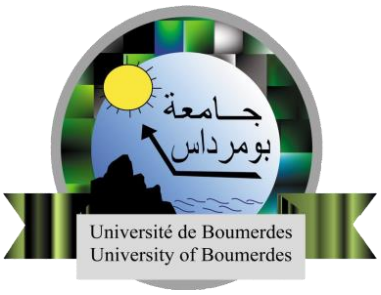
People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University M'Hamed BOUGARA – Boumerdes



Institute of Electrical and Electronic Engineering
Department of Power and Control

Final Year Project Report Presented in Partial Fulfilment of
the Requirements for the Degree of the Master in Power Engineering

Title:

Detection and Classification of Power Quality Disturbances.

Presented by:

- Dekkiche Sarra
- Ouzerala Sara

Supervisor:

Pr.F.Recioui

2022/2023

Abstract

Smart grid equipment has revolutionized various aspects of everyday life. However, a drawback of these systems is their nonlinear behavior, which leads to the injection of electrical disturbances into the power grid. These disturbances can result in voltage and current distortions. Artificial Intelligence (AI) techniques based on automatic feature extraction are utilized to detect and classify power quality disturbances. Subsets of AI; Deep Learning and Machine Learning algorithms such as the hybrid CNN-SVM and the Long Short-Term Memory (LSTM) are used for the detection and classification of these disturbances. The voltage signals obtained from MATLAB simulation serve as training and testing data for identification algorithms aimed at detecting various types of disturbances, such as voltage sag, voltage swell, harmonics, and interruption. The outcomes of the simulations demonstrate that the CNN-SVM technique exhibits higher accuracy in detecting and classifying power quality disturbances compared to the LSTM technique.

Acknowledgement

We would like to extend our heartfelt gratitude and appreciation to all those who have contributed to the successful completion of our Master's thesis.

First and foremost, we are profoundly grateful to the almighty ALLAH for granting us the strength and ability to comprehend, learn, and accomplish this thesis.

We would like to express our deepest thanks to our supervisor, Pr. F. Recioui, for her exceptional guidance, expertise, and unwavering support throughout the entire research process. Her invaluable insights and constructive feedback have been pivotal in shaping this thesis.

Our gratitude extends to the faculty members and mentors who have provided us with their valuable help and profound knowledge during our academic journey. Their dedication to our growth and development has pushed us to new intellectual heights and expanded our horizons.

We are also immensely grateful to our colleagues and friends who have continuously served as a source of inspiration, motivation, and intellectual exchange. Their engaging discussions, innovative ideas, and constant encouragement have played a significant role in shaping our thesis.

Lastly, we would like to acknowledge the unwavering support of our families and loved ones throughout this journey. Their boundless love, understanding, and encouragement have been our source of strength during the most challenging times.

Once again, we express our sincerest appreciation to all those who have contributed to the completion of our Master's thesis.

Dedication

We dedicate this Master's thesis with deep gratitude and appreciation to our loving mothers, fathers, and siblings, whose unwavering support, love, and encouragement have been the foundation of our academic journey. Your belief in our abilities and sacrifices made for our education have shaped us into the individuals we are today. This achievement would not have been possible without you.

To our friends, thank you for standing by our side throughout the writing process. Your words of encouragement and belief in our capabilities have provided us with the strength to overcome challenges and persevere.

Lastly, we express our sincere gratitude to all those who have offered their assistance and support, whether through providing resources, proofreading, or lending a listening ear. Your contributions have lightened our burdens and enriched our work

List of Contents

Abstract…….. IV
Acknowledgements……... IV
Dedication……. IV
List of Contents……. IV
List of Figures……. VI
List of Tables……... VII
List of Abbreviations……. IV
General Introduction…….. 1
Chapter 1 Power Quality Disturbances
1.1 Introduction……. 2
1.2 Definition of Electric Power Quality……... 2
1.3 Power Quality Standards……. 2
1.4 Classification of Power Quality Disturbances……. 3
1.4.1 Transients……... 3
1.4.2 Long-Duration Voltage Variations……. 5
1.4.3 Short-Duration Voltage Variations……. 6
1.4.4 Voltage Imbalance (Voltage Unbalance) …….. 7
1.4.5 Waveform Distortion…… 7
1.4.6 Voltage fluctuation (Flicker) …… 8
1.4.7 Power Frequency Variations…….. 9
1.5 Conclusion……... 10
Chapter 2 Detection and Classification of PQDs
2.1 Introduction……. 11
2.2 Detection of Power Quality Disturbances……... 11
2.2.1 Fourier Transform Based Methods…… 11
2.2.2 Wavelet Transform Based Methods……... 12
2.2.3 Hilbert-Huang Transform…… 12
2.2.4 S-transform…… 12
2.3 Classification of Power Quality Disturbances……. 13
2.3.1 Artificial Intelligence……. 13
2.3.2 Machine Learning…….. 13
2.4 Classification Algorithms…… 15
2.4.1 K-Nearest Neighbor (K-NN) …… 15
2.4.2 Decision Trees……... 16
2.4.3 Random Forest……... 17
2.4.4 Support Vector Machine (SVM) ……... 17
2.5 Artificial Neural Networks (ANNs) …… 18
2.6 Deep Learning……. 18
2.7 Detection and Classification of PQDs Using Deep Learning…… 19
2.7.1 Convolutional Neural Networks (CNNs) …… 19
2.7.2 Long Short-Term Memory (LSTM) ……. 19
2.8 Conclusion……... 20
Chapter 3 CNN, SVM and LSTM
3.1 Introduction …… 21
3.2 Convolutional Neural Network (CNN) …….. 21
3.3 Layers of a Convolutional Neural Network…… 22
3.3.1 Convolutional Layer…… 22
3.3.2 Batch Normalization Layer……... 23

3.3.3 ReLU Layer…….. 24
3.3.4 Pooling Layer…… 24
3.3.5 Flattening…….. 25
3.3.6 Fully Connected Layer…… 25
3.3.7 Softmax Activation Function……. 27
3.4 Loss Function…….. 28
3.5 CNN Architectures…….. 28
3.5.1 ResNet-50……. 29
3.6 Support Vector Machines (SVMs) ……. 30
3.6.1 Kernel Functions…… 30
3.7 Hybrid CNN-SVM…….. 31
3.8 Long Short-Term Memory (LSTM) ……... 32
3.9 Architecture of LSTM……. 32
3.9.1 Forget Gate…… 33
3.9.2 Input Gate…….. 33
3.9.3 Output Gate……... 33
3.10 Performance Evaluation of Classification Models…… 34
3.11 Conclusion……. 35
Chapter 4 Simulation Results & Discussion
4.1 Introduction……. 36
4.2 Generation of PQ Disturbances……... 36
4.3 CNN-SVM Model…… 40
4.4 LSTM Model……... 40
4.5 Proposed Technique……. 41
4.6 Normal Voltage…… 44
4.7 Voltage Sag…….. 46
4.8 Voltage Swell……... 48
4.9 Voltage Interruption……. 50
4.10 Harmonics……. 52
4.11 Sag with Harmonics……... 54
4.12 Swell with Harmonics……... 56
4.13 Interruption with Harmonics……. 58
4.14 All PQ Events without Harmonics…… 60
4.15 All PQ Events with Harmonics…… 63
4.16 CNN-SVM and LSTM Training…… 67
4.17 Discussion…….. 68
4.18 Conclusion……. 69
General Conclusion……... 70
Future Work……... 71
References……. 72

List of Figures

Fig 1.1 Lightning stroke current impulsive transient……... 4
Fig 1.2 Oscillatory transient current caused by back-to-back capacitor switching……. 5
Fig 1.3 Overvoltage signal…… 5
Fig 1.4 Undervoltage signal……. 6
Fig 1.5 Voltage interruption……. 6
Fig 1.6 Voltage sag signal…… 6
Fig 1.7 Voltage swell signal……. 7
Fig 1.8 Typical three-phase voltage unbalance signal…… 7
Fig 1.9 Voltage fluctuation signal…… 9
Fig 1.10 Power frequency variations…… 9
Fig 2.1 K-NN Classifier……... 16
Fig 2.2 Decision Tree Classifier…….. 16
Fig 2.3 Random Forest Classifier…… 17
Fig 2.4 Support Vector Machine Classifier…….. 18
Fig 3.1 The calculation process for the features of the convolutional layer……..…. 22
Fig 3.2 Max pooling and average pooling methods…… 25
Fig 3.3 Flattening the pooled feature map……... 25
Fig 3.4 Visualized dot product of the weights matrix and the input vector……. 26
Fig 3.5 Fully connected layer…….. 27
Fig 3.6 ResNet-50 Architecture……... 29
Fig 3.7 The LSTM Cell……... 32
Fig 3.8 Evaluation Measures of Confusion Matrix…… 35
Fig 4.1 PQ disturbance models…… 37
Fig 4.2 PQ Disturbances without Harmonics. …….. 38
Fig 4.3 PQ Disturbances with Harmonics. …….. 39
Fig 4.4 Flowcharts of the proposed methods: (a) CNN-SVM Method and (b) LSTM Method.. 42
Fig 4.5 Normal voltage: (a) Waveform and (b) CNN-SVM Output……. 45
Fig 4.6 Normal voltage: (a) Waveform, (b) Amplitude, (c) LSTM output 1, and (d) LSTM output 2. ……...…… 45
Fig 4.7 Voltage Sag: (a) Waveform and (b) CNN-SVM Output…….. 47
Fig 4.8 Voltage Sag: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output 2…… 48

Fig 4.9 Voltage Swell: (a) Waveform and (b) CNN-SVM Output…….. 49
Fig 4.10 Voltage Swell: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output 2. ……... 50
Fig 4.11 Voltage Interruption: (a) Waveform and (b) CNN-SVM Output……... 51
Fig 4.12 Voltage Interruption: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output 2.…….. 52
Fig 4.13 Harmonics: (a) Waveform and (b) CNN-SVM Output……. 53
Fig 4.14 Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2…….. 54
Fig 4.15 Sag with Harmonics: (a) Waveform and (b) CNN-SVM Output…….. 55
Fig 4.16 Sag with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.. ……. 56
Fig 4.17 Swell with Harmonics: (a) Waveform and (b) CNN-SVM Output…… 57
Fig 4.18 Swell with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.…… 58
Fig 4.19 Interruption with Harmonics: (a) Waveform and (b) CNN-SVM Output……. 59
Fig 4.20 Interruption with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2. …… 60
Fig 4.21 All PQ Events without Harmonics: (a) Waveform and (b) CNN-SVM Output. …… 62
Fig 4.22 All PQ Events without Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2…….. 63
Fig 4.23 All PQ Events with Harmonics: (a) Waveform and (b) CNN-SVM Output…. 65
Fig 4.24 All PQ Events with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.……. 66
Fig 4.25 CNN-SVM and LSTM Training: a) CNN-SVM Training b) LSTM Training.. 67

List of Tables

Table 1.1 List of some standards on various issues of power quality……... 3

Table 4.1 CNN-SVM Outputs…….. 40

Table 4.2 LSTM Outputs…….. 41

Table 4.3 magnitude, start and end time, duration, and accuracy for the normal waveform…… 44

Table 4.4 magnitude, start and end time, duration, and accuracy for the sag waveform……..... 46

Table 4.5 magnitude, start and end time, duration, and accuracy for the swell waveform……... 49

Table 4.6 magnitude, start and end time, duration, and accuracy for the voltage interruption waveform…… 51

Table 4.7 magnitude, start and end time, duration, and accuracy for the harmonics waveform….. 53

Table 4.8 magnitude, start and end time, duration, and accuracy for the sag with harmonics waveform……... 55

Table 4.9 magnitude, start and end time, duration, and accuracy for the swell with harmonics waveform……... 57

Table 4.10 magnitude, start and end time, duration, and accuracy for the interruption with harmonics waveform……. 59

Table 4.11 The Set Values for Each Disturbance in Both Models…… 61

Table 4.12 magnitude, start and end time, duration, and accuracy for all the PQ events without harmonics waveform: a) CNN-SVM model and b) LSTM model…… 61

Table 4.13 The Set Values for Each Disturbance in Both Models…… 64

Table 4.14 magnitude, start and end time, duration, and accuracy for all the PQ events with harmonics waveform: a) CNN-SVM model and b) LSTM model…… 64

List of abbreviations

PQ	Power Quality
IEEE	Institute of Electrical and Electronics Engineers
IEC	International Electrotechnical Commission
EPQ	Electric Power Quality
RMS	Root Mean Square
AC	Alternating Current
DC	Direct Current
PU	Per Unit
AI	Artificial Intelligence
FT	Fourier Transform
STFT	Short-Time Fourier Transform
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
WT	Wavelet Transform
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
HHT	Hilbert-Huang Transform
EMD	Empirical Mode Decomposition
ML	Machine Learning
SVM	Support Vector Machine
K-NN	K-Nearest Neighbor
DL	Deep Learning
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
BN	Batch Normalization
ReLU	Rectified Linear Unit
MSE	Mean Squared Error
RBF	Radial Basis Function
RES	Renewable Energy Source

General Introduction

The incorporation of renewable energy sources (RESs) plays a significant role in fulfilling the growing electricity demand. However, the integration of these sources can introduce power quality disturbances (PQDs) into the grid, including harmonics, voltage sags, voltage swells, and interruption. These disturbances can impact the reliability of the grid and result in failures in electrical equipment. Electric signals monitoring and real-time identification of these disturbances are necessary for the network's recovery. Various methods have been developed to detect and classify PQDs based on feature extraction. Traditional frequency-domain analysis techniques like Fourier transform (FT) and discrete Fourier transform (DFT) are suitable for stationary signals but may not capture nonstationary signals effectively. The short-time Fourier transform (STFT) and S transform (ST) address some limitations but still have drawbacks in analyzing non-stationary transients. Automatic PQD detection based on Artificial Intelligence (AI) techniques solves the above-mentioned issues. Artificial Intelligence (AI) is a dynamic and interdisciplinary field of computer science that plays an important role in addressing power quality disturbances. Within AI, specialized subfields, namely Machine Learning and Deep Learning, contribute significantly to this domain. Machine Learning empowers computers to learn from data and experiences without explicit programming, leveraging algorithms and statistical models to enhance performance on specific tasks related to power quality analysis. On the other hand, Deep Learning, a subset of Machine Learning, emulates the neural networks of the human brain. By employing Artificial Neural Networks with multiple layers, Deep Learning effectively captures intricate patterns and representations from extensive datasets, making it well-suited for dealing with complex power quality challenges. methods such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Support Vector Machines (SVMs) have been proven successful for the detection and classification of PQD.

This thesis is divided into the following sections: starting by a general introduction, then the first chapter, which describes the power quality problems with their causes and effects. The next chapter presents the most common detection and classification methods. Chapter three explains in detail Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), and Long Short-Term Memory (LSTM) methods. The last chapter deals with the simulation of the hybrid CNN-SVM and LSTM models, then, a discussion about the obtained results. Finally, a general conclusion is reached that summarizes the achievement of this study's objectives.

1.1. Introduction

The quality of electric power has become a significant issue for both utility companies and their customers. As a result, power quality (PQ) studies are becoming more popular. Power line disturbances such as voltage sag/swell with and without harmonics, momentary interruption, harmonic distortion, flicker, notch, spike, and transients typically cause degradation in power quality, resulting in problems such as malfunctions, instabilities, short lifetime, failure of electrical equipment, and so on. Various organizations, such as the IEEE (Institute of Electrical and Electronics Engineers) and the IEC (International Electrotechnical Commission), develop a number of standards and benchmarks that are enforced by customers, utilities, and manufacturers to minimize or eliminate power quality problems.

1.2. Definition of Electric Power Quality

Electric Power Quality (EPQ) is a term that refers to maintaining the near sinusoidal waveform of voltages and currents at rated magnitude and frequency [1]. In IEEE dictionary, power quality is defined as “the concept of powering and grounding sensitive equipment in a matter that is suitable to the operation of the equipment”. Also defined in IEC as “set of parameters defining the properties of the power supply as delivered to the user in normal operating conditions in terms of continuity of supply and characteristics of voltage (magnitude, frequency, waveform)” [2].

1.3. Power Quality Standards

Several standards, such as allowable level of deviations, mitigation, and monitoring, have been developed, amended, advised, and enforced in response to the evolution of technology in order to preserve and measure the level of power quality. Table 1.1 lists some of the current standards on several elements of power quality [3].

Table 1.1 List of some standards on various issues of power quality [3].

Standards	Description
IEEE Standard 519-1992	Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems
IEEE Standard 1159-1995	Recommended Practice for Monitoring Electric Power Quality
IEEE Standard 1100-1999	Recommended Practice for Powering and Grounding Sensitive Electronic Equipment
IEEE Standard 1250-1995	Guide for Service to Equipment Sensitive to Momentary Voltage Disturbances
IEEE Standard 1366-2012	Electric Power Distribution Reliability Indices
IEC 61000-2-2	Compatibility Levels for Low-Frequency Conducted Disturbances and Signaling in Public Supply Systems
IEC 61000-2-4	Compatibility Levels in Industrial Plants for Low-Frequency Conducted Disturbances
IEC 61000-3-2	Limits for Harmonic Current Emissions (Equipment Input Current Up to and Including 16A Per Phase)
IEC 61000-4-15	Flicker-meter Functional and Design Specifications
EN 50160	Voltage Characteristics of Public Distribution Systems

1.4. Classification of Power Quality Disturbances

Different types of electrical disturbances contribute to power quality issues. These disturbances are categorized based on their impact on amplitude, frequency, or both. Power quality events are classified as short, medium, or long, depending on the duration of the disruptions. The disturbances that lead to power quality degradation in a power system, as well as their classification, primarily include:

1.4.1. Transients

Is the sudden change in one steady state operating condition to another, are classified into two categories:

a) Impulsive and b) Oscillatory.

a) Impulsive Transients

An impulsive transient refers to an abrupt, non-power frequency alteration in the steady-state condition of either voltage, current, or both, with a unidirectional polarity, usually either positive or negative. Such transients can stimulate the inherent frequency of power system circuits, leading to the production of oscillatory transients [4]. The most common causes of impulsive transients are lightning strikes, switching of the inductive loads, or switching in the power distribution system. These transients can result in equipment shutdown or damage if the disturbance level is high enough [5].

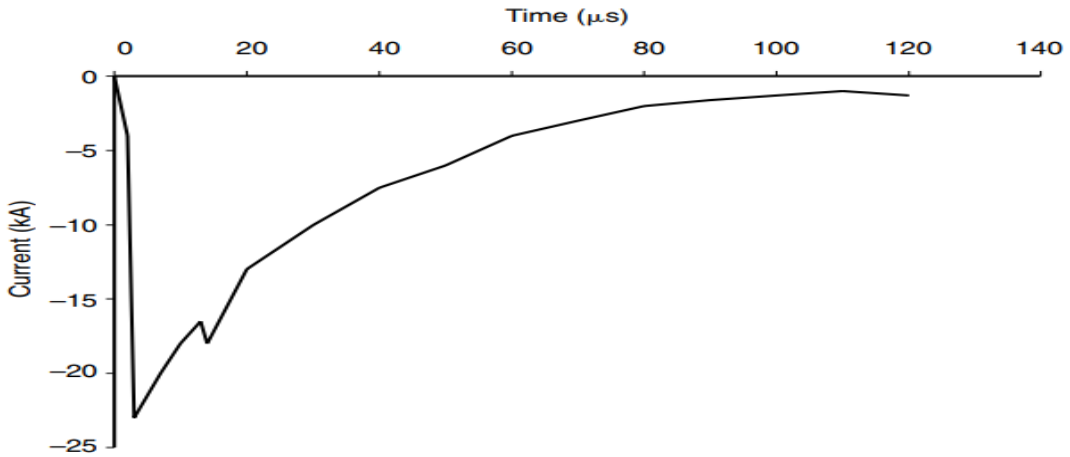


Fig 1.1 Lightning stroke current impulsive transient [6].

b) Oscillatory Transients

An oscillatory transient is an abrupt and non-power frequency alteration in the steady-state condition of either voltage, current, or both, that involves both positive and negative polarity values. Such transients can arise from the energization of transformers [5].

Oscillatory transients are classified into three types:

1. High-frequency Transients These have frequency components greater than 500 kHz and a typical duration measured in microseconds (or several cycles of the principal frequency).

2. Medium-frequency Transients These have frequency components between 5 and 500 kHz with duration measured in the tens of microseconds (or several cycles of the principal frequency).

