L. V. Hmurcik, "Energy management in a microgrid with distributed energy resources," *Energy Conversion and Management*, vol. 78, pp. 297–305, 2014.
- [10] E. A. Soto, L. B. Bosman, E. Wollega, and W. D. Leon-Salas, "Peer-to-peer energy trading: A review of the literature," *Applied Energy*, vol. 283, p. 116268, 2021.
- [11] A. Chandra, G. K. Singh, and V. Pant, "Protection of ac microgrid integrated with renewable energy sources—a research review and future trends," *Electric Power Systems Research*, vol. 193, p. 107036, 2021.
- [12] M. R. Khalid, I. A. Khan, S. Hameed, M. S. J. Asghar, and J.-S. Ro, "A comprehensive review on structural topologies, power levels, energy storage systems, and standards for electric vehicle charging stations and their impacts on grid," *IEEE Access*, vol. 9, pp. 128 069–128 094, 2021.
- [13] L. Wang, Z. Qin, T. Slangen, P. Bauer, and T. van Wijk, "Grid impact of electric vehicle fast charging stations: Trends, standards, issues and mitigation measures—an overview," *IEEE Open Journal of Power Electronics*, vol. 2, pp. 56–74, 2021.
- [14] I. M. Mehedi, M. Ahmadipour, Z. Salam, H. M. Ridha, H. Bassi, M. J. H. Rawa, M. Ajour, A. Abusorrah, and M. P. Abdullah, "Optimal feature selection using modified cuckoo search for classification of power quality disturbances," *Applied Soft Computing*, vol. 113, p. 107897, 2021.
- [15] O. P. Mahela, A. G. Shaik, and N. Gupta, "A critical review of detection and classification of power quality events," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 495–505, 2015.
- [16] P. Dash, M. Padhee, and T. Panigrahi, "A hybrid time–frequency approach based fuzzy logic system for power island detection in grid connected distributed generation," *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 453–464, 2012.

-
- [17] T. Karthick, K. Chandrasekaran *et al.*, “Design of iot based smart compact energy meter for monitoring and controlling the usage of energy and power quality issues with demand side management for a commercial building,” *Sustainable Energy, Grids and Networks*, vol. 26, p. 100454, 2021.
- [18] N. K. Singh and V. Mahajan, “End-user privacy protection scheme from cyber intrusion in smart grid advanced metering infrastructure,” *International Journal of Critical Infrastructure Protection*, vol. 34, p. 100410, 2021.
- [19] F. Al-Turjman and M. Abujubbeh, “Iot-enabled smart grid via sm: An overview,” *Future Generation Computer Systems*, vol. 96, pp. 579–590, 2019.
- [20] 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.
- [21] U. Dwivedi and S. Singh, “Denoising techniques with change-point approach for wavelet-based power-quality monitoring,” *IEEE transactions on power delivery*, vol. 24, no. 3, pp. 1719–1727, 2009.
- [22] H.-T. Yang and C.-C. Liao, “A de-noising scheme for enhancing wavelet-based power quality monitoring system,” *IEEE Transactions on Power Delivery*, vol. 16, no. 3, pp. 353–360, 2001.
- [23] U. Singh and S. N. Singh, “Optimal feature selection via nsga-ii for power quality disturbances classification,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 2994–3002, 2017.
- [24] 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, no. 5, p. 1238, 2021.
- [25] A. Yılmaz, A. Küçüker, and G. Bayrak, “Automated classification of power quality disturbances in a sofc&pv-based distributed generator using a hybrid machine learning method with high noise immunity,” *International Journal of Hydrogen Energy*, vol. 47, no. 45, pp. 19 797–19 809, 2022.
- [26] 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.