3. Low-frequency Transients These have frequency components less than 5 kHz and a duration from 0.3 to 50ms [6].

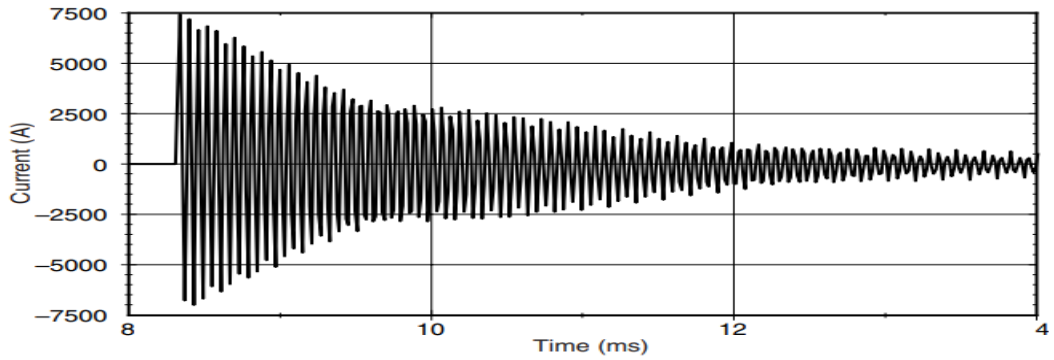


Fig 1.2 Oscillatory transient current caused by capacitor switching [6].

1.4.2. Long-Duration Voltage Variations

When the rms value of voltage deviates for a duration more than 1 minute, it is termed as long duration voltage variation.

a) Over-voltage

Over-voltage refers to a rise in the root mean square (rms) alternating current (ac) voltage, exceeding 110 percent at the power frequency, and lasting for more than one minute. Such over-voltages can result from load switching, capacitor energization, or improper tap settings on transformers, and may cause detrimental harm to machines or other equipment linked to the system [5].

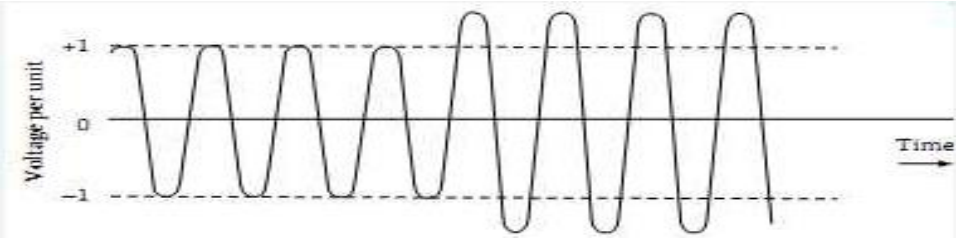


Fig 1.3 Over-voltage signal [7].

b) Under-voltage

An under-voltage is a decrease in the rms ac voltage to less than 90 percent at the power frequency for a duration longer than 1 min [5]. Turning on a load or de-energizing a capacitor can lead to under-voltages. Additionally, overloaded circuits can also cause under-voltages, resulting in augmented losses in the power system and motor heating.

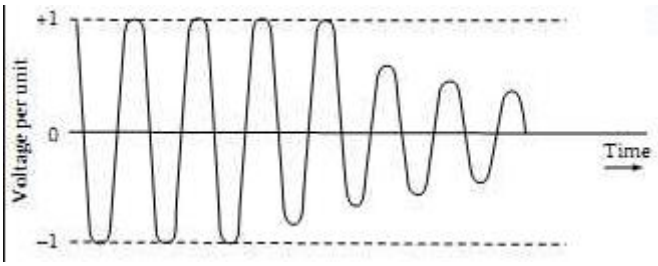


Fig 1.4 Under-voltage signal [7].

c) Sustained interruption

When the supply voltage becomes zero for a period in excess of 1 min, the long-duration voltage variation is considered a sustained interruption [6]. Power system faults are among the most frequent reasons for sustained interruptions, leading to production losses.

1.4.3. Short-Duration Voltage variations

When the rms value of voltage deviates for a duration less than 1 minute, it is termed as short duration voltage variation.

a) Interruption

An interruption happens when the supply voltage or load current drops to below 0.1 pu, and the duration does not exceed one minute. Such interruptions can occur due to power system faults, equipment failures, or control malfunctions [8].

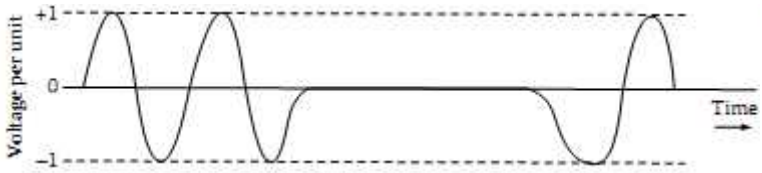


Fig 1.5 Voltage interruption [7].

b) Sags (dips)

A sag refers to a decline in the root mean square (rms) voltage or current at the power frequency, ranging from 0.1 to 0.9 pu, and lasting from 0.5 cycle to one minute. This phenomenon is also known as "dip," as per the IEC definition, and the two terms are used interchangeably. However, "sag" is the preferred term in the U.S. power quality community. The voltage sag happens between the fault initiation and the protective device operation [5]. Common sources of sags are utility faults, energization of heavy loads or starting of large motors.

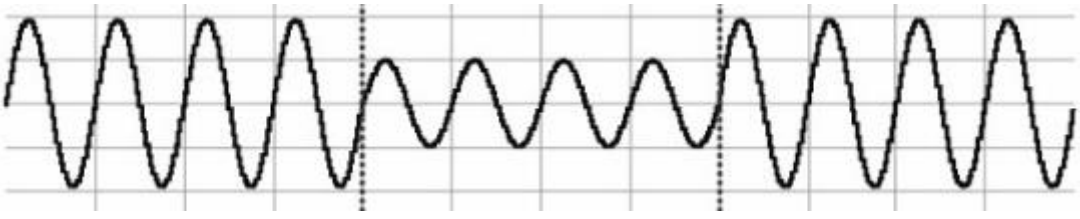


Fig 1.6 Voltage sag signal [9].

c) Swells

A swell is characterized by a rise in the root mean square (rms) voltage or current at the power frequency, ranging from 1.1 to 1.8 pu, and lasting from 0.5 cycle to one minute. Temporary voltage increase on the faultless phase during single line to ground (SLG) fault is responsible for voltage swells. Additionally, switching off a large load or energizing a large capacitor bank can also cause swells [6].

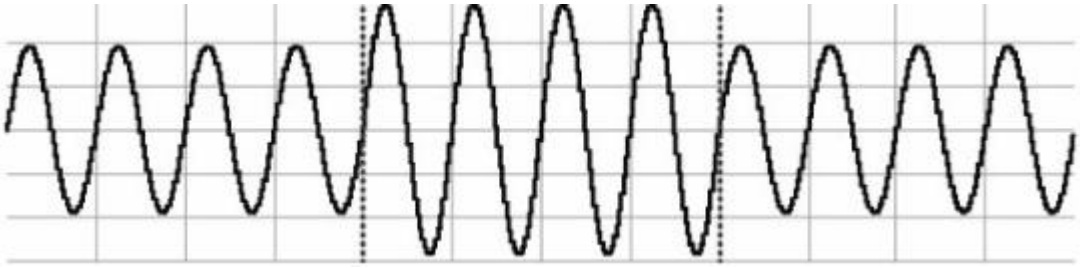


Fig 1.7 Voltage swell signal [9].

1.4.4. Voltage Imbalance (Voltage Unbalance)

Voltage unbalance is defined as the highest deviation from the mean of the three-phase voltages or currents, divided by the mean of the three-phase voltages or currents, and expressed as a percentage. This power quality issue occurs due to an uneven distribution of loads across the three phases. Voltage unbalance can also arise from blown fuses in one phase of a three-phase capacitor bank. In some cases, severe voltage unbalance (more than 5 percent) can occur due to single-phasing conditions [10].

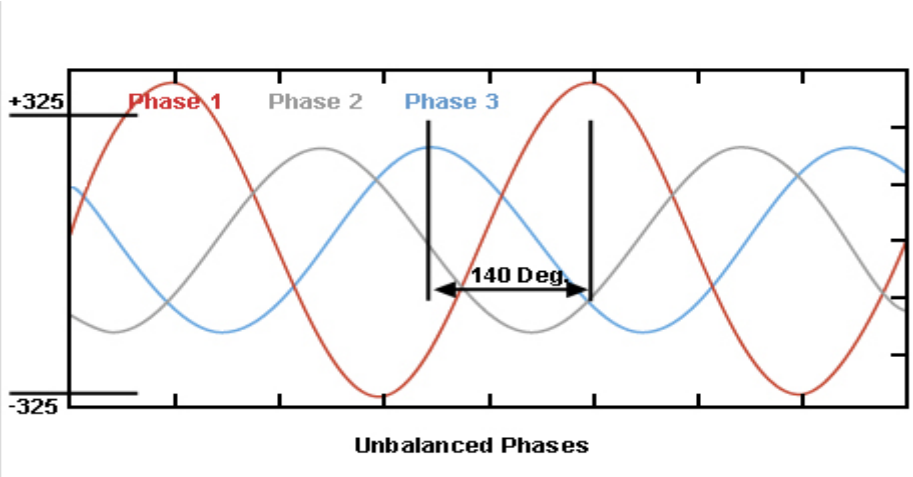


Fig 1.8 Typical three-phase voltage unbalance signal [11].

1.4.5. Waveform Distortion

Waveform distortion is defined as a steady-state deviation from an ideal sine wave of power frequency. There are five types of waveform distortion:

a) DC Offset

DC offset is the existence of a direct current voltage or current in an alternating current power system. This phenomenon can arise due to geomagnetic disturbances or asymmetry of electronic power converters. DC offset leads to increased heating and saturation in transformers. Moreover, it can cause electrolytic erosion of grounding electrodes and other connectors [12].

b) Harmonics

A harmonic refers to a component whose frequency is a multiple number of the fundamental frequency, and the harmonic number represents its frequency. The harmonic order is the ratio of the harmonic frequency to the fundamental frequency. Harmonics can be generated due to non-linearity, excessive use of semiconductor-based switching devices, and various design constraints. The presence of harmonics can have negative impacts on generation, transmission, and distribution systems, as well as on consumer equipment [13].

c) Interharmonics

Interharmonics are components with a frequency that is not a multiple of the fundamental frequency. Nonlinear loads, static frequency converters, cycloconverters, and arching devices are the primary sources of interharmonic waveform distortion. The presence of interharmonics can result in acoustic noise in power equipment [14].

d) Notching

Notching is a power quality issue caused by a brief switching disturbance in the normal power voltage waveform that lasts less than half a cycle. This disturbance initially has an opposite polarity to the waveform, and as a result, the peak value of the disturbance voltage is subtracted from the normal waveform. Notching typically arises due to the regular operation of electronic devices during the commutation of current from one phase to another. This type of disturbance can cause damage to capacitive components [15].

e) Noise

Electrical noise refers to undesired electrical signals that cause adverse effects in the circuits of control systems where they occur. Power electronic devices, arcing equipment, loads with solid-state rectifiers, and switching power supplies are some of the primary sources of noise in power systems. The presence of electrical noise can lead to capacitor overloading and disturbances in appliances [16].

1.4.6. Voltage Fluctuation (Flicker)

Voltage fluctuation is an electromagnetic phenomenon that results in random changes in voltage. Typically, the magnitude of these changes is within the range of 0.9 to 1.1 volts. One

of the primary reasons for voltage fluctuation is load changes. Flicker, as defined in electrical engineering, is a short-term fluctuation of voltage in the power supply system. This can cause lamps to flicker as their brightness is directly proportional to the applied voltage. Flicker can also disrupt the function of power supplies or TV receivers and, in severe cases, even cause damage to them [17].

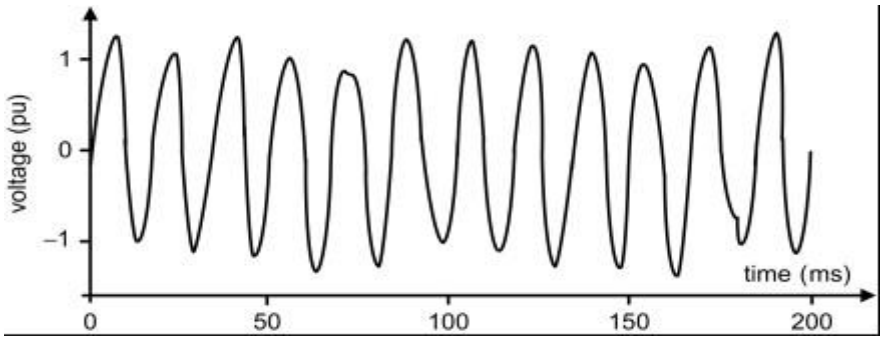


Fig 1.9 Voltage fluctuation signal [18].

1.4.7. Power Frequency Variations

Power frequency deviation refers to an increase or decrease in the frequency from its nominal value. The duration of such deviations can range from several cycles to several hours. These frequency variations that go beyond the acceptable limits for normal steady-state operation of the power system can be caused by faults on the bulk power transmission system, a large block of load being disconnected, or a significant source of generation going offline. These disturbances can result in damage to generators and turbine shafts [5].

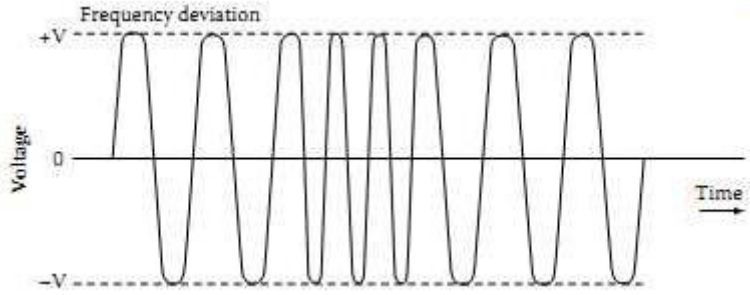


Fig 1.10 Power frequency variations [7].

1.5. Conclusion

Power quality pertains to the attributes of electrical power that affect the performance and dependability of electrical equipment. Inadequate power quality can result in equipment damage, downtime, and reduced efficiency, which can cause significant economic losses for businesses and industries. Therefore, it is crucial to monitor the quality of electricity supplied to the consumers. The various disturbances affecting electrical power can be analyzed and monitored using different techniques.

2.1 Introduction

The increasing occurrence of power quality signal-related issues has raised concerns for utilities and customers. Therefore, power quality monitoring systems employ various signal-processing techniques to identify the disturbances. Artificial Intelligence techniques can enhance the accuracy in detection speed and classification.

2.2 Detection of Power Quality Disturbances

The detection of power quality disturbances involves analyzing and monitoring various parameters and waveforms of the electrical signals to identify any deviations from normal operating conditions. These disturbances can manifest as voltage sags or swells, harmonics, transients, flicker, voltage interruptions, or other anomalies that affect the quality and reliability of the power supply.

Detection techniques in power quality analysis often rely on sophisticated signal processing algorithms such as Fourier transform, Wavelet transform, Hilbert-Huang transform, and S-transform.

2.2.1 Fourier Transform Based Methods

The Fourier transform (FT) is a well-known technique for frequency domain analysis, which represents a signal as a sum of sinusoidal terms with different frequencies. While FT is suitable for stationary signals and extracting the spectrum at specific frequencies, it cannot resolve any temporary information associated with fluctuations. To address this limitation, the short time Fourier transform (STFT) divides the signal into small segments, assuming each segment is stationary. STFT determines the sinusoidal frequency and phase contents of local sections of the signal, as they change over time and extracts several frames of signals analyzed with a window that moves with time. This allows for the identification of the relationship between the variance of frequency and time. However, analyzing non-stationary signals with STFT is difficult. Discrete STFT is used for time-frequency analysis of non-stationary signals by decomposing the time-varying signals into time-frequency domain components. On the other hand, the discrete Fourier transform (DFT) represents discrete signals that repeat themselves in a periodic fashion from negative to positive infinity, while the fast Fourier transform (FFT) achieves equivalent results as the DFT by significantly reducing the computational time required [19].

2.2.2 Wavelet Transform Based Methods

The wavelet transform (WT) has proven to be a powerful signal processing technique in various research fields. Its advantage lies in its ability to preserve time and frequency information more efficiently than other methods, such as the Fourier transform. Therefore, the WT technique, which uses a variable length window, solves the main drawback of the fixed window in the STFT, making it more suitable for non-stationary signals. There are two algorithms of the Wavelet transform: the Continuous Wavelet transform (CWT) and the Discrete Wavelet transform (DWT).

The CWT transfers the raw signals from the time domain to a continuous time-scale domain, similar to the time-frequency domain in the STFT. The scales in the WT correspond to the frequency contents of the signal, but each scale does not have a fixed relation with each individual frequency in the signal. On the other hand, the DWT decomposes the raw signals into a specific number of scales (discrete), as long as sufficient information can be obtained from its analysis [20],[21].

2.2.3 Hilbert-Huang Transform

The Hilbert-Huang Transform (HHT) is a powerful tool for analyzing power quality disturbances. One of the main advantages of the HHT is its ability to analyze nonlinear and non-stationary signals [22]. The HHT is a two-step method for analyzing signals, where the first step is empirical mode decomposition (EMD), and the second step is the Hilbert transform [23]. Several studies have proposed the use of the HHT in power quality analysis [24].

2.2.4 S-transform

S-transform is a mathematical tool used in power quality analysis to extract disturbance signals and obtain the time-frequency matrix with characteristics of the disturbance signals [25]. The S transform is a time-frequency analysis technique that uses a sliding window and a Fourier transform simultaneously to analyze the time and frequency content of a signal. Unlike the short-time Fourier transform (STFT), which uses a fixed window size, the S transform uses a variable window size that adapts to the local frequency content of the signal. This makes it particularly useful for analyzing non-stationary signals, such as power quality signals [25],[26]. The S transform is used to detect voltage sags, swells, interruptions, and other power quality disturbances [27].

2.3 Classification of Power Quality Disturbances

Classification of power quality disturbances involves applying Artificial Intelligence (AI) techniques, such as machine learning algorithms or deep learning models, to automatically categorize and label different types of disturbances or anomalies in an electrical power system.

2.3.1 Artificial Intelligence

The general definition of artificial intelligence (AI) can be the automation of activities that are associated with human thinking such as decision-making, problem solving, learning, perception, and reasoning [28].

In the context of power quality, AI, particularly machine learning and deep learning, plays an important role in detection and analysis. Machine learning algorithms are trained on labeled power quality data to classify different types of disturbances. Deep learning, with its ability to extract intricate representations and handle large datasets, can further enhance power quality analysis. Deep neural networks can process mixed power quality disturbances, automatically learn hierarchical features, and make precise predictions. They can capture long-term dependencies, discern subtle variations, and differentiate between various types of disturbances [29].

2.3.2 Machine Learning

Machine learning is a branch of artificial intelligence (AI) that enables computer systems to automatically improve their performance on a specific task through experience. This is achieved by feeding large amounts of data into algorithms, allowing them to identify patterns and make predictions or decisions without being explicitly programmed to do so [30]. Machine learning algorithms play an important role in power quality analysis by automatically detecting, classifying, and predicting disturbances. By training models on labeled power quality data, machines can learn to recognize patterns and features associated with different types of disturbances. These algorithms can then analyze real-time power quality signals, identify the presence of specific disturbances, and provide timely alerts or take appropriate actions. Two main strategies employed in ML: supervised and unsupervised learning. Two other categories incorporated into ML: semi-supervised learning and reinforcement learning.

a) Supervised Learning

Supervised learning is one of the most widely practiced branches of machine learning that uses labeled training data to help models make predictions or decisions. It is based on cultivating data and generating an output from the previous labeled data. That means the input data consists of labeled examples: each data point is a pair of data examples (input object) and target label (desired to be predicted) [31]. Supervised learning provides the means to address two fundamental tasks in machine learning: classification and regression.

- **Classification Problem**

Classification is the process of assigning categorical labels or classes to input data based on their features [32]. This technique is commonly used in power quality monitoring to identify different types of disturbances, such as voltage sags, swells, harmonics, or transients. The input data includes features that describe the characteristics of the power quality events, and the output is a predicted class label. The classification algorithm learns from labeled training data and creates a model that can classify new, unseen instances into the appropriate class. Popular algorithms for classification include decision trees, support vector machines (SVMs), and neural networks [33].

- **Regression Problem**

Regression is the process of finding a model or function for distinguishing the data into continuous real values instead of using classes or discrete values. This technique is used when the goal is to estimate a quantity rather than assigning categorical labels. In power quality monitoring, regression can be applied to predict quantitative measures such as the magnitude or duration of a disturbance. The input data includes features that describe the properties of the power quality events, and the output is a numerical value. Regression algorithms learn from training data and create a model that can make predictions on new, unseen instances. Common regression algorithms include linear regression, decision trees, and random forests [34].

b) Unsupervised Learning

Unsupervised learning refers to algorithms that learn patterns from unlabeled data. In contrast to supervised learning where models learn to map the input to the output, unsupervised learning algorithms are allowed to classify, label, and/or group the data points contained within the data sets without having any external guidance in performing that task [35]. Unsupervised learning includes clustering algorithms such as k-means and dimensional reduction techniques.

c) Semi-Supervised Learning

Semi-supervised learning is a broad category of machine learning techniques that utilizes both labeled and unlabeled data. It is a hybrid technique between supervised and unsupervised learning [36].

d) Reinforcement Learning

Reinforcement learning is a type of machine learning which is trained by interacting with a real-time environment, and adjusting its actions accordingly [37]. A solution is found by the computer employing trial and error to find an optimal outcome. Reinforcement learning has been applied in robotics, game theory, and data science applications [38].

2.4 Classification Algorithms

The main task of power quality (PQ) disturbance recognition is to accurately identify the type of disturbance and provide references for parameter estimation and control strategies in power systems [39]. Due to the increasing demand for pure power, continuous monitoring and classification of these disturbances are required. To classify these disturbances, various algorithms have been developed to relate signal characteristics with the group they belong to. The performance of these methodologies depends on the extracted features and the classifier used. If the disturbance characteristics are not accurately captured, the performance of the classification algorithm will be limited. Therefore, it is important to select appropriate features and classifiers to achieve accurate and reliable classification of power quality disturbances. These algorithms include Decision Trees, Random Forests, K-NN, and Support Vector Machines (SVMs), among others [40].

2.4.1 K-Nearest Neighbor (K-NN)

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. The K-NN algorithm can be used for regression as well as for classification, but is mostly used for classification problems [41]. The algorithm works by storing all the available data and then, for a new data point, finding the K data points in that storage that are closest to it in terms of distance. The prediction is then based on the majority class of the K nearest neighbors or the average of their values, depending on the problem that needs to be solved (classification or regression) [31].

Some of the disadvantages of K-NN algorithms are: the constant need to determine the value of K which may be complex sometimes, and the computation cost is high due to the calculation of the distance between the data points for all the training samples [41].

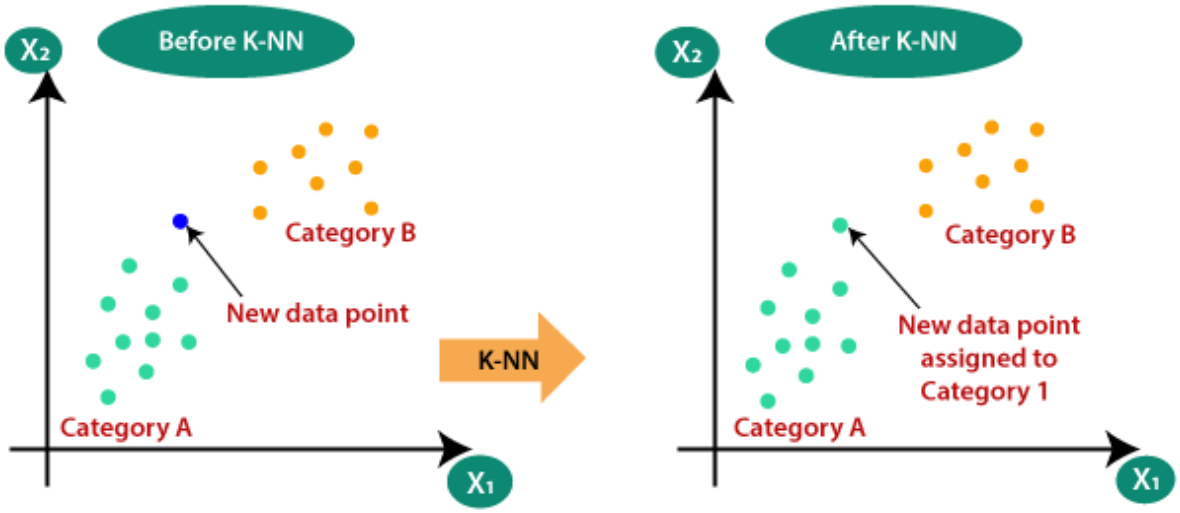


Fig 2.1 K-NN Classifier [41]

2.4.2 Decision Trees

A decision tree is a supervised learning algorithm that can be utilized for both classification and regression tasks [42]. It has a tree-like structure, which consists of a root node, branches, internal nodes, and leaf nodes. Each internal node represents a "test" on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules [43]. Decision trees can be used to create a training model that can be used to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data).

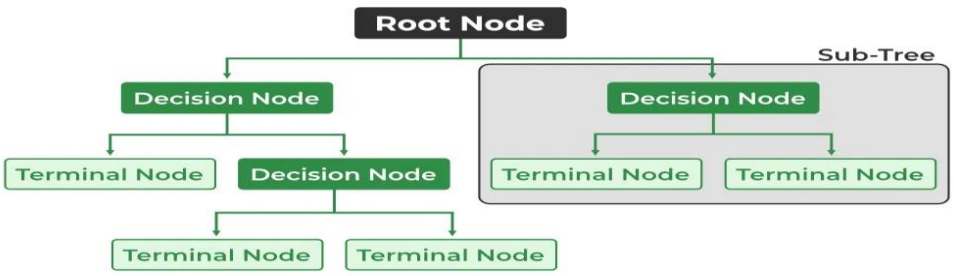


Fig 2.2 Decision Tree Classifier [43].

2.4.3 Random Forest

Random forest is a popular machine-learning algorithm that belongs to the supervised learning technique and can be used for both classification and regression problems in machine learning [44]. The random forest algorithm consists of many decision trees, and the output of the random forest is the class selected by most trees for classification tasks, while for regression tasks, the output of the random forest is the mean or average prediction of the individual trees [45]. Random decision forests correct for decision trees' habit of over-fitting to their training set. The greater the number of trees in the forest, the higher its accuracy and problem-solving ability, which prevents the problem of overfitting [44].

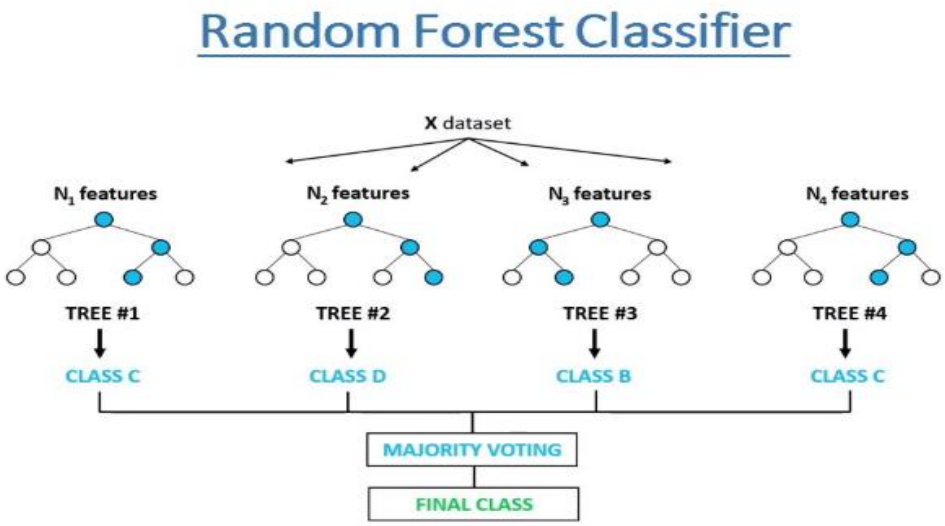


Fig 2.3 Random Forest Classifier [46]

2.4.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful and widely used supervised machine learning algorithm that is particularly effective for classification and regression tasks. SVMs are based on the concept of finding an optimal hyperplane that maximally separates different classes of data points in a high-dimensional feature space. The fundamental idea behind SVM is to transform the input data into a higher-dimensional feature space, where it becomes easier to find a hyperplane that can effectively separate the data points belonging to different classes. This transformation is accomplished using a kernel function, which computes the similarity between pairs of data points in the original feature space. The key objective of SVM is to find the hyperplane that maximizes the margin, which is the distance between the hyperplane and the nearest data points of each class. This margin allows for better generalization and robustness

of the model. The data points that are closest to the hyperplane and have the largest impact on its position are called support vectors [31].

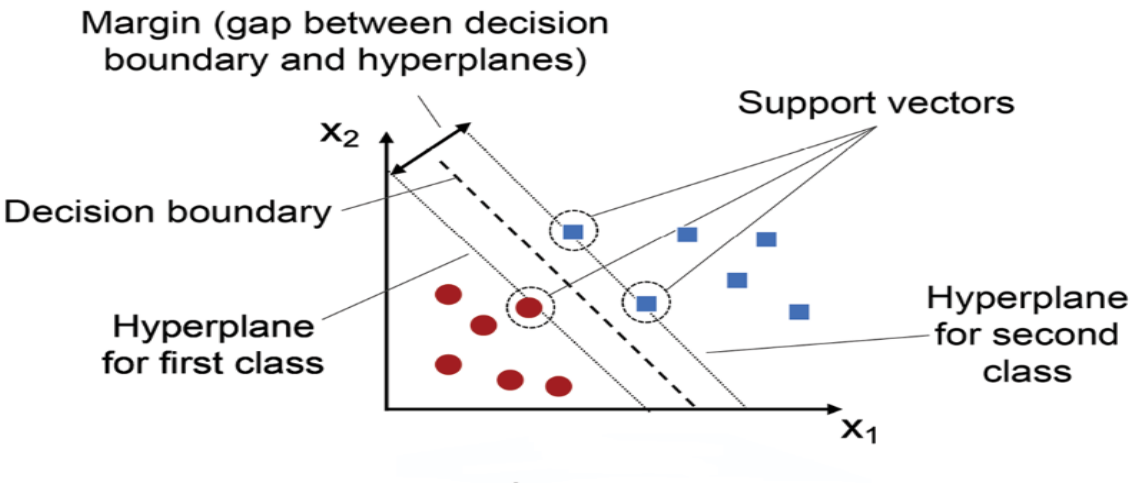


Fig 2.4 Support Vector Machine Classifier [47].

2.5 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, emulating the way that biological neurons signal to one another. ANNs are composed of nodes and layers containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. ANNs offer the ability to learn from data and capture complex relationships, making them suitable for handling the nonlinear and dynamic nature of power quality phenomena [49].

2.6 Deep Learning

Deep learning is a subset of machine learning that uses artificial neural networks to learn and make decisions based on data. It involves constructing multiple layers of artificial neurons that can extract higher-level features from raw input data, such as images or sound. Deep learning is used in a variety of applications, including image and speech recognition, and natural language processing. DL algorithms can effectively handle the complexity and variability of power quality signals. By training deep neural networks on labeled power quality data, the models can learn to automatically recognize and classify different types of disturbances. The multiple layers of the network allow for the extraction of intricate representations, capturing both global and local patterns within the data. This enables deep learning models to detect subtle

variations and distinguish between various types of power quality disturbances with high accuracy [48].

2.7 Detection and Classification of PQDs Using Deep Learning

Detection and classification of power quality disturbances using deep learning involves leveraging deep neural network models to automatically identify and categorize different types of disturbances or anomalies in an electrical power system. Some deep learning techniques commonly used in power quality analysis include: Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM).

2.7.1 Convolutional Neural Networks (CNNS)

Convolutional Neural Network (CNN) is an artificial neural network. Different layers are stacked together to form the CNN architecture. Commonly, four types of layers used to build convolutional neural networks are: convolutional layers, pooling layers, nonlinear or activation layers and fully connected layers. Four hyper parameters of CNN are: number of filters (K), their spatial extent (F), stride (S) and zero padding (P). The convolutional layer plays a key role in feature extraction. It contains filters that are convolved with the input and features are extracted. Each feature of a layer receives an input from a small neighborhood of the previous layer, known as the receptive field. CNN may contain one or more convolutional layers. In the case of more than one convolutional layer, the activation map from one layer is fed as input to the next convolutional layer and so on. Second layer of a CNN contains activation functions for the identification of like features. The third layer, the pooling layer is the sub sampling or down sampling layer. In the convolutional layer, the spatial volume of the input is preserved. The pooling layer shrinks the input volume spatially. It reduces the resolution of the features. Down sampling operations are performed on every single activation map independently. Two types of pooling that are commonly used are: average pooling and max pooling. The fully connected layer is the final layer of the CNN. This layer contains neurons that connect to the entire input volume as in an ordinary neural network. The specified target output scores are obtained from this layer. [51]

2.7.2 Long Short-Term Memory (LSTM)

A **Recurrent Neural Network (RNN)** is a type of artificial neural network; it uses sequential data or time series data. While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of RNNs depends on the prior elements within the

sequence. RNNs are commonly used for power quality event detection and classification, where the goal is to identify the occurrence and duration of disturbances in a given time series and assign a class or label to each time series segment [50]. However, traditional RNNs suffer from the "vanishing gradient" problem, where the gradients diminish exponentially as they propagate back through time. This makes it difficult for the network to capture and propagate long-range dependencies in the data [52]. **Long short-term memory (LSTM)**, on the other hand, is a specific type of RNN architecture designed to address the vanishing gradient problem and handle long-term dependencies more effectively [53]. Unlike standard feedforward neural networks, LSTMs have feedback connections, which allow them to process entire sequences of data, such as speech or video and not just single data points like images [54]. It introduces memory cells and gating mechanisms to the network. LSTM networks have memory cells that can store and output information over extended periods. Each memory cell has three main components: an input gate, a forget gate, and an output gate. These gates regulate the flow of information into and out of the memory cell, allowing the network to decide how much information should be remembered, forgotten, and output at each time step. This enables LSTMs to capture and retain long-term dependencies in the data more effectively compared to traditional RNNs [55]

2.8 Conclusion

In this chapter, the focus was placed on signal processing techniques that were used to extract optimal features from disturbance signals. The use of artificial intelligence techniques for classifying disturbances and the importance of both classification and regression in power quality for identifying and quantifying different types of disturbances were also covered. Additionally, the application of deep learning techniques in power quality analysis was discussed. Convolutional Neural Networks (CNN), Support Vector Machines (SVMs), and Long Short-Term Memory (LSTM) techniques are widely employed in the field for their effectiveness in accurately detecting and classifying power quality disturbances.

3.1. Introduction

The CNN, SVM, and LSTM are methods used for the detection and classification of power quality disturbances. CNNs are well suited for capturing spatial dependencies and local patterns in power quality disturbances. while, LSTM is suitable for temporal data and capturing temporal dependencies. SVM, on the other hand, is a powerful classification algorithm known for its ability to handle high-dimensional data and non-linear decision boundaries.

3.2. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN), as a deep learning method, has been widely used in the field of pattern recognition and image classification. The architecture of CNN is analogous to the connective pattern of neurons in the human brain. A typical CNN consists of two components: a feature extractor and a classifier. The feature extractor is used to filter input images into feature maps that represent a variety of features of the image. These features may include corners, lines, edges, etc., which are relatively invariant to position shifting or distortions. The output of the feature extractor is a low-dimensional vector composed of these features. Then this vector is fed into the classifier, which is usually based on traditional Artificial Neural Networks. The purpose of this classifier is to decide the probability of categories that the input (e.g. Image) might belong to [56].

Time-domain signals are used as images of the input dataset for a 2D Convolutional Neural Network (CNN). In this approach, each signal is treated as a 1D image with one dimension representing time and the other dimension representing the signal amplitude. To use time-domain signals as input to a 2D CNN, each signal needs to be converted into an image format.

One common approach is to use windowing by dividing each sampled signal into fixed-length segments by choosing a window size (number of samples) that captures the relevant temporal information, and then stack the frames together to form a 2D image.

The signal is represented as a 2D image, 2D CNN architectures can be applied to extract features and perform classification, regression, or other signal processing tasks. The advantage of using the time-domain signals as input to a 2D CNN is that it preserves the temporal information of the signal, which can be important for certain signal processing tasks.

3.3 Layers of a Convolutional Neural Network (CNN)

One of the key elements that make CNNs powerful is their unique architecture, which is characterized by multiple layers. Each layer in a CNN plays a specific role in extracting and transforming features from input images. These layers work in a hierarchical manner, progressively learning more abstract and complex representations as information flows through the network. The core layers in a typical CNN architecture include convolutional layers, batch normalization layers, Rectified linear unit layers, pooling layers, and fully connected layers.

3.3.1 Convolutional Layer

The principle of the convolution is to slide across the input image from left to right and from top to bottom using a specific size window. The sliding stride usually uses the same distances in the height and width directions. The sliding window in the CNN is called the filter (or kernel), and the area slipped by the filter is called the receptive field. The matrix and the pixel values of the images multiply when the convolutional layer passes the filter (also called the dot product); then, the multiplied values are added, and a deviation value is added, as in Equation (3.1). Every position of the input image has a value; when the filter is slipped over all of the positions, a feature map is generated (Fig 3.1). Filters can be regarded as feature identifiers, which are used to extract various features from the image. The more filters there are, the deeper the feature maps are, and it is easier to identify the images. Therefore, the purpose of the CNN's convolutional layer is to retain the spatial arrangement of the images and to obtain the local images to be the input features. The filter can acquire each local feature from the images to be the input of the next layer. During the training process of the convolutional layer, all filters of the shapes in the images according to the input patterns are determined. This way, the CNN retains the total calculation of each local area, and the convolutional layers concentrate on obtaining the local features of the images [57].

$$y = \sum(xij.fij) + b \tag{3.1}$$

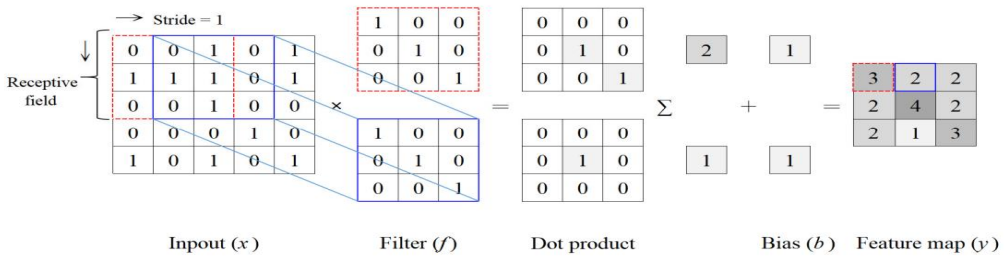


Fig 3.1 The calculation process for the features of the convolutional layer [57]

In addition to determining the size of the filter, strides and paddings must be set up. The distance that the filter moves is the stride, and the convolutional layer is the method for the CNN to extract the features of the input images by moving one unit each time to convolute the input data. When the unit of the stride movement is larger, the overlap rate of the receptive field decreases, which can reduce the calculation cost and could also lose the features of the input data. In addition, padding sets the pixel value to zero at the four boundaries of the input image, which allows the input and output image sizes to maintain the same size, and the output size is shown in Equation (2) [57].

$$O = \frac{M-F+2P}{S} + 1 \tag{3.2}$$

Where O, M, and F are the size of the output image, the input image, and the filter size, respectively; P and S are the padding and the stride, respectively.

3.3.2 Batch Normalization Layer

Batch normalization is a technique for training deep neural networks that standardizes the inputs to mini-batch layers. This process has the effect of stabilizing the learning process and reduces the number of training epochs significantly. BN is used to improve the training speed and the performance of the CNN model and reduce the sensitivity to the initialization of the weights. It also acts as a regularizer by reducing over-fitting and improving the generalization performance of the network [58]. The Batch Normalization (BN) helps to make the outputs from the upper layer conform to Gaussian distribution and its formula of BN is given as follows:

$$\widehat{x}_i = \frac{x_i - \mu_B}{\sqrt{{\sigma_B}^2 + \varepsilon}} \tag{3.3}$$

$$y_i = \gamma \widehat{x}_i + \beta$$

where, $\widehat{x}_i$ represents the normalized values of the mini-batch, ε is a small constant (epsilon) added for numerical stability to avoid division by zero, μ is the mean value of data $\widehat{x}_i$, and σ is the standard deviation. y_i represents the output of the batch normalization layer, γ is a learnable scale parameter, and β is a learnable shift parameter.

After the BN layers are integrated into CNN, the data from the upper layer is normalized. Compared to the CNN without BN layers, the CNN with BN layers can achieve much higher training efficiency. Therefore, the BN layer has become a standard configuration in DL [58].

3.3.3 Rectified Linear Unit (ReLU) Layer

A nonlinear activation function is used after each convolutional layer to introduce nonlinear features into the model. The sigmoid, tanh and rectified linear unit (ReLU) are the typical activation functions used by Convolutional Neural Networks. ReLU is introduced as a nonlinear activation function because it can increase the training speed and reduce the vanishing gradient without significant changes in accuracy. Since the gradients of ReLU are 0 for negative input and 1 for positive input, the calculation speed when using the sigmoid function is much faster. Equation (3.4) shows the ReLU function, where x and y are the input and output of the ReLU activation layer, respectively. The gradients of ReLU are always 0 and 1. A value less than 0 is output as 0, and a value greater than 0 is output directly.

$$\text{ReLU} = \max(0, x)$$

$$\frac{dy}{dx} = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{3.4}$$

3.3.4 Pooling Layer

Pooling reduces the dimensionality of the data obtained from the previous layer by down-sampling each feature map independently. This operation reduces the number of parameters which shortens the training time and combats over-fitting. The two commonly used pooling operations are :

• **Average Pooling** It calculates the average value for a specified window length of the feature map.

• **Max Pooling** It involves calculating the maximum value for a fixed window length of the feature map.

Among them, Max-pooling retains the maximum value, ignoring other values, which can reduce the impact of noise, improve the robustness of the model, reduce the number of model parameters, help reduce model overfitting problems, and be more suitable for power quality classification problems [59].