-
- [27] K. Alanne and A. Saari, "Distributed energy generation and sustainable development," *Renewable and sustainable energy reviews*, vol. 10, no. 6, pp. 539–558, 2006.
- [28] A. Dagar, P. Gupta, and V. Niranjana, "Microgrid protection: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 149, p. 111401, 2021.
- [29] M. Ahmed, L. Meegahapola, A. Vahidnia, and M. Datta, "Stability and control aspects of microgrid architectures—a comprehensive review," *IEEE Access*, vol. 8, pp. 144 730–144 766, 2020.
- [30] J. J. Moreno Escobar, O. Morales Matamoros, R. Tejeida Padilla, I. Lina Reyes, and H. Quintana Espinosa, "A comprehensive review on smart grids: Challenges and opportunities," *Sensors*, vol. 21, no. 21, p. 6978, 2021.
- [31] L. Ahmethodzic and M. Music, "Comprehensive review of trends in microgrid control," *Renewable Energy Focus*, vol. 38, pp. 84–96, 2021.
- [32] F. Garcia-Torres, S. Vazquez, I. M. Moreno-Garcia, A. Gil-de Castro, P. Roncero-Sanchez, and A. Moreno-Munoz, "Microgrids power quality enhancement using model predictive control," *Electronics*, vol. 10, no. 3, p. 328, 2021.
- [33] A. Muhtadi, D. Pandit, N. Nguyen, and J. Mitra, "Distributed energy resources based microgrid: Review of architecture, control, and reliability," *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 2223–2235, 2021.
- [34] A. Yilmaz and G. Bayrak, "A real-time uwt-based intelligent fault detection method for pv-based microgrids," *Electric Power Systems Research*, vol. 177, p. 105984, 2019.
- [35] P. Aramwanid and I. Boonyaroonate, "Power quality impact study and analysis of electrical power efficacy in sugar industry," in *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, 2015, pp. 1–4.
- [36] H. Das and L. Saikia, "Ethernet based smart energy meter for power quality monitoring and enhancement," in *2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE)*. IEEE, 2017, pp. 187–191.

-
- [37] F.-C. Argatu, V. Brezoianu, V. V. Argatu, B.-A. Enache, F.-C. Adochiei, and T. Icleanu, "Power quality analyzer for smart grid-smart home applications," in *2019 54th International Universities Power Engineering Conference (UPEC)*. IEEE, 2019, pp. 1–4.
- [38] B. Eristi and H. Eristi, "Classification of power quality disturbances in solar pv integrated power system based on a hybrid deep learning approach," *International Transactions on Electrical Energy Systems*, vol. 2022, 2022.
- [39] D. Gu, Y. Gao, Y. Li, Y. Zhu, and C. Wu, "A novel label-guided attention method for multilabel classification of multiple power quality disturbances," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4698–4706, 2021.
- [40] IEEE, "Ieee recommended practice for monitoring electric power quality," *IEEE Std 1159-1995*, pp. 1–80, 1995.
- [41] S. Upadhyaya, C. Bhende, S. Mohanty, and R. Pati, "Evaluation of power quality disturbance in pv-connected ieee-14 bus system using lifting-based wavelet transform and random forest," *Electrical Engineering*, pp. 1–10, 2022.
- [42] B. Singh, A. Chandra, and K. Al-Haddad, *Power quality: problems and mitigation techniques*. John Wiley & Sons, 2014.
- [43] R. Kapoor, R. Gupta, S. Jha, R. Kumar *et al.*, "Detection of power quality event using histogram of oriented gradients and support vector machine," *Measurement*, vol. 120, pp. 52–75, 2018.
- [44] I. D. Nicolae, A. A. Qays, P. M. Nicolae, and R. F. Marinescu, "Complex power quality analysis of electric waveforms affected by noise," in *2019 8th International Conference on Modern Power Systems (MPS)*. IEEE, 2019, pp. 1–6.
- [45] Y. Wang, Q. Li, F. Zhou, Y. Zhou, and X. Mu, "A new method with hilbert transform and slip-svd-based noise-suppression algorithm for noisy power quality monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 4, pp. 987–1001, 2018.
- [46] J. Liu and G. Ni, "An efficient optimal algorithm for high frequency in wavelet based image reconstruction," *Journal of Computational Analysis and Applications*, vol. 28, no. 5, pp. 865–878, 2020.

-
- [47] M. Coto-Jiménez, J. Goddard-Close, and F. Martínez-Licona, “Improving automatic speech recognition containing additive noise using deep denoising autoencoders of lstm networks,” in *International Conference on Speech and Computer*. Springer, 2016, pp. 354–361.
- [48] S. Santoso, W. M. Grady, E. J. Powers, J. Lamoree, and S. C. Bhatt, “Characterization of distribution power quality events with fourier and wavelet transforms,” *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 247–254, 2000.
- [49] G. Heydt, P. Fjeld, C. Liu, D. Pierce, L. Tu, and G. Hensley, “Applications of the windowed fft to electric power quality assessment,” *IEEE Transactions on Power Delivery*, vol. 14, no. 4, pp. 1411–1416, 1999.
- [50] F. Jurado and J. R. Saenz, “Comparison between discrete stft and wavelets for the analysis of power quality events,” *Electric Power Systems Research*, vol. 62, no. 3, pp. 183–190, 2002.
- [51] S. H. Jaramillo, G. Heydt, and E. O’Neill-Carrillo, “Power quality indices for aperiodic voltages and currents,” *IEEE Transactions on Power Delivery*, vol. 15, no. 2, pp. 784–790, 2000.
- [52] M. J. Afroni, D. Sutanto, and D. Stirling, “Analysis of nonstationary power-quality waveforms using iterative hilbert huang transform and sax algorithm,” *IEEE Transactions on Power Delivery*, vol. 28, no. 4, pp. 2134–2144, 2013.
- [53] J. Chen, T. Zhao, Y. Yang, and D. Zhang, “Feature extraction and evaluation of electricity load data with high precision,” in *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE, 2017, pp. 24–28.
- [54] F. Zhao and R. Yang, “Power quality disturbance recognition using s-transform,” in *2006 IEEE Power Engineering Society General Meeting*. IEEE, 2006, pp. 7–pp.
- [55] O. Cortes-Robles, E. Barocio, J. Segundo, D. Guillen, and J. Olivares-Galvan, “A qualitative-quantitative hybrid approach for power quality disturbance monitoring on microgrid systems,” *Measurement*, vol. 154, p. 107453, 2020.

-
- [56] M. Farhoumandi, Q. Zhou, and M. Shahidehpour, "A review of machine learning applications in iot-integrated modern power systems," *The Electricity Journal*, vol. 34, no. 1, p. 106879, 2021.
- [57] D. Li, T. Wang, W. Pan, X. Ding, and J. Gong, "A comprehensive review of improving power quality using active power filters," *Electric Power Systems Research*, vol. 199, p. 107389, 2021.
- [58] R. Igual and C. Medrano, "Research challenges in real-time classification of power quality disturbances applicable to microgrids: A systematic review," *Renewable and Sustainable Energy Reviews*, vol. 132, p. 110050, 2020.
- [59] M. Szmajda, K. Górecki, and J. Mroczka, "Dft algorithm analysis in low-cost power quality measurement systems based on a dsp processor," in *2007 9th International Conference on Electrical Power Quality and Utilisation*. IEEE, 2007, pp. 1–6.
- [60] T. Zhu, S. Tso, and K. Lo, "Wavelet-based fuzzy reasoning approach to power-quality disturbance recognition," *IEEE Transactions on Power Delivery*, vol. 19, no. 4, pp. 1928–1935, 2004.
- [61] C. Lee, J. Lee, J.-O. Kim, and S. W. Nam, "Feature vector extraction for the automatic classification of power quality disturbances," in *Proceedings of 1997 IEEE International Symposium on Circuits and Systems. Circuits and Systems in the Information Age ISCAS'97*, vol. 4. IEEE, 1997, pp. 2681–2684.
- [62] M. Lopez-Ramirez, E. Cabal-Yepez, L. M. Ledesma-Carrillo, H. Miranda-Vidales, C. Rodriguez-Donate, and R. A. Lizarraga-Morales, "Fpga-based online pqd detection and classification through dwt, mathematical morphology and svd," *Energies*, vol. 11, no. 4, p. 769, 2018.
- [63] 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, no. 3, pp. 296–306, 2009.
- [64] C.-T. Hsieh, J.-M. Lin, and S.-J. Huang, "Slant transform applied to electric power quality detection with field programmable gate array design enhanced,"