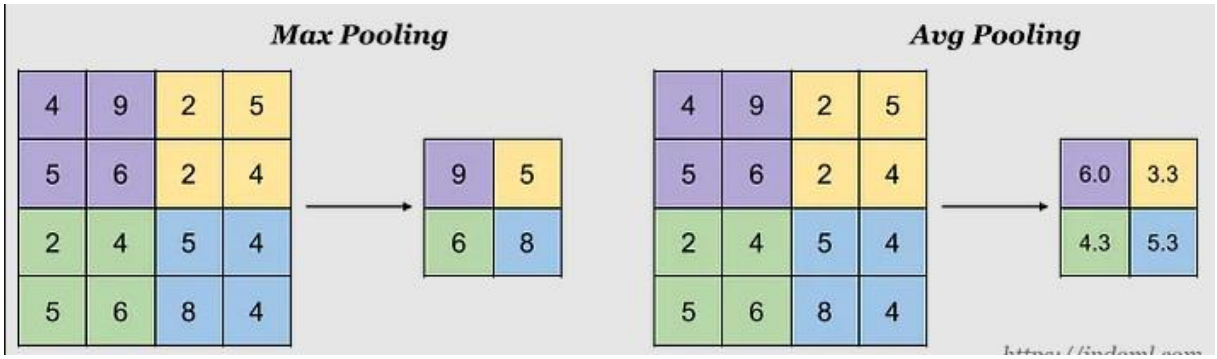


Fig 3.2 Max pooling and average pooling methods [60].

3.3.5 Flattening:

The process of pooling involves down-sampling the feature maps generated by convolutional layers, which helps to reduce the dimensionality of the input and extract important features. However, the resulting 2-dimensional arrays are not suitable for feeding directly into a fully connected layer, which requires a 1-dimensional input. To address this issue, the 2-dimensional arrays are flattened into a single linear vector. This is achieved by simply stacking the rows of the array on top of each other, resulting in a long continuous vector. The flattened vector is then fed as input to the fully connected layer, which performs the final classification of the image [61].



Fig 3.3 Flattening the pooled feature map [61].

3.3.6 Fully Connected Layer

Neural networks are a set of dependent nonlinear functions. Each individual function consists of a neuron (or a perceptron). The neurons in a neural network function as individual processing units, taking input from the previous layer and applying a linear transformation to it through a weights matrix. This output is then passed through a non-linear activation function, such as a

sigmoid or ReLU, which introduces non-linearity into the model. This nonlinearity is critical to the network's ability to learn complex patterns and relationships in the input data [62].

$$\boldsymbol{y_{jk}(x)} = f\left(\sum_{i=1}^{n_H} w_{jk}x_i + w_{j0}\right) \qquad (3.5)$$

Taking the dot product between the weights matrix W and the input vector x. The bias term (W0) can be added inside the nonlinear function. It does not affect the output sizes or decision-making and is just another weight.

A layer in a fully connected neural network with an input size of nine and an output size of four can be visualized as follows:

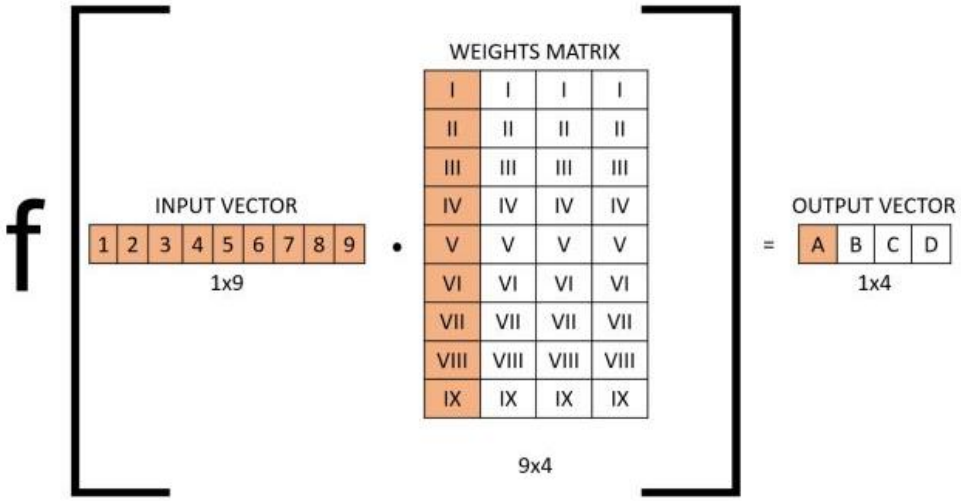


Fig 3.4 Visualized dot product of the weights matrix and the input vector [62]

The activation function “f” wraps the dot product between the input of the layer and the weights matrix of that layer. Note that the columns in the weights matrix would all have different numbers and would be optimized as the model is trained.

The input is a 1x9 vector, the weights matrix is a 9x4 matrix. By taking the dot product and applying the non-linear transformation with the activation function, an output vector (1x4) is obtained. This layer can also be visualized the following way:

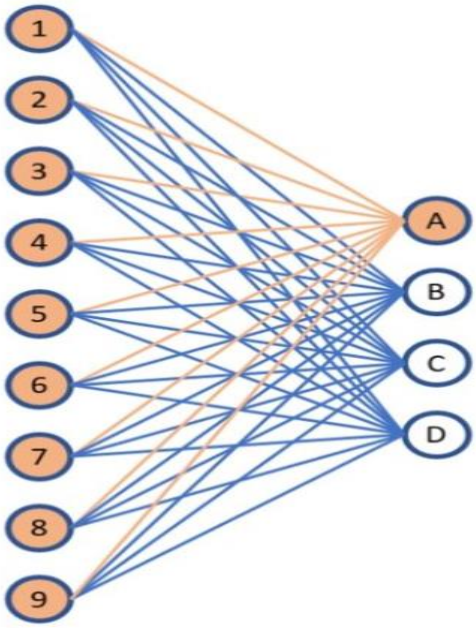


Fig 3.5 Fully connected layer [62].

The image above shows why these kinds of layers are called “fully connected” or sometimes “densely connected.” All possible connections layer-to-layer are present, meaning every input of the input vector influences every output of the output vector. However, not all weights affect all outputs. Looking at the lines between each node above, the orange lines represent the first neuron (or perceptron) of the layer. The weights of this neuron only affect output A, and do not have an effect on outputs B, C or D [62]

3.3.7 Softmax Activation Function

In Convolutional Neural Networks (CNNs), the Softmax function is commonly used as the final activation function in the output layer to convert the raw output of the network into a probability distribution over multiple classes. The Softmax function takes a vector of real numbers as input and normalizes it into a probability distribution. The output of the Softmax function is a vector of the same dimensionality as the input vector, with each element representing the probability of the corresponding class [63]. The Softmax function has several properties that make it suitable for multi-class classification tasks. It ensures that the probabilities sum up to 1, making it a valid probability distribution. It also amplifies the differences between the input values, emphasizing the highest value and suppressing the lower values. This property helps in identifying the most likely class based on the network's output [64]. In CNNs, the Softmax function is typically used in combination with a loss function like cross-entropy loss. The

network's parameters are trained to minimize the loss, which effectively maximizes the probability of the correct class and minimizes the probabilities of incorrect classes.

3.4 Loss Function

Choosing a suitable loss function is crucial for training Convolutional Neural Networks (CNNs) and optimizing their parameters. In general, loss functions are used to quantify the difference between the expected outcome and the outcome produced by the machine learning model [65]. The most common types of loss functions used in neural networks are cross-entropy and mean squared error.

Cross-Entropy Loss: Cross-entropy loss is widely used for multi-class classification tasks [66]. It quantifies the dissimilarity between the predicted class probabilities and the true class labels. The loss is minimized when the predicted probabilities match the true class labels. Cross-entropy loss is typically combined with the Softmax activation function in the output layer [67].
Root Mean Squared Error (RMSE) Loss: RMSE is commonly used for regression tasks, where the goal is to predict a continuous value. It calculates the square root of the average of the squared differences between the predicted values and the true values [68]. It provides an estimation of how well the model is able to predict the target value (accuracy). The smaller the RMSE, the better the model is at predicting the target values [69].

Root mean squared error can be expressed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}} \qquad (3.6)$$

While these are some commonly used loss functions in CNNs, there are other specialized loss functions tailored for specific tasks and scenarios. The choice of the loss function depends on the nature of the problem and the desired properties of the network's output. The right loss function can help the network train more efficiently and produce accurate predictions [69].

3.5 CNN Architectures

Various CNN architectures have been developed over the years. Some of the most common CNN architectures are: LeNet-5, AlexNet, VGG-16, and ResNet-50. The latter architecture is considered in this work [70].

3.5.1 ResNet-50

When deeper networks start converging, a degradation problem is exposed; with the network depth increasing, accuracy gets saturated and then degrades rapidly. Unexpectedly, such degradation is not caused by overfitting (usually indicated by lower training error and higher testing error) since adding more layers to a suitably deep network leads to higher training error.

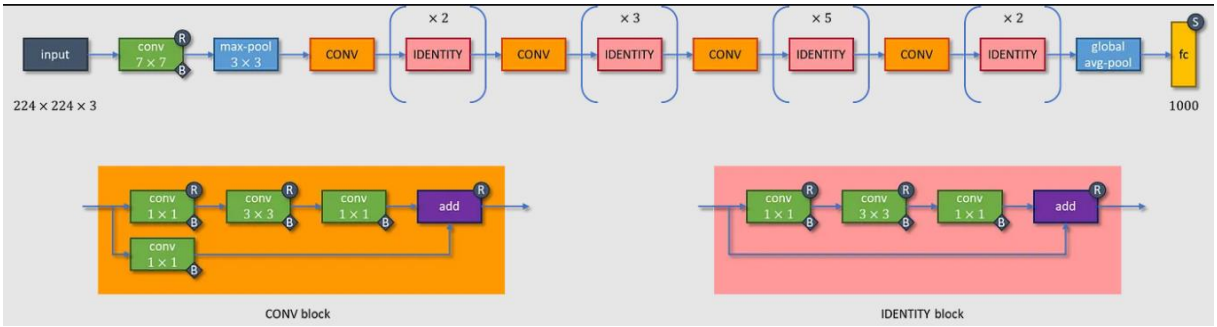


Fig 3.6 ResNet-50 Architecture [70].

The degradation problem is addressed by introducing bottleneck residual blocks. There are 2 kinds of residual blocks :

1. Identity block: consists of 3 convolution layers with 1×1, 3×3, and 1×1 kernel sizes, all of which are equipped with BN. The ReLU activation function is applied to the first two layers, while the input of the identity block is added to the last layer before applying ReLU.

2. Convolution block: same as identity block, but the input of the convolution block is first passed through a convolution layer with 1×1 kernel size and BN before being added to the last convolution layer of the main series.

Notice that both residual blocks have 3 layers. In total, ResNet-50 has 26 million parameters and 50 layers:

- 1 convolution layer with BN then ReLU is applied, followed by
- 9 layers that consist of 1 convolution block and 2 identity blocks, followed by
- 12 layers that consist of 1 convolution block and 3 identity blocks, followed by
- 18 layers that consist of 1 convolution block and 5 identity blocks, followed by
- 9 layers that consist of 1 convolution block and 2 identity blocks, followed by
- 1 fully connected layer with softmax.

The first convolution layer is followed by a 3×3 max-pooling and the last identity block is followed by a global-average-pooling. Default ResNet-50 accepts colored images with dimensions 224×224 and outputs one of the 1000 classes.

3.6 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are a type of supervised learning algorithms that can be used for classification or regression tasks. The main idea behind SVMs is to find a hyperplane that maximally separates the different classes in the training data. This is done by finding the hyperplane that has the largest margin, which is defined as the distance between the hyperplane and the closest data points from each class. Once the hyperplane is determined, new data can be classified by determining on which side of the hyperplane it falls. SVMs are particularly useful when the data has many features, and/or when there is a clear margin of separation in the data. In addition, it is capable of solving linear and non-linear problems with precision as well as handling binary and multiclass characterization, and its noise-insensitive state vectors make it an optimal choice for classification tasks. SVM utilizes a non-linear mapping of input vectors to a high-dimensional feature space in pattern recognition, which results in the construction of a linear decision surface. This transformation of input vectors to the feature space F changes the original nonlinear separation problem into a linear separation problem. The ultimate aim is to find the optimal hyperplane in the feature space that linearly separates the two classes. SVM achieves this by maximizing the distance between the closest points to the hyperplane [71]. To classify nonlinear applications, SVM requires kernel functions.

3.6.1 Kernel Functions

In Support Vector Machine (SVM) classification, the kernel function is used to transform the input data into a higher dimensional feature space where it can be more easily separated by a linear decision boundary (hyperplane). When using SVM for power quality disturbances classification, the choice of the kernel function can greatly affect the performance of the classifier [72]. The commonly used kernel function in SVM for power quality disturbances classification is **Radial Basis Function (RBF) Kernel.**

The RBF (Radial Basis Function) kernel is a popular kernel function used in machine learning and specifically in support vector machines (SVMs). It is commonly used for classification and regression problems. The RBF kernel is defined as:

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \, ||\mathrm{x} - \mathrm{x}'||^2) \qquad (3.7)$$

where x and x' are two input data points, ||.|| is the Euclidean distance between them, and γ is a parameter that controls the width of the kernel function. The RBF kernel function takes the input data points as arguments and returns a scalar value representing the similarity between them [71]. The RBF kernel has several properties that make it a popular choice in machine learning, including:

1. Flexibility: The RBF kernel can model complex nonlinear relationships between the input features and the output variable.

2. Universality: The RBF kernel can approximate any continuous function to arbitrary precision, given enough training data.

3. Computational efficiency: The RBF kernel is computationally efficient to evaluate [71].

The choice of the γ parameter can have a significant impact on the performance of the RBF kernel. A large value of γ will result in a narrower kernel, which may lead to over-fitting, while a small value of γ will result in a wider kernel, which may lead to under-fitting. The γ parameter can be tuned using cross-validation or other hyperparameter optimization techniques [71].

In general, the RBF kernel is the default choice and often performs well, but it is always a good practice to try different kernels and compare their performance on the specific dataset.

3.7 Hybrid CNN-SVM

A hybrid CNN-SVM approach combines the strengths of Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) to improve the detection and classification of power quality disturbances. This approach leverages the feature extraction capabilities of CNNs and the robust classification capabilities of SVMs [73]. In this approach, the CNN is first employed to extract relevant features from power quality waveforms. The CNN learns hierarchical representations and captures local patterns and spatial dependencies within the data. The output of the CNN is a feature representation that encodes important characteristics of the power quality signals. These extracted features from the CNN are then used as input to an SVM classifier. SVMs are known for their ability to handle high-dimensional data and nonlinear decision boundaries. The SVM analyzes the feature representation and trains a classification model to accurately classify the power quality disturbances into their respective classes, such as sags, swells, harmonics, or interruptions. By combining CNN and SVM, this hybrid approach takes advantage of the CNN's ability to automatically learn relevant features

from raw data and the SVM's strong classification performance. The hybrid CNN-SVM approach offers several advantages. Firstly, it can handle complex and non-linear relationships within the power quality data, as the CNN learns hierarchical representations and the SVM can model intricate decision boundaries. Secondly, the CNN component helps in reducing the dimensionality of the data and extracting meaningful features, enabling the SVM to achieve better classification performance. Lastly, the hybrid approach can handle imbalanced datasets and mitigate over-fitting by leveraging the regularization properties of the SVM. Overall, the hybrid CNN-SVM approach provides a powerful and effective solution for the detection and classification of power quality disturbances [73].

3.8 Long Short-Term Memory (LSTM)

A recurrent neural network (RNN) is a type of artificial neural network designed to process sequential data by introducing a feedback loop in its architecture. A drawback of RNNs lies in their inability to retain long-term dependencies effectively due to the issue of the vanishing gradient [74]. One solution to this problem in RNNs is to use a variant of the basic RNN architecture, such as long short-term memory (LSTM). LSTMs are designed to handle long-term dependencies in sequential data, which makes them well-suited for analyzing power quality data [75].

3.9 Architecture of LSTM

LSTMs address the problem of the vanishing gradient by introducing a memory cell, which is a container that can hold information for an extended period of time. The memory cell is controlled by three gates: the input gate, the forget gate, and the output gate. These gates decide what information to add to, remove from, and output from the memory cell [76].

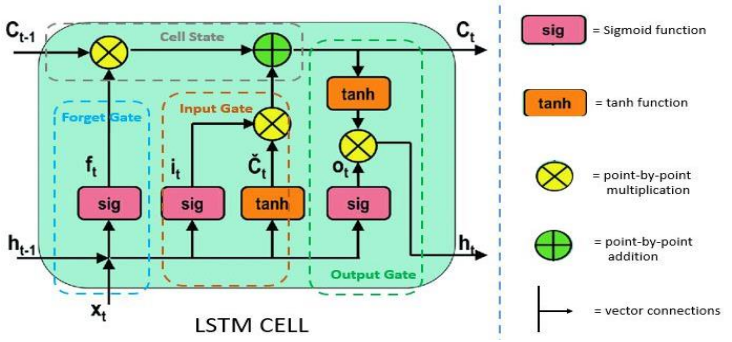


Fig 3.7 The LSTM Cell [81].

3.9.1 Forget Gate

The forget gate determines which information from the previous cell state should be forgotten and which information should be retained for the current time step. The forget gate takes the previous hidden state h(t-1) and the current input X(t) as inputs and outputs a number between 0 and 1 for each number in the cell state. A value of 0 means that the information should be completely forgotten, while a value of 1 means that the information should be completely retained. The forget gate is implemented using a sigmoid activation function, which squashes the output between 0 and 1 [77],[78].

3.9.2 Input Gate

The input gate determines which information from the current input should be added to the cell state. The current state X(t) and previously hidden state h(t-1) are passed into the second sigmoid function. The values are transformed between 0 and 1. A value of 0 means that the information should not be added to the cell state, while a value of 1 means that the information should be completely added to the cell state. Next, the same information of the hidden state and current state will be passed through the tanh function. To regulate the network, the tanh operator will create a vector C(t) with all the possible values between -1 and 1. The output values generated form the activation functions are ready for point-by-point multiplication [79],[80].

3.9.3 Output Gate

The output gate determines the value of the next hidden state. This state contains information on previous inputs. First, the values of the current state and previous hidden state are passed into the third sigmoid function. Then the new cell state generated from the cell state is passed through the tanh function. Both these outputs are multiplied point-by-point. Based upon the final value, the network decides which information the hidden state should carry. This hidden state is used for prediction. Finally, the new cell state and new hidden state are carried over to the next time step.

To conclude, the forget gate determines which relevant information from the prior steps is needed. The input gate decides what relevant information can be added from the current step, and the output gates finalize the next hidden state [81].

LSTMs excel in capturing and modeling the complex temporal dependencies present in power quality data. By explicitly addressing the issue of long-term dependencies, LSTMs offer a powerful solution for analyzing and predicting power quality parameters. Their ability to retain

information over extended periods allows for more accurate forecasting and detection of abnormalities, contributing to improved monitoring, control, and management of power systems. LSTMs have proven to be a valuable tool in enhancing power quality analysis and ensuring the reliability and stability of electrical grids [82],[83].

3.10 Performance Evaluation of Classification Models

There are many performance evaluation measures when it comes to selecting a classification model. Amongst them is the confusion matrix.

Confusion matrix

The confusion matrix is a size n x n table to view the results of predictive models, it is a tool for measuring the performance of classification models to 2 or more classes. In confusion matrix of binary classification, we get four possibilities:

• **True Positive TP** A true positive is an outcome where the model correctly predicts the positive class.

• **True Negative TN** A true negative is an outcome where the model correctly predicts the negative class.

• **False positive FP** A false positive is an outcome where the model incorrectly predicts the positive class.

• **False negative FN** A false negative is an outcome where the model incorrectly predicts the negative class. [84]

These four outcomes can be used to calculate the accuracy and error:

• **Accuracy:** This term tells us how many right classifications were made out of all the classifications. [85]

It can be written as:

$$Accuracy = \frac{TP+\mathrm{TN}}{total\ number\ of\ given\ predictions} \times 100\% \qquad (3.8)$$

• **Error:**

It is the number of false predictions over the total number of given predictions [86]

$$Error = \frac{FP+FN}{total\ number\ of\ given\ predictions} * 100\% = 100\% - accuracy\ (\%) \qquad (3.9)$$

The next figure shows the evaluation measures in a confusion matrix.