International Journal of Electrical Power & Energy Systems, vol. 32, no. 5, pp. 428–432, 2010.

- [65] A. Rodríguez, J. Aguado, F. Martín, J. López, F. Muñoz, and J. Ruiz, “Rule-based classification of power quality disturbances using s-transform,” *Electric power systems Research*, vol. 86, pp. 113–121, 2012.
- [66] R. Bhavani and N. R. Prabha, “A hybrid classifier for power quality (pq) problems using wavelets packet transform (wpt) and artificial neural networks (ann),” in *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*. IEEE, 2017, pp. 1–7.
- [67] M. Manjula, S. Mishra, and A. Sarma, “Empirical mode decomposition with hilbert transform for classification of voltage sag causes using probabilistic neural network,” *International Journal of Electrical Power & Energy Systems*, vol. 44, no. 1, pp. 597–603, 2013.
- [68] 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, pp. 246–259, 2017.
- [69] 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.
- [70] N. Huang, S. Zhang, G. Cai, and D. Xu, “Power quality disturbances recognition based on a multiresolution generalized s-transform and a pso-improved decision tree,” *Energies*, vol. 8, no. 1, pp. 549–572, 2015.
- [71] 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, pp. 88 380–88 392, 2019.
- [72] 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.

-
- [73] 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.
- [74] S. Naderian and A. Salemnia, "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, 2017.
- [75] 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.
- [76] T. A. Muthusamy and N. Ramanathan, "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, pp. 308–313, 2018.
- [77] C. Yan, Y. Tu, X. Wang, Y. Zhang, X. Hao, Y. Zhang, and Q. Dai, "Stat: Spatial-temporal attention mechanism for video captioning," *IEEE transactions on multimedia*, vol. 22, no. 1, pp. 229–241, 2019.
- [78] A. Zeyer, P. Bahar, K. Irie, R. Schlüter, and H. Ney, "A comparison of transformer and lstm encoder decoder models for asr," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2019, pp. 8–15.
- [79] G. Sun, C. Zhang, and P. C. Woodland, "Transformer language models with lstm-based cross-utterance information representation," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 7363–7367.
- [80] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [81] 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.

-
- [82] 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, no. 6, p. e12010, 2019.
- [83] W. L. R. Junior, F. A. S. Borges, R. d. A. L. Rabelo, B. V. A. de Lima, and J. E. A. de Alencar, "Classification of power quality disturbances using convolutional network and long short-term memory network," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–6.
- [84] Q.-m. Zhang and H.-j. Liu, "Application of ls-svm in classification of power quality disturbances," *Proceedings-Chinese Society Of Electrical Engineering*, vol. 28, no. 1, p. 106, 2008.
- [85] P. Janik and T. Lobos, "Automated classification of power-quality disturbances using svm and rbf networks," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1663–1669, 2006.
- [86] N. Mohan, K. Soman, and R. Vinayakumar, "Deep power: Deep learning architectures for power quality disturbances classification," in *2017 International Conference on Technological Advancements in Power and Energy (TAP Energy)*. IEEE, 2017, pp. 1–6.
- [87] 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.
- [88] H. Sindi, M. Nour, M. Rawa, Ş. Öztürk, and K. Polat, "A novel hybrid deep learning approach including combination of 1d power signals and 2d signal images for power quality disturbance classification," *Expert Systems with Applications*, vol. 174, p. 114785, 2021.
- [89] M. A. Rodriguez, J. F. Sotomonte, J. Cifuentes, and M. Bueno-López, "Power quality disturbance classification via deep convolutional auto-encoders and stacked lstm recurrent neural networks," in *2020 International Conference on Smart Energy Systems and Technologies (SEST)*. IEEE, 2020, pp. 1–6.