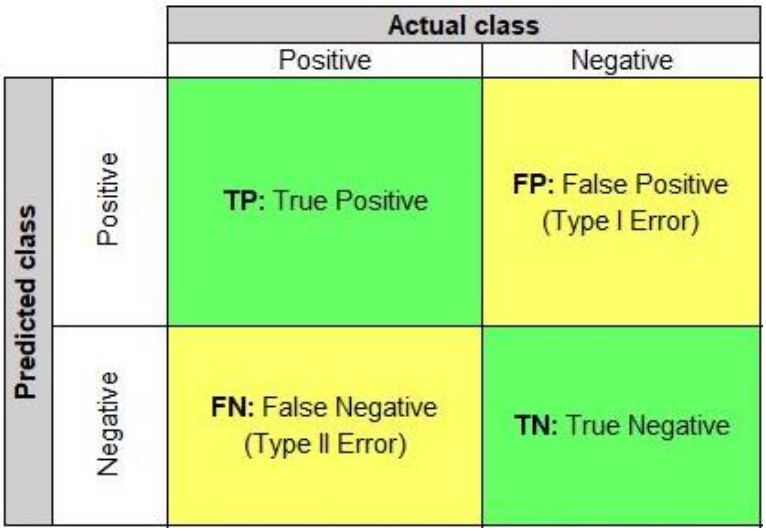


Fig 3.8 Evaluation Measures of Confusion Matrix [85]

3.11 Conclusion

This chapter explored the application of Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), and Long Short-Term Memory (LSTM) for the detection and classification of power quality disturbances. The application of these methods yielded valuable insights into achieving accurate detection and classification of such disturbances. These models offer advanced capabilities for power system monitoring and maintenance, facilitating more efficient and effective analysis of power quality data. The simulation and evaluation of the CNN-SVM and LSTM models can be performed to assess their respective performance and accuracy.

4.1 Introduction

In this chapter, MATLAB is employed to simulate PQD signals such as Sag, swell, interruption, and harmonics, which serve as inputs for evaluating the performance of the CNN-SVM and LSTM models. The primary objective of this evaluation is to accurately identify key characteristics of each disturbance, including its amplitude, start time, end time, and duration. By leveraging the capabilities of MATLAB simulation, the proposed detection methods aim to enhance the accuracy and effectiveness of PQD analysis and classification.

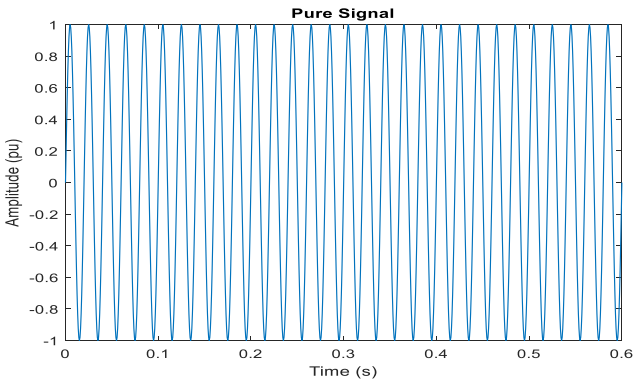
4.2 Generation of PQ Disturbances

Eight PQDs are considered in this chapter, which include 5 single PQDs (normal, sag, swell, harmonics, and interruption) and 3 multiple PQDs (sag with harmonics, swell with harmonics, and interruption with harmonics). The proposed PQD detection method uses the dataset, which is obtained from mathematical models given in Fig 4.1. These simulated models serve as reliable representations of actual power quality disturbances, aiding in comprehensive analysis and research.

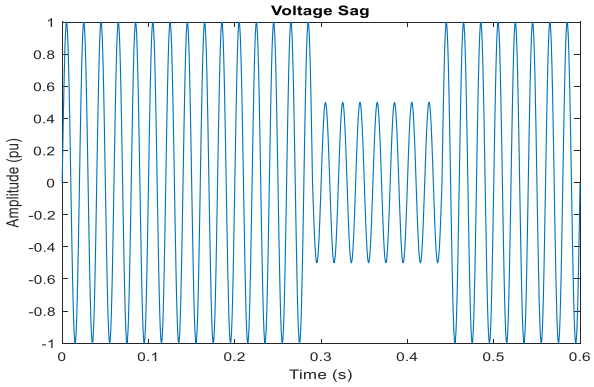
The parameters of the mathematical model are varied as per the direction of IEEE 1159 standard. In all the mathematical models of PQDs, the parameter “A” indicates the amplitude of the waveform and is set to 1 per unit. The factor “α” shows the amplitude of the disturbance. The step function u(t) of the model presents the time span of PQD in a signal. Harmonic disturbance data use the 3rd, 5th, and 7th orders. The sampling frequency of 3200 Hz is set because electric power signal recording equipment mostly uses the chosen sampling frequency. The generated signal contains 640 sampling points (10 cycles) with the fundamental frequency of 50 Hz. In each PQD case, 2082 data have been generated by varying parameters, and hence, a total of 16,656 (2082 ×8 PQDs) data instances have been generated. From this data, 70% are used for training and 30% for testing. Each disturbance starts from 0.29s and ends at 0.44s. The characteristic PQ signals simulated by mathematical model equations are shown in Fig 4.2 and Fig 4.3.

PQ disturbance	Class symbol	Model	Parameters
Pure signal	S1	$f(t) = A\sin(\omega t)$	Frequency: 50 Hz, $A = 1$
Sag	S2	$f(t) = A(1 - \alpha(u(t - t_1) - u(t - t_2)))\sin(\omega t)$	$0.1 \le \alpha \le 0.9$, $T \le t_2 - t_1 \le 9T$
Swell	S3	$f(t) = A(1 + \alpha(u(t - t_1) - u(t - t_2)))\sin(\omega t)$	$0.1 \le \alpha \le 0.8$, $T \le t_2 - t_1 \le 9T$
Interruption	S4	$f(t) = A(1 - \alpha(u(t - t_1) - u(t - t_2)))\sin(\omega t)$	$0.9 \le \alpha \le 1$, $T \le t_2 - t_1 \le 9T$
Harmonic	S5	$f(t) = A(\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t))$	$0.05 \le \alpha_3 \le 0.15$, $0.05 \le \alpha_5 \le 0.15$, $0.05 \le \alpha_7 \le 0.15$, $\Sigma\alpha_i^2 = 1$
Transient	S6	$f(t) = (\sin(\omega t) + \alpha_{osc}\exp(-(t - t_1)/\tau_{osc}))\sin(\omega_{nosc}(t - t_1))$	$\tau_{osc} = 0.008$–0.04 sec, $\omega_{nosc} = 100$–400 Hz
Sag + harmonic	S7	$f(t) = A(1 - \alpha(u(t - t_1) - u(t - t_2)))(\alpha_1 \sin(\omega t) + \alpha_3 \sin(\omega t) + \alpha_5 \sin(5\omega t))$	$0.1 \le \alpha \le 0.9$, $T \le t_2 - t_1 \le 9T$, $0.05 \le \alpha_3 \le 0.15$, $0.05 \le \alpha_5 \le 0.15$, $0.05 \le \alpha_7 \le 0.15$, $\Sigma\alpha_i^2 = 1$
Swell + harmonic	S8	$f(t) = A(1 + \alpha(u(t - t_1) - u(t - t_2)))(\alpha_1 \sin(\omega t) + \alpha_3 \sin(\omega t) + \alpha_5 \sin(5\omega t))$	$0.1 \le \alpha \le 0.8$, $T \le t_2 - t_1 \le 9T$, $0.05 \le \alpha_3 \le 0.15$, $0.05 \le \alpha_5 \le 0.15$, $0.05 \le \alpha_7 \le 0.15$, $\Sigma\alpha_i^2 = 1$
Flicker	S9	$f(t) = (1 + \alpha_f \sin(\beta_f \omega t))\sin(\omega t)$	$\alpha_f = 0.1$–0.2, $\beta_f = 5$–10 Hz

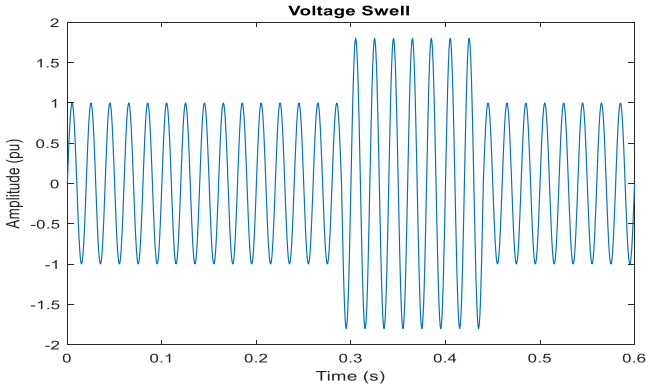
Fig 4.1 PQ disturbance models [87]



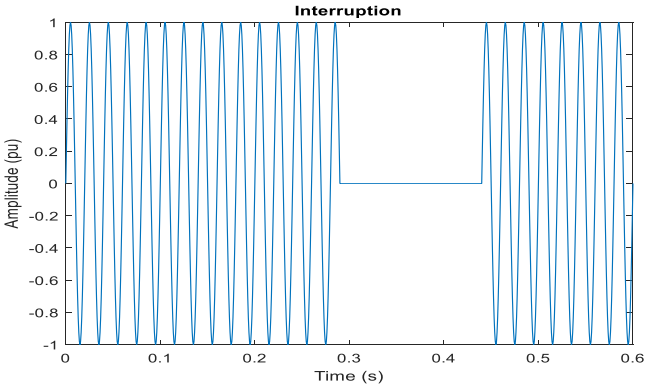
a) Normal Waveform.



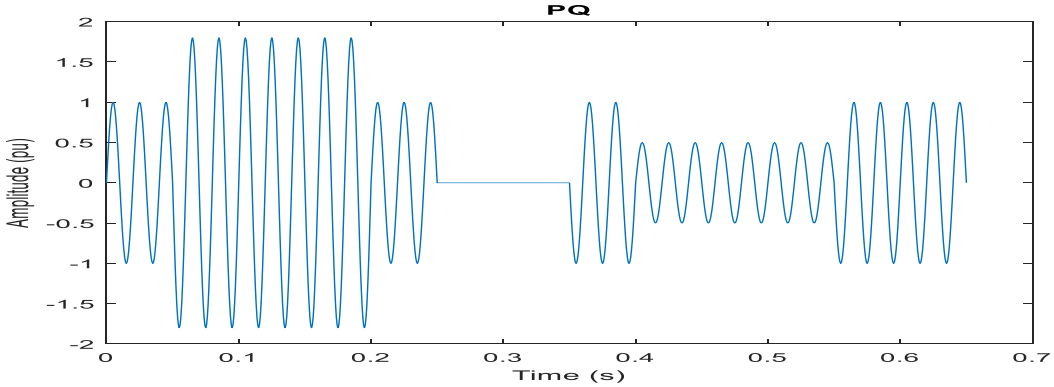
b) Voltage Sag.



c) Voltage Swell.

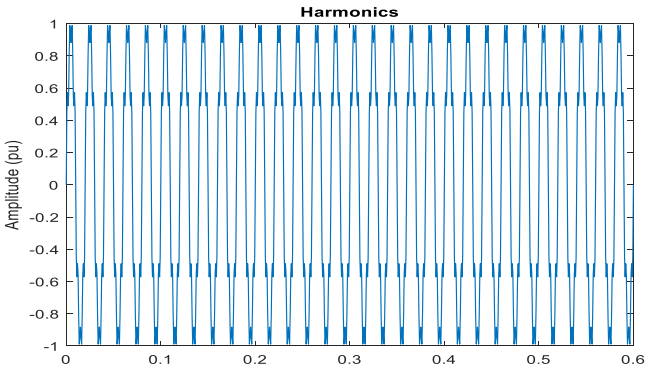


d) Voltage Interruption.

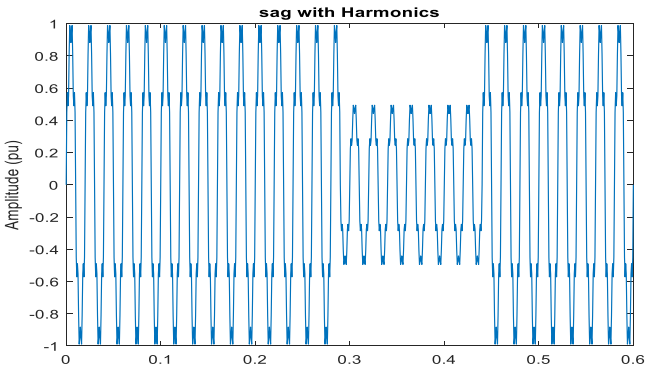


e) All PQ Events.

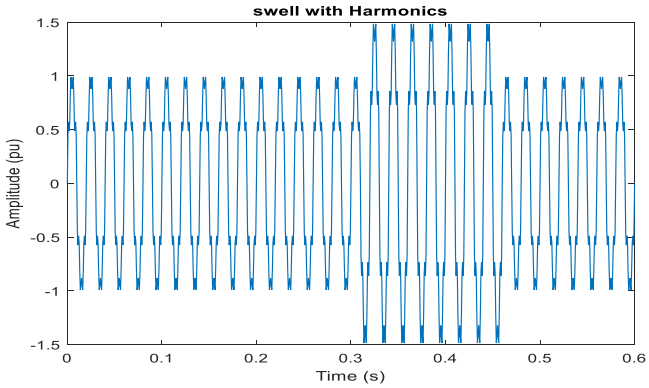
Fig 4.2 PQ Disturbances without Harmonics.



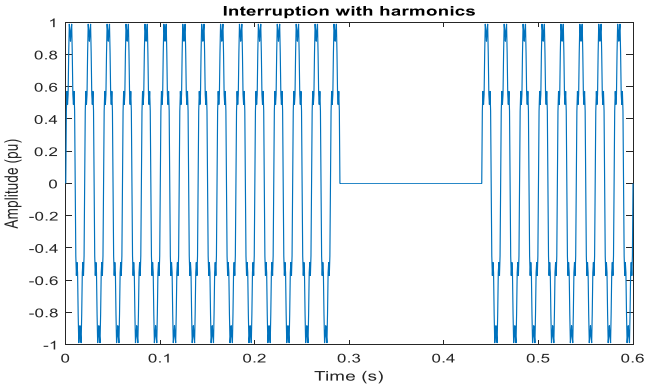
a) Harmonics.



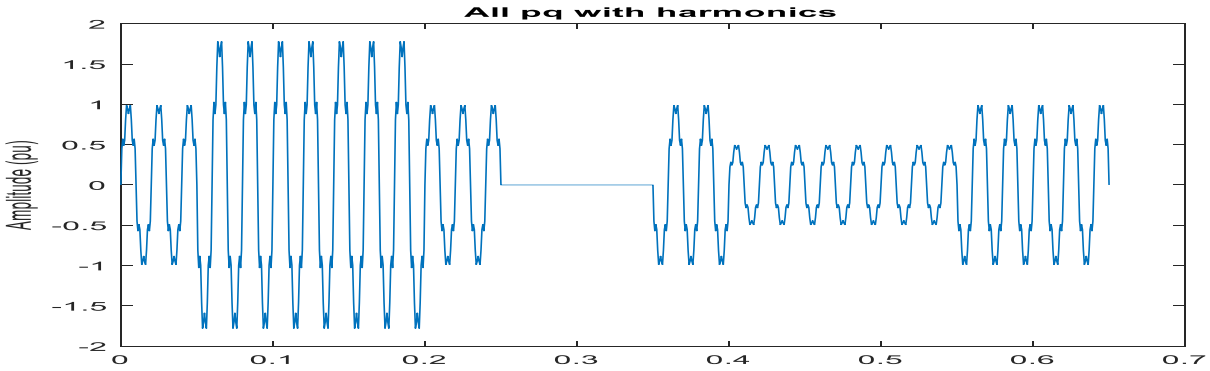
b) Voltage Sag.



c) Voltage Swell.



d) Voltage Interruption.



e) All PQ Events.

Fig 4.3 PQ Disturbances with Harmonics.

4.3 CNN-SVM model

The CNN model used in this chapter uses a pretrained ResNet-50 architecture model with the SVM classifier. The default values are used for this model, which are a stride of 2 and a padding size of 1. The stride determines the distance between consecutive filter placements, and the padding size determines the number of pixels added to the edge of an image to preserve its size after convolution. The filter (convolution kernel) used is of size 3×3. The cross-entropy loss function is used, and the classification layer is substituted with SVM for classification. Each image has 224x224 pixels as the input of CNN model. The output of the CNN-SVM system can assume values between 0 and 3 belonging to: normal, harmonics, sag, swell, and interruption. as shown in Table 4.1.

Table 4.1 CNN-SVM Outputs.

PQ Disturbance	CNN-SVM Output
Normal	1
Harmonics	3
Sag	0.5
Swell	1.8
Interruption	0

4.5 LSTM Model

The LSTM (Long Short-Term Memory) model used in this chapter is based on the traditional LSTM architecture. The default values are used for this model, including a single LSTM layer with a hidden size of 200 units and a Softmax activation function for classification.

For detecting and classifying the problem waveforms, the LSTM has two outputs:

• The first output of the LSTM function is for the waveform class which is defined by four sets (normal, interruption, sag, and swell). Any output value, which is not belonging to these sets, represents the distortion. The first output of the LSTM network system can assume values between 0 and 3 as shown in Table 4.2.

- The second output of the LSTM function is for harmonics indication, which is partitioned into two function sets. The labels of these sets are Pure and Harmonics as shown in Table 4.2.

Table 4.2 LSTM Outputs.

PQ Disturbance	LSTM Output
First LSTM Output	
Normal	1
Harmonics	3
Sag	0.5
Swell	1.8
Interruption	0
Second LSTM Output	
Pure	0
Harmonics	1

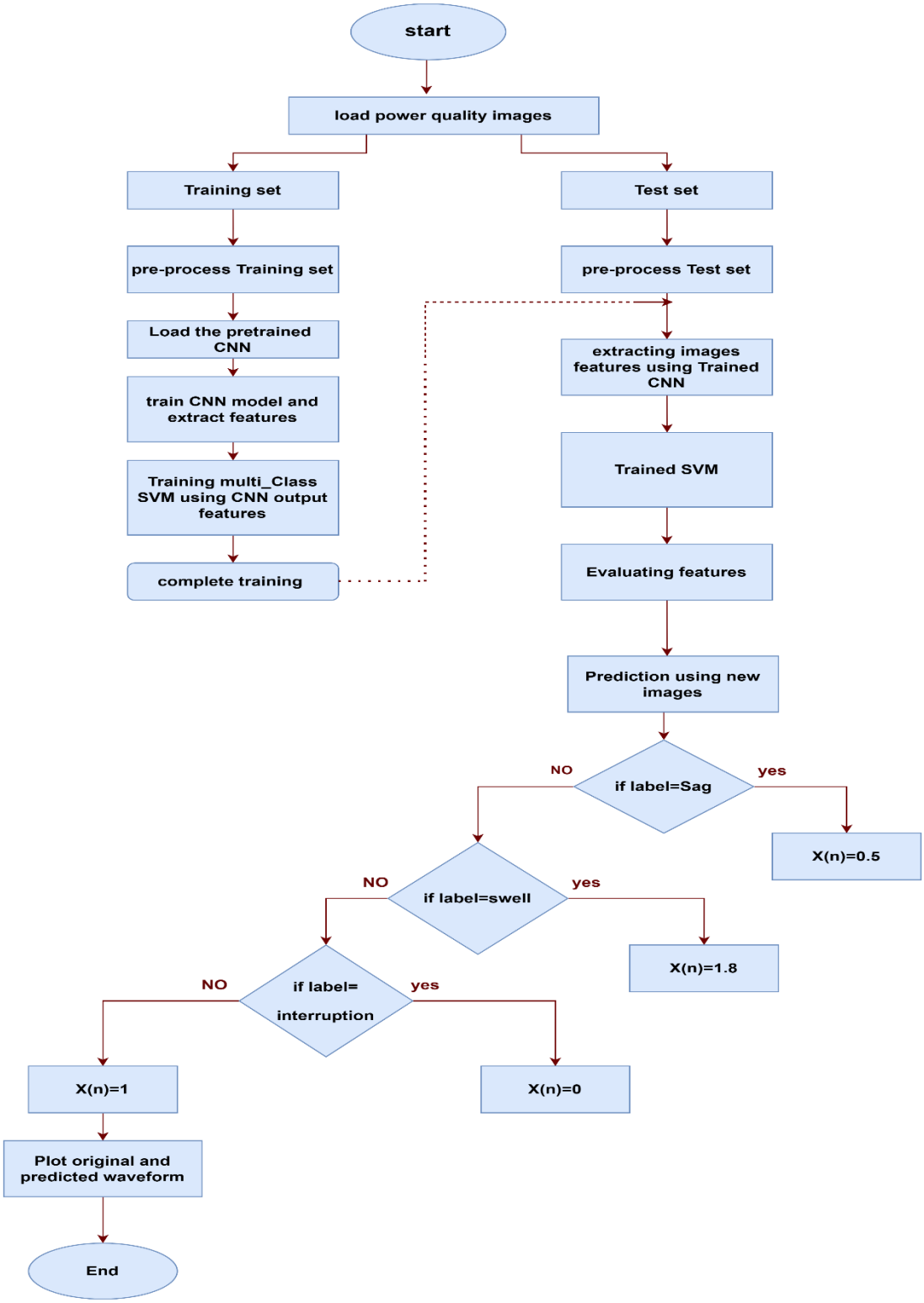
4.4 Proposed Technique

The CNN-SVM and LSTM architectures are represented as flowcharts as shown in Fig 4.4 (a) and (b), respectively.

CNN-SVM: The flowchart starts by loading the generated PQ data that is used to train the CNN model in order to extract important features. These features are utilized for training the SVM, which classifies the new image data into four classes.

LSTM: The flowchart starts by loading the pretrained LSTM model. Then, PQ signals are generated for different events. By using predict function, these signals are classified into four classes.

a) CNN-SVM Method.



b) LSTM Method

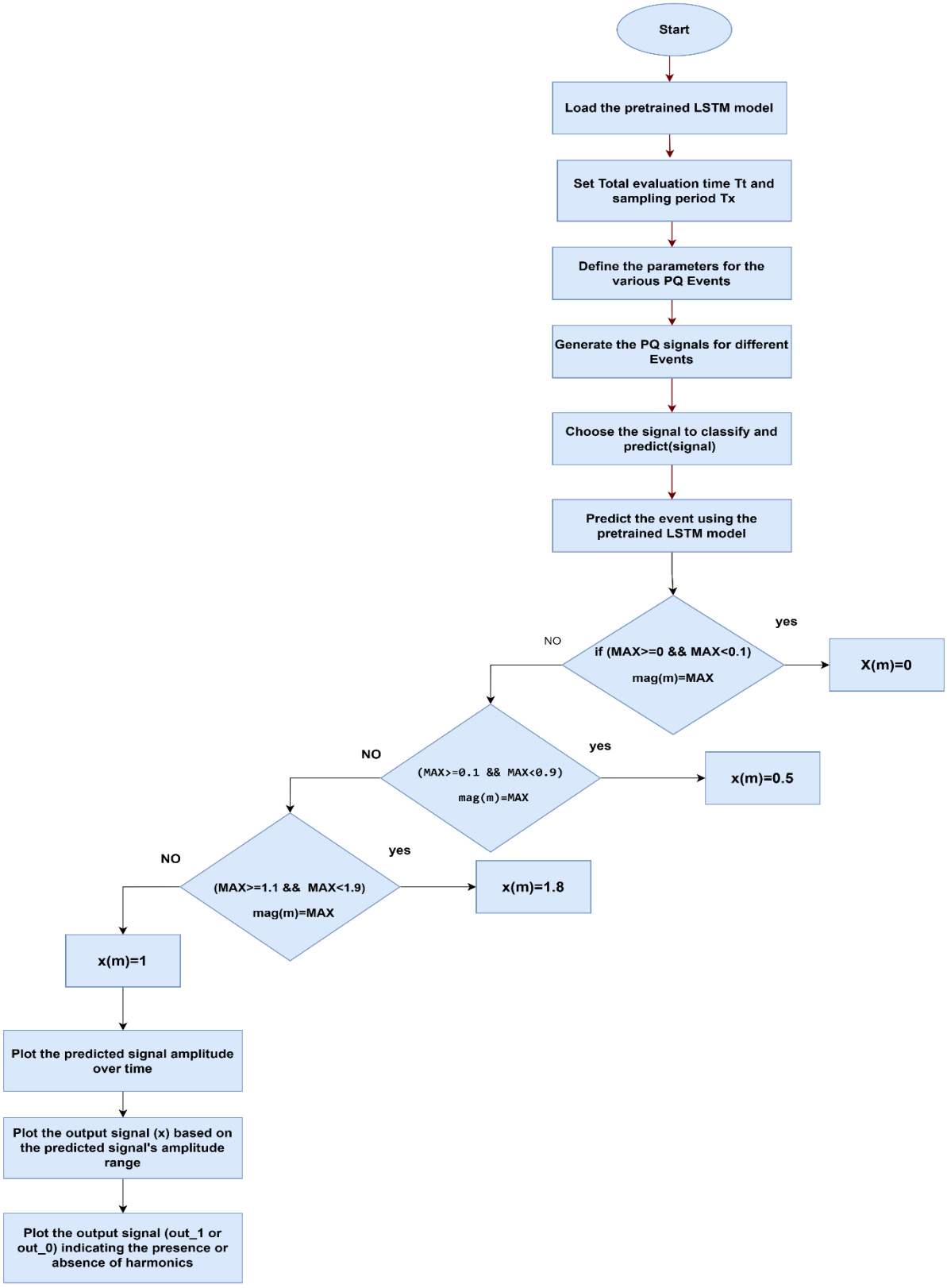


Fig 4.4 Flowcharts of the proposed methods: (a) CNN-SVM Method and (b) LSTM Method.

4.6 Normal voltage

The baseline scenario for voltage is referred to as the normal voltage, where the waveform remains pure without any variations in the magnitude of the RMS voltage.

CNN-SVM: Fig 4.4 (a) and (b) shows the typical voltage waveform and the CNN-SVM output, respectively. The CNN output equals 1, indicating that this is a pure sinwave.

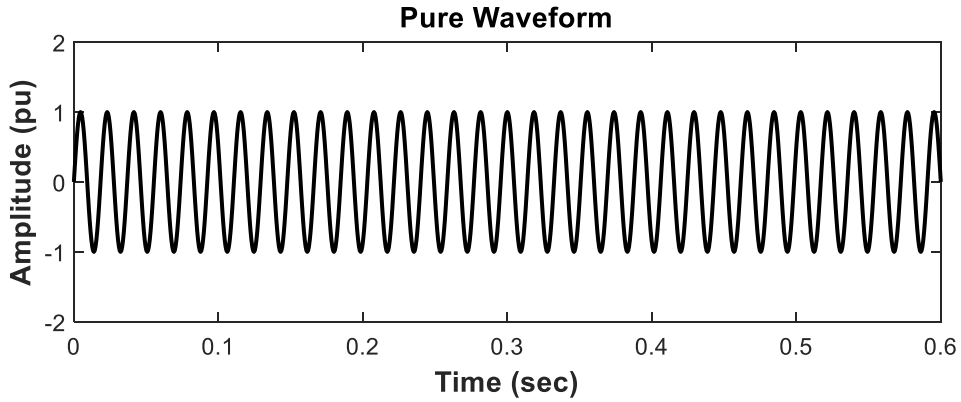
LSTM: Fig 4.5(a) and (b) shows the voltage waveform and its amplitude, respectively. Fig 4.5(c) and (d) shows the two LSTM outputs. The LSTM output 1 equals one, indicating that there is no disturbance, and the LSTM output 2 equals zero, indicating that this waveform has no harmonic distortion.

The magnitude, start and end time, duration, and accuracy for the normal waveform of each model are shown in Table 4.3.

Table 4.3 magnitude, start and end time, duration, and accuracy for the normal waveform.

	Magnitude (pu)	**Start Time**	**End Time**	**Duration (s)**	**Accuracy**
CNN-SVM	1	0	0.6	0.6	100%
LSTM	1.00193	0	1	1	100%

a)



b)

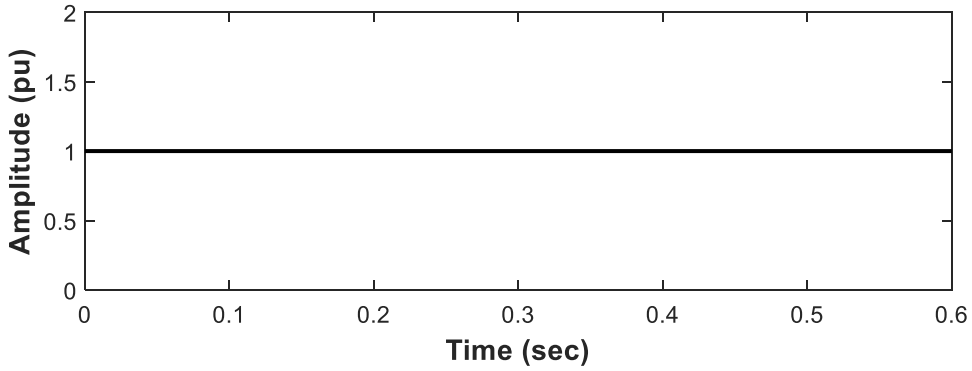
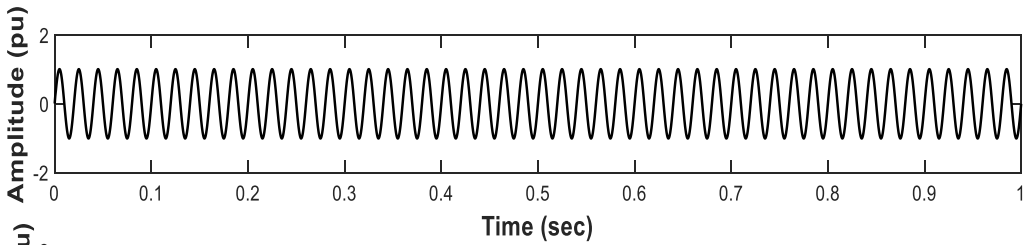
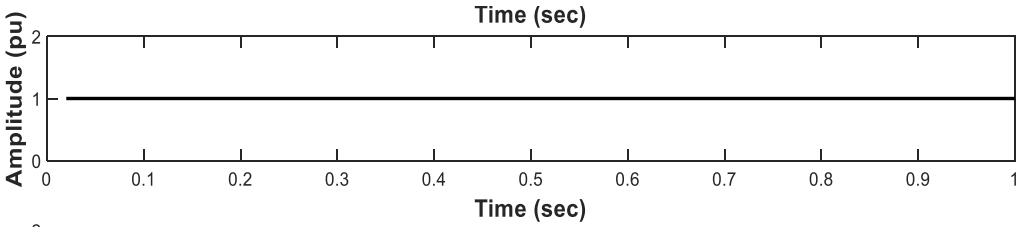


Fig 4.5 Normal voltage: (a) Waveform and (b) CNN-SVM Output.

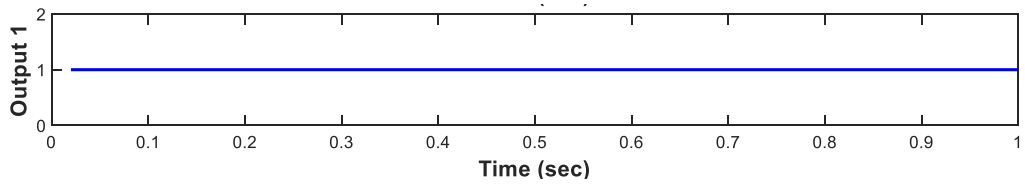
a)



b)



c)



d)

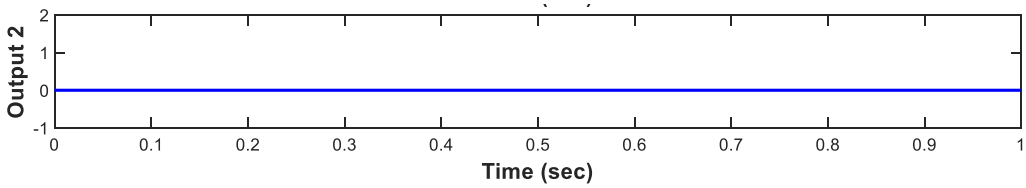


Fig 4.6 Normal voltage: (a) Waveform, (b) Amplitude, (c) LSTM output 1, and (d) LSTM output 2.

4.7 Voltage sag

CNN-SVM: The voltage sag waveform and its CNN-SVM output are shown in Fig 4.6 (a) and (b), respectively. The CNN-SVM model outputs 0.5 during the time of the disturbance indicating that there is a voltage sag, and outputs 1 when the waveform is normal.

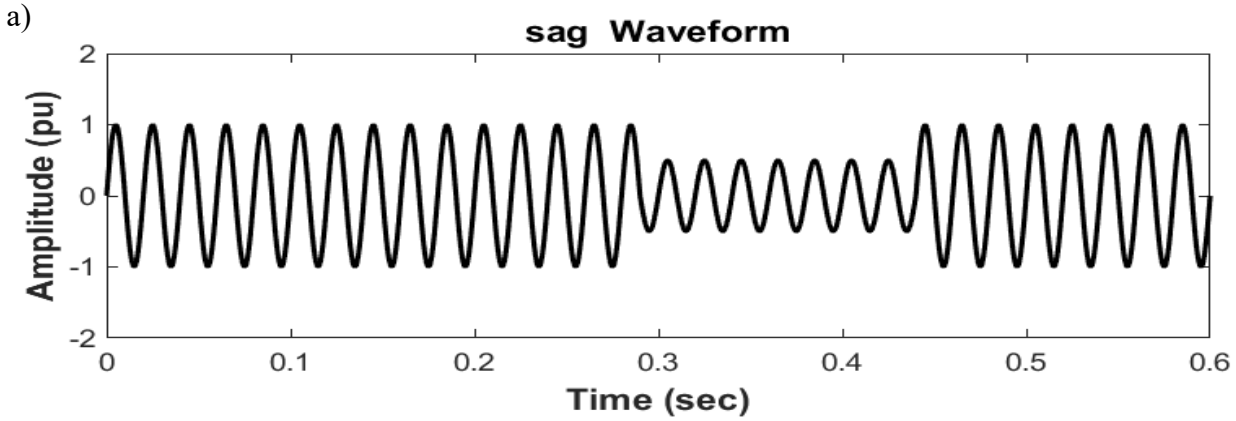
LSTM: The voltage sag waveform and its amplitude are shown in Fig 4.7 (a) and (b), respectively. The two LSTM outputs are shown in Fig 4.7 (c) and (d). The LSTM output 1 equals 0.5, which indicates that the waveform is a sag waveform. The LSTM output 2 equals zero, meaning that there is no harmonic distortion in this waveform.

The magnitude, start and end time, duration, and accuracy for the sag waveform of each model are shown in Table 4.4.

Table 4.4 magnitude, start and end time, duration, and accuracy for the sag waveform.

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
CNN-SVM	0.5	0.2944	0.4450	0.1506	98.52%
LSTM	0.5011	0.3062	0.4437	0.1375	97.98%

a)



b)

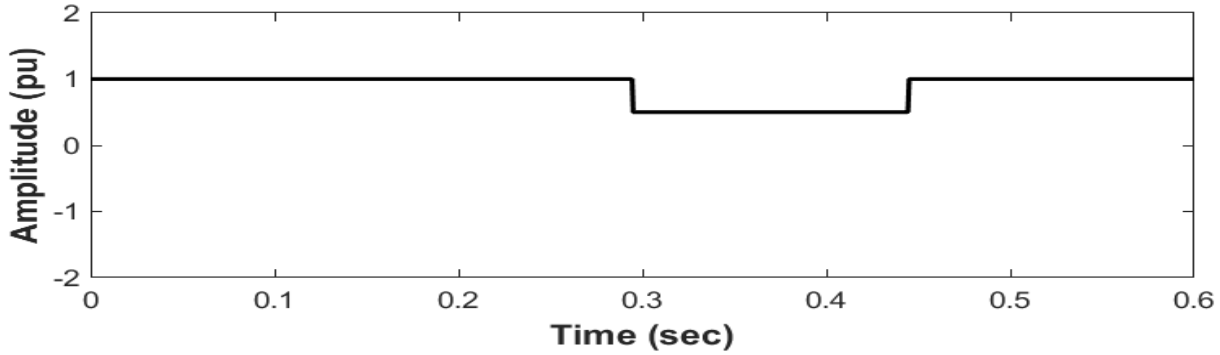


Fig 4.7 Voltage Sag: (a) Waveform and (b) CNN-SVM Output.

a)

b)

c)

d)

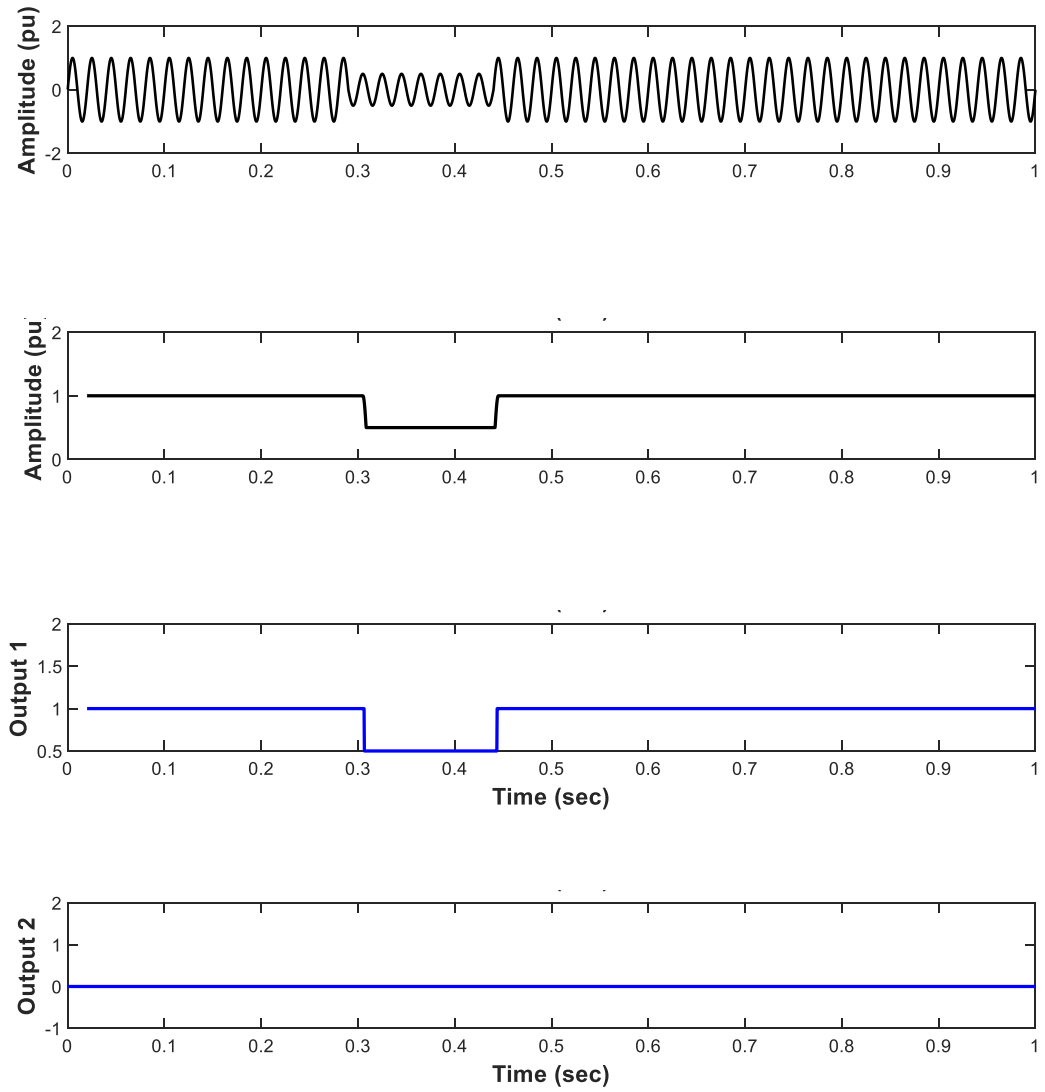


Fig 4.8 Voltage Sag: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output 2.

4.8 Voltage swell

CNN-SVM: The voltage swell waveform and its CNN-SVM output are shown in Fig 4.8 (a) and (b), respectively. The CNN-SVM outputs a 1.8 during the time of the disturbance, indicating that there is a voltage swell, and a 1 when the waveform is normal.

LSTM: The voltage swell waveform and its amplitude are shown in Fig 4.9 (a) and (b), respectively. The two LSTM outputs are shown in Fig 4.9 (c) and (d). The LSTM output 1 equals to 1.8, showing that the waveform is a sag waveform. The LSTM output 2 equals to 0, indicating that there is no harmonic distortion in this waveform.

The magnitude, start and end time, duration, and accuracy for the swell waveform of each model are shown in Table 4.5.

Table 4.5 magnitude, start and end time, duration, and accuracy for the swell waveform.