-
- [90] C. I. Garcia, F. Grasso, A. Luchetta, M. C. Piccirilli, L. Paolucci, and G. Talluri, “A comparison of power quality disturbance detection and classification methods using cnn, lstm and cnn-lstm,” *Applied Sciences*, vol. 10, no. 19, p. 6755, 2020.
- [91] A. Bagheri, I. Y. Gu, M. H. Bollen, and E. Balouji, “A robust transform-domain deep convolutional network for voltage dip classification,” *IEEE Transactions on Power Delivery*, vol. 33, no. 6, pp. 2794–2802, 2018.
- [92] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, “Power quality disturbance waveform recognition using wavelet-based neural classifier. i. theoretical foundation,” *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 222–228, 2000.
- [93] —, “Power quality disturbance waveform recognition using wavelet-based neural classifier. ii. application,” *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 229–235, 2000.
- [94] R. Kumar, B. Singh, and D. T. Shahani, “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, 2015.
- [95] Z. Chen, M. Li, T. Ji, and Q. Wu, “Detection and classification of power quality disturbances in time domain using probabilistic neural network,” in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 1277–1282.
- [96] 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, no. 11, pp. 9715–9723, 2019.
- [97] D. Lee, P. Srikantha, and D. Kundur, “Online power quality disturbance classification with recurrent neural network,” in *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (Smart-GridComm)*. IEEE, 2018, pp. 1–6.

-
- [98] M. Masoum, S. Jamali, and N. Ghaffarzadeh, "Detection and classification of power quality disturbances using discrete wavelet transform and wavelet networks," *IET Science, Measurement & Technology*, vol. 4, no. 4, pp. 193–205, 2010.
- [99] P. Kanirajan and V. S. Kumar, "Power quality disturbance detection and classification using wavelet and rbfnn," *Applied Soft Computing*, vol. 35, pp. 470–481, 2015.
- [100] R. Singh, S. Nigam, A. K. Singh, and M. Elhoseny, "Wavelet transforms: From classical to new generation wavelets," in *Intelligent Wavelet Based Techniques for Advanced Multimedia Applications*. Springer, 2020, pp. 13–29.
- [101] 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.
- [102] W. Qiu, Q. Tang, J. Liu, Z. Teng, and W. Yao, "Power quality disturbances recognition using modified s transform and parallel stack sparse auto-encoder," *Electric Power Systems Research*, vol. 174, p. 105876, 2019.
- [103] Y. Deng, L. Wang, H. Jia, X. Tong, and F. Li, "A sequence-to-sequence deep learning architecture based on bidirectional gru for type recognition and time location of combined power quality disturbance," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4481–4493, 2019.
- [104] 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.
- [105] U. Singh and S. N. Singh, "A new optimal feature selection scheme for classification of power quality disturbances based on ant colony framework," *Applied Soft Computing*, vol. 74, pp. 216–225, 2019.
- [106] C. Zhao, K. Li, Y. Li, L. Wang, Y. Luo, X. Xu, X. Ding, and Q. Meng, "Novel method based on variational mode decomposition and a random discriminative projection extreme learning machine for multiple power quality disturbance recognition," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2915–2926, 2018.