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
CNN-SVM	1.8	0.2944	0.4437	0.1493	98.51%
LSTM	1.7944	0.2949	0.4586	0.1637	98.33%

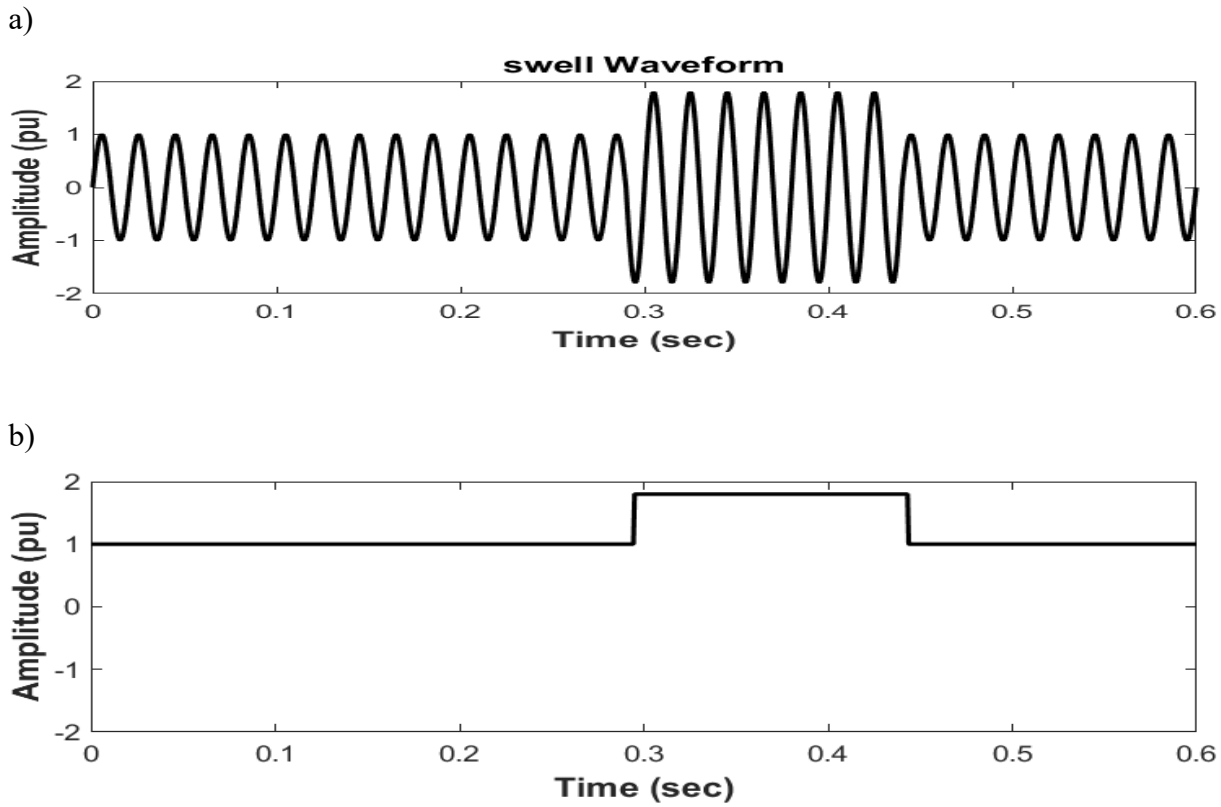
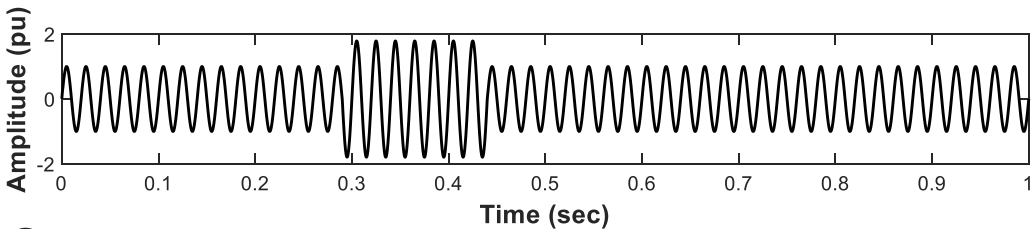
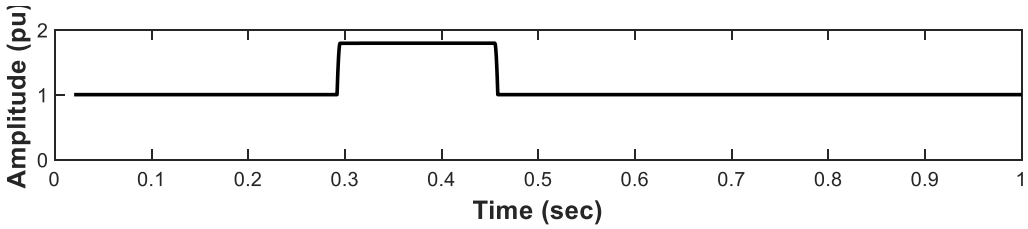


Fig 4.9 Voltage Swell: (a) Waveform and (b) CNN-SVM Output.

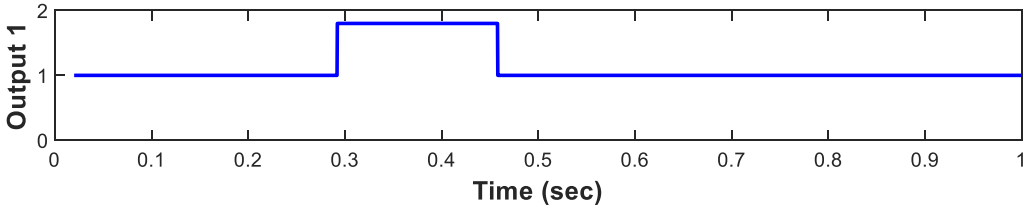
a)



b)



c)



d)

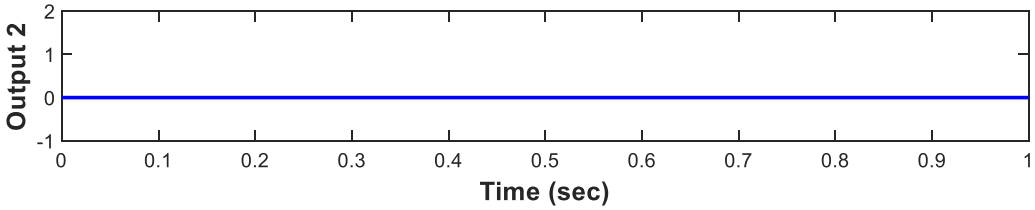


Fig 4.10 Voltage Swell: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output 2.

4.9 Voltage interruption

CNN-SVM: Fig 4.10 (a) and (b) depicts the voltage interruption waveform and its CNN-SVM output, respectively. During the disturbance time, the CNN-SVM output equals 0, indicating that the waveform is interrupted. Outside the range of the disturbance, the model's output equals 1, stating that the waveform is normal.

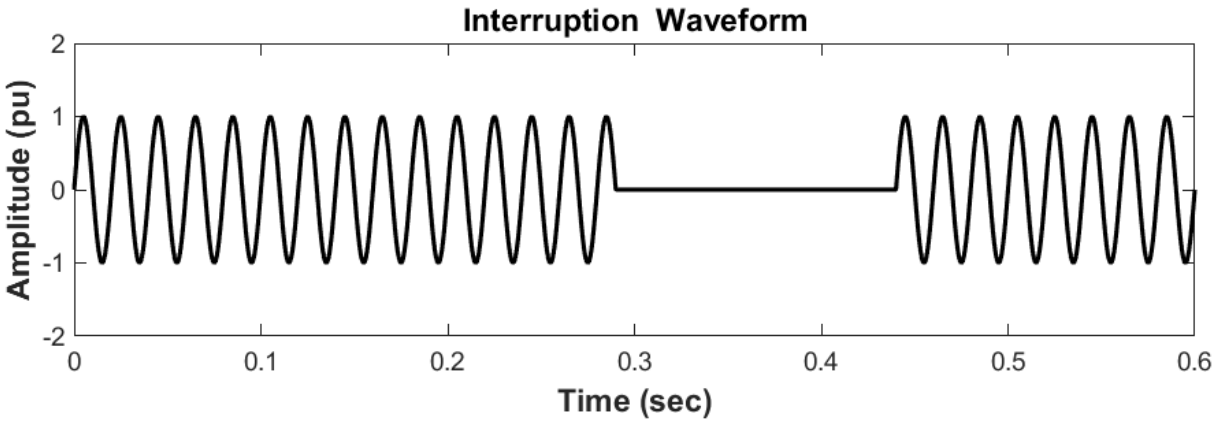
LSTM: Fig 4.11 (a) and (b) depicts the voltage interruption waveform and its amplitude, respectively. Fig 4.11 (c) and 4.6 (d) illustrates the two LSTM outputs. During the interruption time, the LSTM output 1 equals zero, indicating that the waveform is interrupted. The LSTM output 2 is zero, indicating that this waveform has no harmonic distortion.

The magnitude, start and end time, duration, and accuracy for the voltage interruption of each model are shown in Table 4.6.

Table 4.6 magnitude, start and end time, duration, and accuracy for the voltage interruption waveform.

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
CNN-SVM	0	0.2962	0.4437	0.1475	97.90%
LSTM	0.0003	0.3062	0.4437	0.1375	94.70%

a)



b)

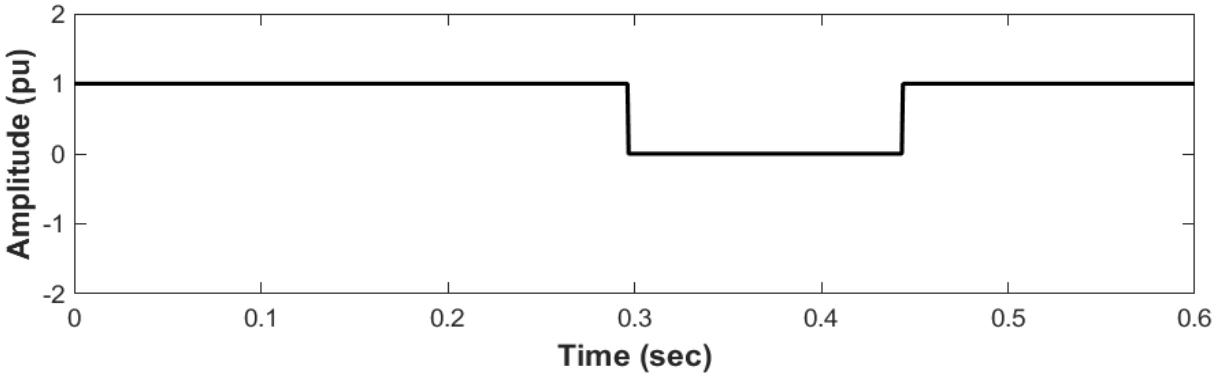


Fig 4.11 Voltage Interruption: (a) Waveform and (b) CNN-SVM Output.

a)

b)

c)

d)

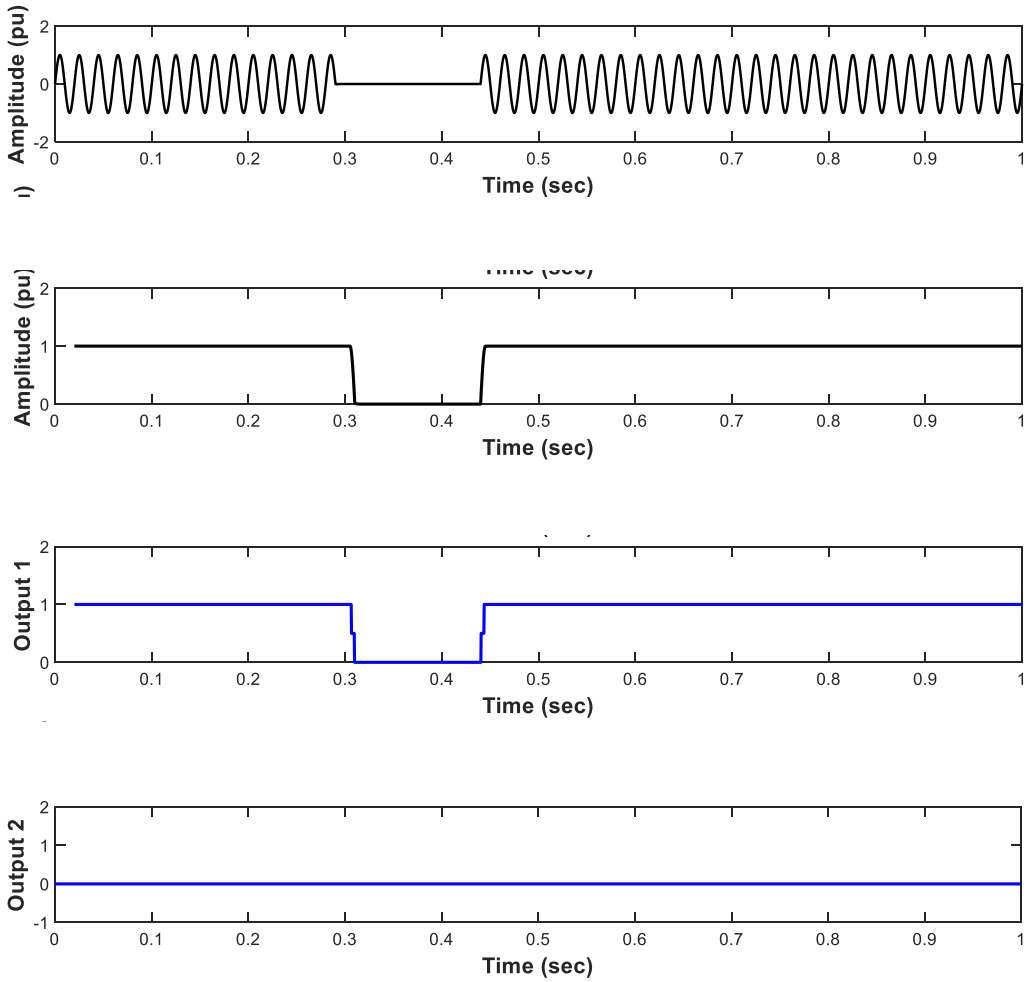


Fig 4.12 Voltage Interruption: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.

4.10 Harmonics

CNN-SVM: The voltage distortion waveform (harmonics) and its CNN-SVM output are shown in Fig 4.12 (a) and (b), respectively. The CNN-SVM output equals 3, which indicates the presence of harmonics.

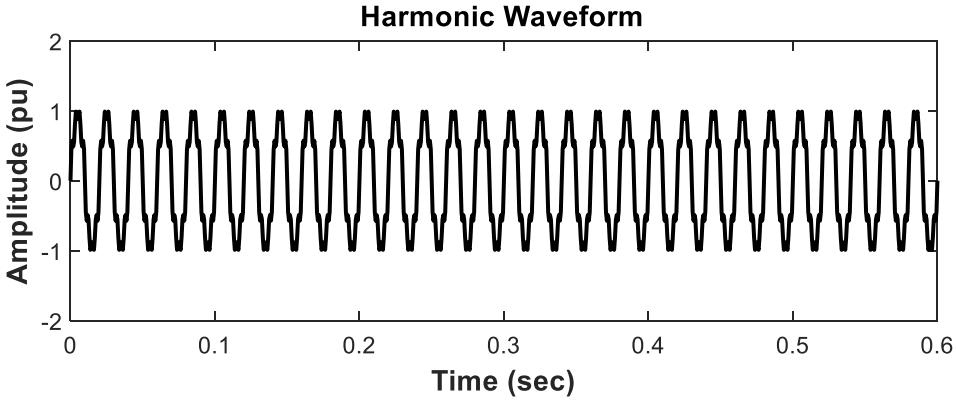
LSTM: The voltage distortion waveform and its amplitude are shown in Fig 4.13(a) and (b), respectively. The two LSTM outputs are shown in Fig 4.13 (c) and (d). The LSTM output 1 equals to 3 and the LSTM output 2 equals to one, which means that the voltage waveform contains harmonics.

The magnitude, start and end time, duration, and accuracy for the harmonics waveform of each model are shown in Table 4.7.

Table 4.7 magnitude, start and end time, duration, and accuracy for the harmonics waveform.

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
CNN-SVM	1	0	0.6	0.6	100%
LSTM	1.00861	0	1	1	100%

a)



b)

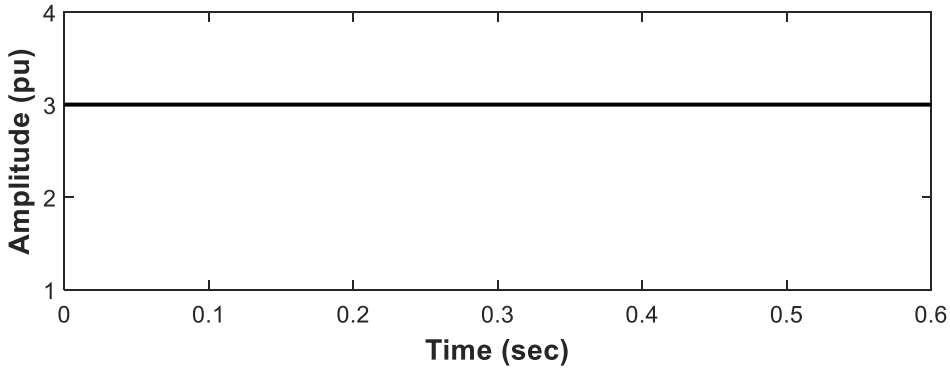
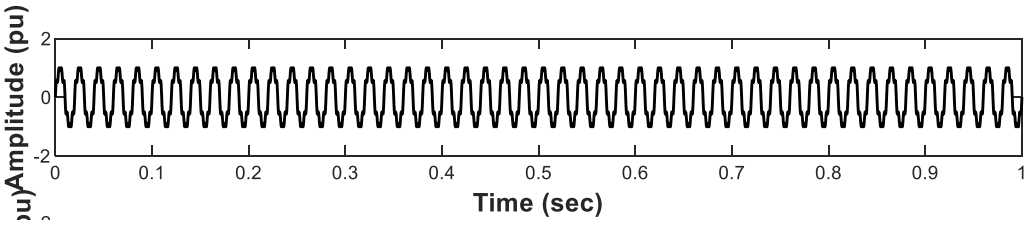
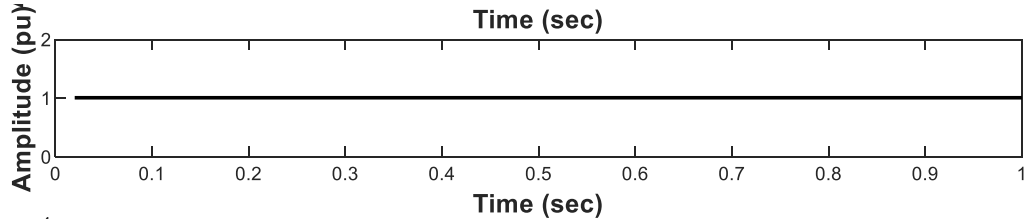


Fig 4.13 Harmonics: (a) Waveform and (b) CNN-SVM Output.

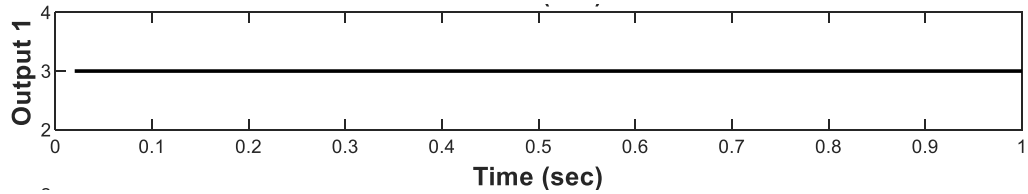
a)



b)



c)



d)

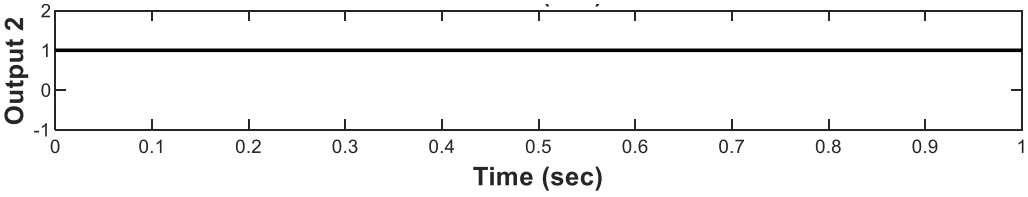


Fig 4.14 Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.

4.11 Sag with harmonics

CNN-SVM: The voltage sag with harmonics waveform and its CNN-SVM output are shown in Fig 4.14 (a) and (b), respectively. The model outputs a 0.5 indicating that there is a sag in the waveform, and a 3 outside the range of the sag, showing that the waveform is distorted.

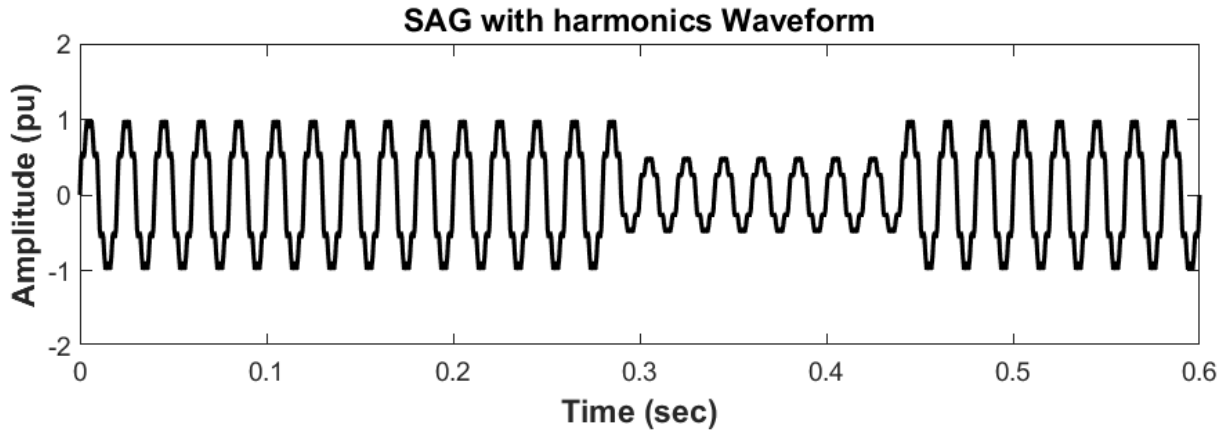
LSTM: The voltage sag with harmonics waveform and its amplitude are shown in Fig 4.15 (a)and (b), respectively. The two LSTM outputs are shown in Fig 4.15 (c) and (d). The LSTM output 1 equals to 0.5 during the time the sag is present in the waveform. The LSTM output 2 equals 1, which means that this sag waveform is distorted.