-
- [107] P. M. Ramos, T. Radil, F. M. Janeiro, and A. C. Serra, "Dsp based power quality analyser using new signal processing algorithms for detection and classification of disturbances in a single-phase power system," *Metrology and Measurement Systems*, vol. 14, no. 4, pp. 483–494, 2007.
- [108] T. Radil, P. M. Ramos, F. M. Janeiro, and A. C. Serra, "Pq monitoring system for real-time detection and classification of disturbances in a single-phase power system," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 8, pp. 1725–1733, 2008.
- [109] M. Zhang, K. Li, and Y. Hu, "A real-time classification method of power quality disturbances," *Electric power systems Research*, vol. 81, no. 2, pp. 660–666, 2011.
- [110] S. He, K. Li, and M. Zhang, "A real-time power quality disturbances classification using hybrid method based on s-transform and dynamics," *IEEE transactions on instrumentation and measurement*, vol. 62, no. 9, pp. 2465–2475, 2013.
- [111] J. Li, Z. Teng, Q. Tang, and J. Song, "Detection and classification of power quality disturbances using double resolution s-transform and dag-svms," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 10, pp. 2302–2312, 2016.
- [112] F. A. Borges, R. A. Fernandes, I. N. Silva, and C. B. Silva, "Feature extraction and power quality disturbances classification using smart meters signals," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 824–833, 2015.
- [113] E. G. Ribeiro, T. M. Mendes, G. L. Dias, E. R. Faria, F. M. Viana, B. H. Barbosa, and D. D. Ferreira, "Real-time system for automatic detection and classification of single and multiple power quality disturbances," *Measurement*, vol. 128, pp. 276–283, 2018.
- [114] M. Sahani and P. K. Dash, "Fpga-based online power quality disturbances monitoring using reduced-sample hht and class-specific weighted rvfln," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4614–4623, 2019.
- [115] —, "Deep convolutional stack autoencoder of process adaptive vmd data with robust multikernel rvfln for power quality events recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

-
- [116] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *arXiv preprint arXiv:1409.0473*, 2014.
- [117] C. Jin, Z. Shi, W. Li, and Y. Guo, “Bidirectional lstm-crf attention-based model for chinese word segmentation,” *arXiv preprint arXiv:2105.09681*, 2021.
- [118] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, “Visual attention network,” *arXiv preprint arXiv:2202.09741*, 2022.
- [119] C. Subakan, M. Ravanelli, S. Cornell, M. Bronzi, and J. Zhong, “Attention is all you need in speech separation,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 21–25.
- [120] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” *Neurocomputing*, vol. 452, pp. 48–62, 2021.
- [121] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [122] A. Luo, Q. Xu, F. Ma, and Y. Chen, “Overview of power quality analysis and control technology for the smart grid,” *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 1, pp. 1–9, 2016.
- [123] E. Hossain, M. R. Tür, S. Padmanaban, S. Ay, and I. Khan, “Analysis and mitigation of power quality issues in distributed generation systems using custom power devices,” *Ieee Access*, vol. 6, pp. 16 816–16 833, 2018.
- [124] K. Schipman and F. Delincé, “The importance of good power quality,” *ABB Power Qual. Prod., Charleroi, Belgium, ABB Review*, 2010.
- [125] F. Yazdi and S. Hosseinian, “A novel “smart branch” for power quality improvement in microgrids,” *International Journal of Electrical Power & Energy Systems*, vol. 110, pp. 161–170, 2019.
- [126] S. Choe and J. Yoo, “Wavelet packet transform modulus-based feature detection of stochastic power quality disturbance signals,” *Applied Sciences*, vol. 11, no. 6, p. 2825, 2021.

-
- [127] A. Yılmaz, A. Küçüker, G. Bayrak, D. Ertekin, M. Shafie-Khah, and J. M. Guerrero, “An improved automated pqd classification method for distributed generators with hybrid svm-based approach using un-decimated wavelet transform,” *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107763, 2022.
- [128] C.-C. Liao, H.-T. Yang, and H.-H. Chang, “Denoising techniques with a spatial noise-suppression method for wavelet-based power quality monitoring,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 6, pp. 1986–1996, 2011.
- [129] M. Uyar, S. Yildirim, and M. T. Gencoglu, “An effective wavelet-based feature extraction method for classification of power quality disturbance signals,” *Electric power systems Research*, vol. 78, no. 10, pp. 1747–1755, 2008.
- [130] D. H. Chiam, K. H. Lim, and K. H. Law, “Global attention-based lstm for noisy power quality disturbance classification,” *International Journal of Systems, Control and Communications*, vol. 14, no. 1, pp. 22–39, 2023.
- [131] J. KU and B. C. Kovoov, “A wavelet-based hybrid multi-step wind speed forecasting model using lstm and svr,” *Wind Engineering*, vol. 45, no. 5, pp. 1123–1144, 2021.
- [132] J. Liang, S. Elangovan, and J. Devotta, “A wavelet multiresolution analysis approach to fault detection and classification in transmission lines,” *International Journal of Electrical Power & Energy Systems*, vol. 20, no. 5, pp. 327–332, 1998.
- [133] K. Silva, B. A. Souza, and N. S. Brito, “Fault detection and classification in transmission lines based on wavelet transform and ann,” *IEEE Transactions on Power Delivery*, vol. 21, no. 4, pp. 2058–2063, 2006.
- [134] D. H. Chiam, K. H. Lim, and K. H. Law, “Noise level evaluation on power quality disturbances classification,” in *International Conference on Green Energy, Computing and Sustainable Technology (GECOST 2021)*, 2021, p. (Accepted).
- [135] 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, no. 13, p. 4676, 2022.

-
- [136] R. Smolenski, P. Szczesniak, W. Drozd, 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, p. 111988, 2022.
- [137] S. P. Bihari, P. K. Sadhu, K. Sarita, B. Khan, L. Arya, R. Saket, and D. Kothari, “A comprehensive review of microgrid control mechanism and impact assessment for hybrid renewable energy integration,” *IEEE Access*, 2021.
- [138] 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, pp. 1335–1352, 2019.
- [139] H. Sindi, M. Nour, M. Rawa, Ş. Öztürk, and K. Polat, “An adaptive deep learning framework to classify unknown composite power quality event using known single power quality events,” *Expert Systems with Applications*, vol. 178, p. 115023, 2021.
- [140] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, “Deep learning methods and applications for electrical power systems: A comprehensive review,” *International Journal of Energy Research*, vol. 44, no. 9, pp. 7136–7157, 2020.
- [141] J. Vig and Y. Belinkov, “Analyzing the structure of attention in a transformer language model,” *arXiv preprint arXiv:1906.04284*, 2019.
- [142] 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, no. 2, pp. 924–930, 1996.
- [143] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [144] H. Yan, B. Deng, X. Li, and X. Qiu, “Tener: adapting transformer encoder for named entity recognition,” *arXiv preprint arXiv:1911.04474*, 2019.
- [145] 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 *International conference on machine learning*. PMLR, 2019, pp. 3744–3753.

-
- [146] 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.
- [147] A. Olabi and M. A. Abdelkareem, “Renewable energy and climate change,” *Renewable and Sustainable Energy Reviews*, vol. 158, p. 112111, 2022.
- [148] T. Hai, J. Zhou, and K. Muranaka, “Energy management and operational planning of renewable energy resources-based microgrid with energy saving,” *Electric Power Systems Research*, vol. 214, p. 108792, 2023.
- [149] V. Vinothkumar *et al.*, “Recent trends in power quality improvement using custom power devices and its performance analysis,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 7, pp. 1686–1695, 2021.
- [150] R. P. K. Naidu and S. Meikandasivam, “A comprehensive review on custom power devices,” *System*, vol. 20, p. 21, 2020.
- [151] O. N. Pardo-Zamora, R. d. J. Romero-Troncoso, J. R. Millan-Almaraz, D. Morinigo-Sotelo, R. A. Osornio-Rios, and J. A. Antonino-Daviu, “Power quality disturbance tracking based on a proprietary fpga sensor with gps synchronization,” *Sensors*, vol. 21, no. 11, p. 3910, 2021.
- [152] S. Baraskar, “Assessment of power quality disturbance by the discrete wavelet packet transform implemented on the fpga platform,” in *2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*. IEEE, 2021, pp. 1–4.
- [153] M. M. Luiz, T. F. Duque, A. H. Almeida, E. B. Kapisch, L. R. Silva, and M. A. Lima, “Power quality parameters calculation using fpga embedded parallel processors in compliance with the iec 61000-4-30 standard,” *Journal of Control, Automation and Electrical Systems*, pp. 1–12, 2022.
- [154] Z. Shen, C. Wu, L. Wang, and G. Zhang, “Real-time energy management for microgrid with ev station and chp generation,” *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1492–1501, 2021.

-
- [155] M. Chen, Z. Shen, L. Wang, and G. Zhang, "Intelligent energy scheduling in renewable integrated microgrid with bidirectional electricity-to-hydrogen conversion," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2212–2223, 2022.
- [156] F. R. Albogamy, M. Y. I. Paracha, G. Hafeez, I. Khan, S. Murawwat, G. Rukh, S. Khan, and M. U. A. Khan, "Real-time scheduling for optimal energy optimization in smart grid integrated with renewable energy sources," *IEEE Access*, vol. 10, pp. 35 498–35 520, 2022.
- [157] S. M. Halpin and A. Card, "Power quality," in *Power Electronics Handbook*. Elsevier, 2011, pp. 1179–1192.
- [158] H. D. Tafti, A. Ahmad, L. Callegaro, G. Konstantinou, and J. E. Fletcher, "Sensitivity of commercial rooftop photovoltaic inverters to grid voltage swell," in *2021 IEEE 12th Energy Conversion Congress & Exposition-Asia (ECCE-Asia)*. IEEE, 2021, pp. 308–313.
- [159] D.-M. Lee, T. G. Habetler, R. G. Harley, T. L. Keister, and J. R. Rostron, "A voltage sag supporter utilizing a pwm-switched autotransformer," *IEEE transactions on power electronics*, vol. 22, no. 2, pp. 626–635, 2007.
- [160] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [161] A. Chaman and I. Dokmanic, "Truly shift-invariant convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3773–3783.
- [162] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *IEEE digital signal processing workshop*, vol. 86. Citeseer, 1998, pp. 120–131.

-
- [163] H. H. Tan and K. H. Lim, “Vanishing gradient mitigation with deep learning neural network optimization,” in *2019 7th international conference on smart computing & communications (ICSCC)*. IEEE, 2019, pp. 1–4.
- [164] K. H. Lim, H. H. Tan, and H. G. Harno, “Approximate greatest descent in neural network optimization,” *Numerical Algebra, Control and Optimization*, vol. 8, no. 3, pp. 327–336, 2018.
- [165] H. H. Tan, K. H. Lim, and H. G. Harno, “Stochastic diagonal approximate greatest descent in neural networks,” in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 1895–1898.
- [166] J. Mohanty, M. Panda, M. Das, P. Dash, and P. Pradhan, “Sensitive load management in captive power plant—aluminium smelter,” in *Advances in Energy Technology: Proceedings of ICAET 2020*. Springer, 2021, pp. 211–220.
- [167] C.-J. Wong, J. Shi, J. Bao, B. J. Welch, M. Skyllas-Kazacos, A. Jassim, M. Mahmoud, and K. Nikandrov, “A smart individual anode current measurement system and its applications,” in *Light Metals 2023*. Springer, 2023, pp. 43–51.
- [168] C.-J. Wong, J. Bao, M. Skyllas-Kazacos, A. Jassim, M. Mahmoud, and A. Arkhipov, “Monitoring cell conditions and anode freeze dissolution with model-based soft sensor after anode change,” in *Light Metals 2023*. Springer, 2023, pp. 87–94.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.