The magnitude, start and end time, duration, and accuracy for the sag with harmonics waveform of each model are shown in Table 4.8.

Table 4.8 magnitude, start and end time, duration, and accuracy for the sag with harmonics waveform.

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
CNN-SVM	0.5	0.2925	0.4475	0.1550	99.14%
LSTM	0.5043	0.3065	0.4434	0.1369	94.62%

a)



b)

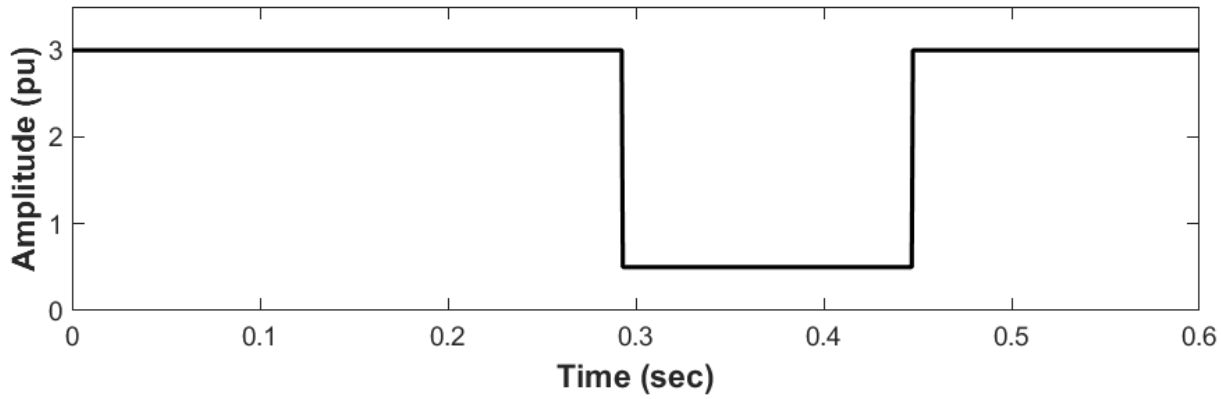
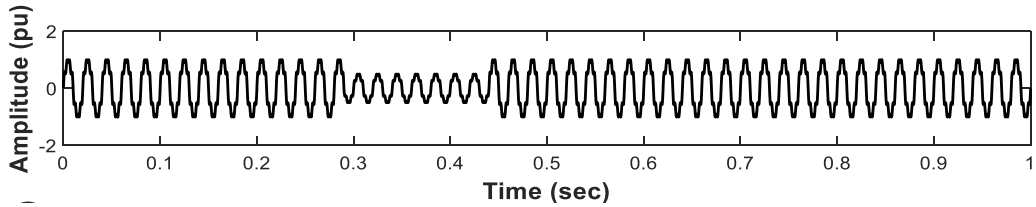
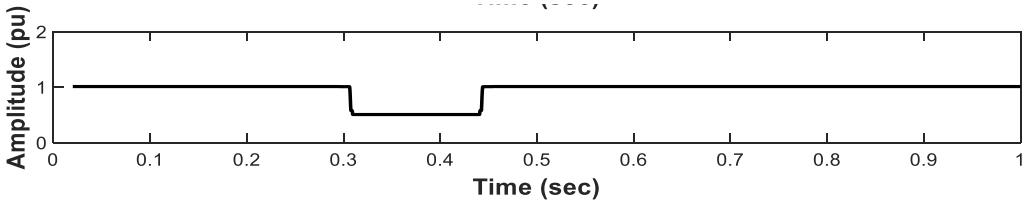


Fig 4.15 Sag with Harmonics: (a) Waveform and (b) CNN-SVM Output.

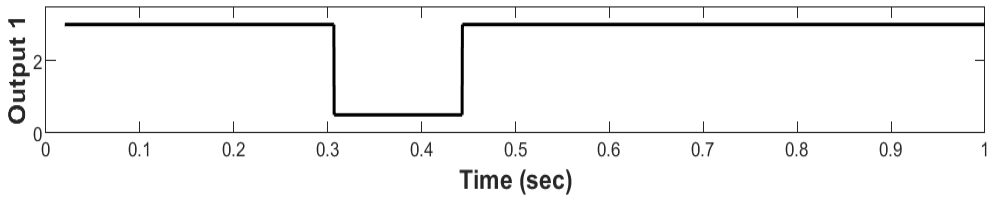
a)



b)



c)



d)

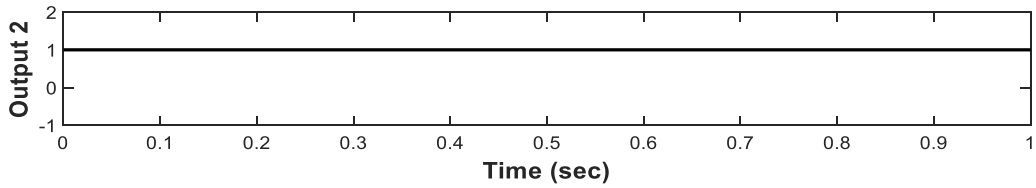


Fig 4.16 Sag with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.

4.12 Swell with harmonics

CNN-SVM: Fig 4.16 (a) and (b) depicts the swell with harmonics waveform and its CNN-SVM output, respectively. The output is 1.8 during the presence of swell in the waveform, and a 3 outside the range of the disturbance, indicating that there is a distortion in the waveform.

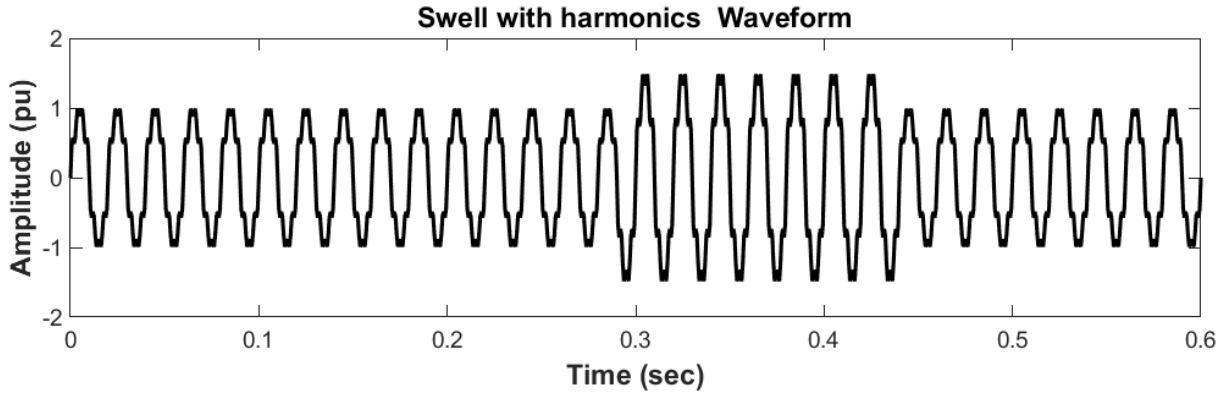
LSTM: Fig 4.17 (a) and (b) depicts the voltage swell with harmonics waveform and its amplitude, respectively. Fig 4.17 (c) and (d) illustrate the two LSTM outputs. The LSTM output 1 equals 1.5 when there is a swell in the waveform. This swell waveform is distorted because the LSTM output 2 equals one.

The magnitude, start and end time, duration, and accuracy for the swell with harmonics waveform of each model are shown in Table 4.9.

Table 4.9 magnitude, start and end time, duration, and accuracy for the swell with harmonics waveform.

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
CNN-SVM	1.5	0.2925	0.4544	0.1619	99.14%
LSTM	1.5116	0.2928	0.4568	0.1640	99.04%

a)



b)

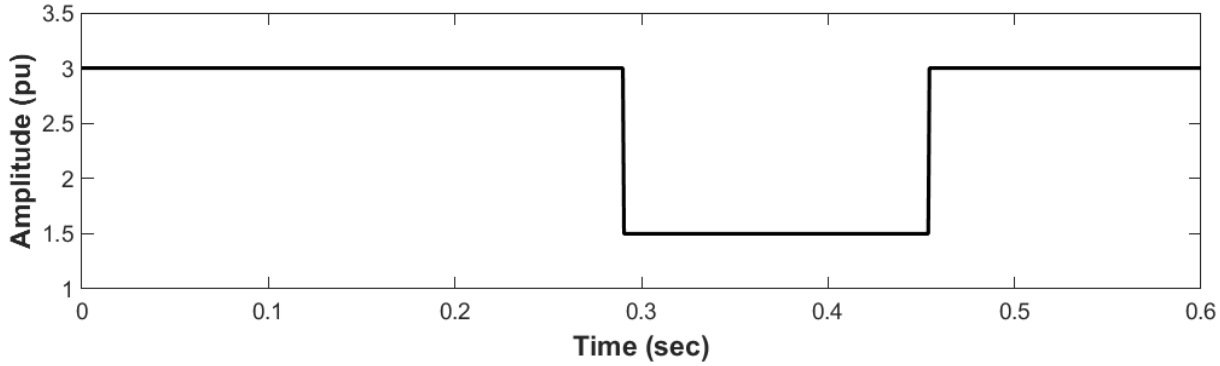
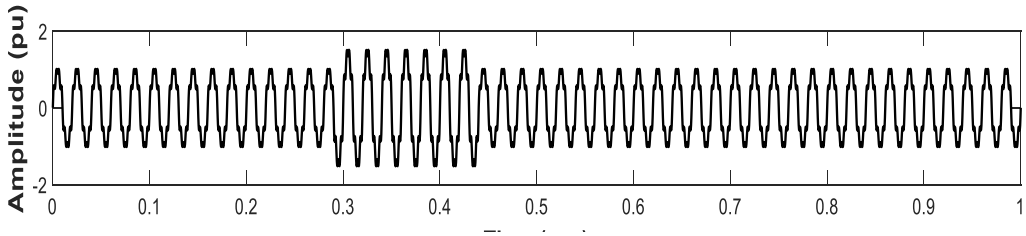
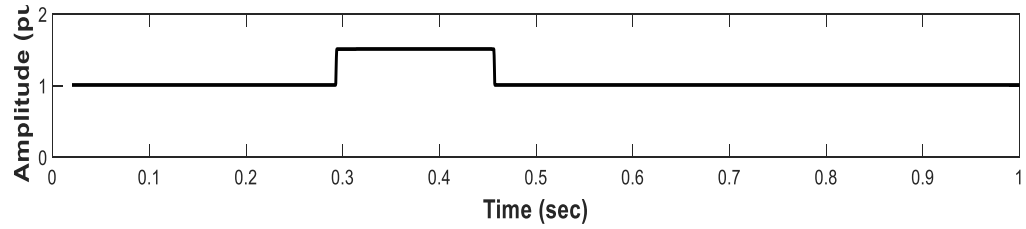


Fig 4.17 Swell with Harmonics: (a) Waveform and (b) CNN-SVM Output.

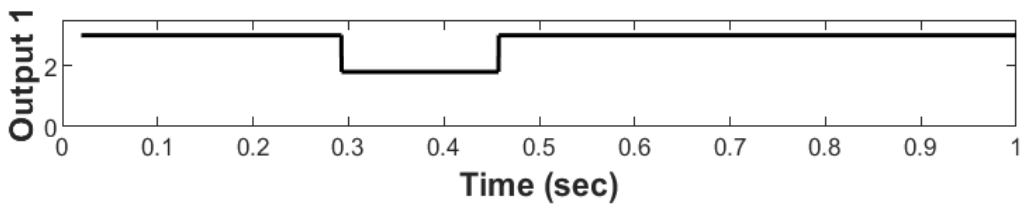
a)



b)



c)



d)

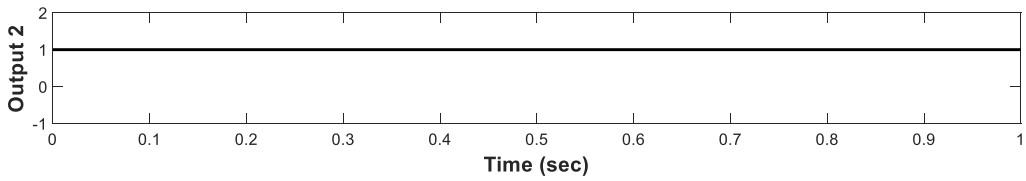


Fig 4.18 Swell with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.

4.13 Interruption with harmonics

CNN-SVM: The voltage interruption with harmonics waveform and its CNN-SVM output are shown in Fig 4.18 (a) and (b), respectively. During the time of the interruption, the CNN-SVM outputs a 0.5, and 3 outside the range of the disturbance, showing that the waveform is distorted.

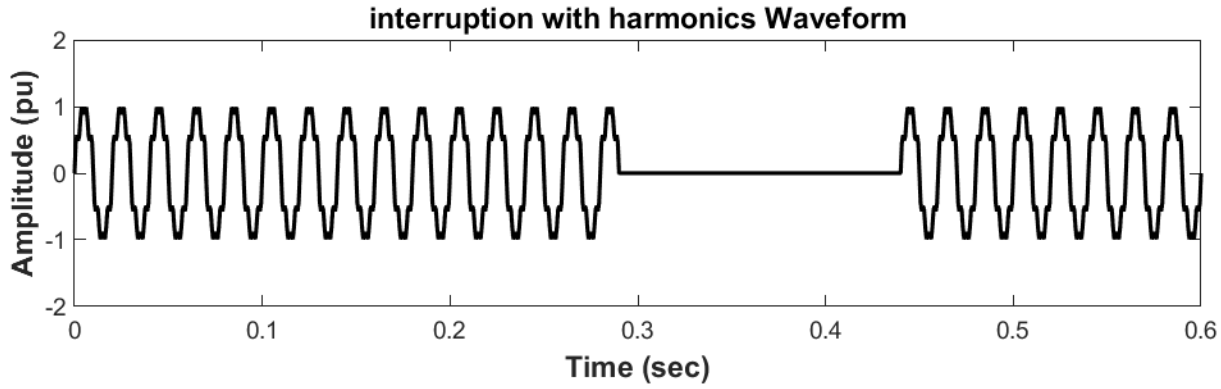
LSTM: The voltage interruption with harmonics waveform and its amplitude are shown in Fig 4.19 (a) and (b), respectively. The two LSTM outputs are shown in Fig 4.19 (c) and (d). The LSTM output 1 equals zero, which means that there is an interruption in the waveform. The LSTM output 2 equals one, meaning that this swell waveform is distorted.

The magnitude, start and end time, duration, and accuracy for the interruption with harmonics waveform of each model are shown in Table 4.10.

Table 4.10 magnitude, start and end time, duration, and accuracy for the interruption with harmonics waveform.

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
CNN-SVM	0	0.2969	0.4431	0.1463	97.68%
LSTM	0.0254	0.3065	0.4434	0.1369	94.62%

a)



b)

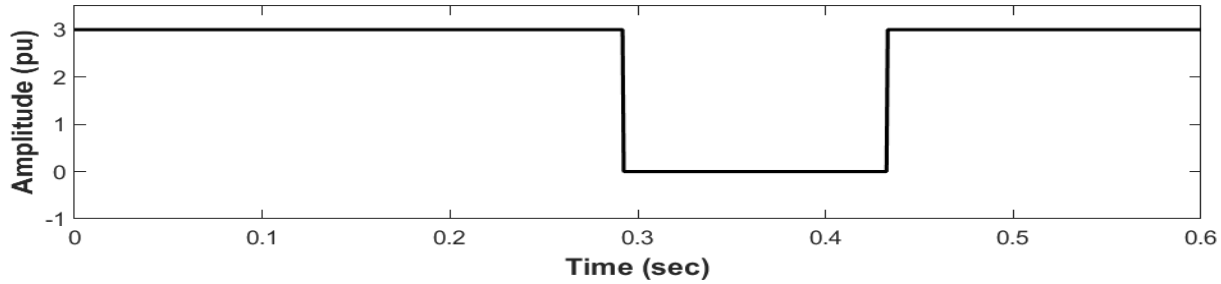


Fig 4.19 Interruption with Harmonics: (a) Waveform and (b) CNN-SVM Output.

a)

b)

c)

d)

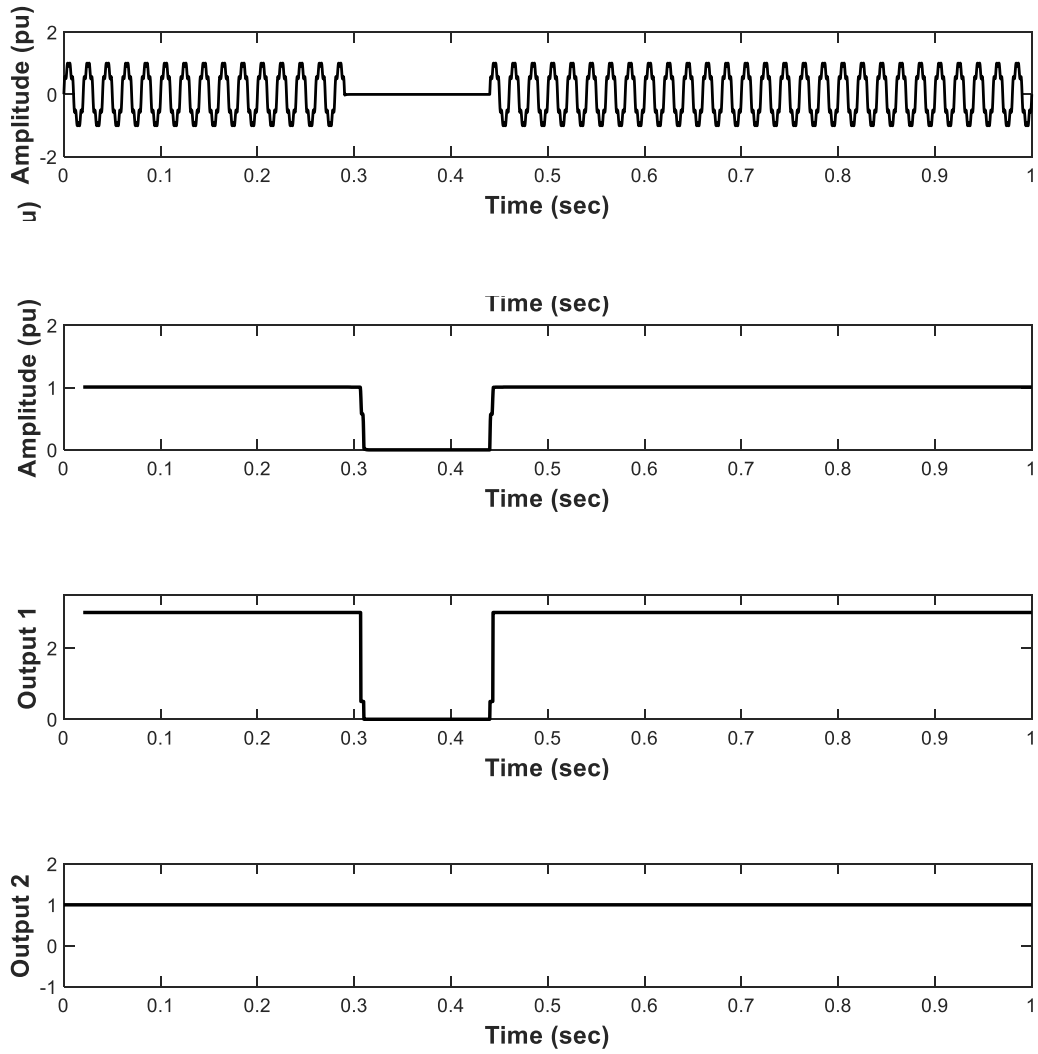


Fig 4.20 Interruption with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.

4.14 All PQ events without harmonics

CNN-SVM: Fig 4.20 (a) and (b) shows all the power quality events (normal voltage, swell, interruption, and sag) and their CNN-SVM outputs, respectively. The model classifies the disturbances according to their given label.

LSTM: The waveform, amplitude, and LSTM outputs 1 and 2 of voltage of all PQ events without harmonics waveform are shown in Fig 4.21 (a), (b), (c), and (d), respectively.

The LSTM output 1 equals the amplitude of each event in its interval (i.e LSTM output 1 equals 0.5 in the interval of the voltage sag and 1.5 for the voltage swell). The LSTM output 2 equals zero, which means that this waveform is not distorted.

Table 4.11 shows the set values for the start, end time, and duration of each disturbance in both models. The predicted magnitude, start and end time, duration, and accuracy for all the PQ events without harmonics waveform of each model are shown in Table 4.12.

Table 4.11 The Set Values for Each Disturbance in Both Models.

	Magnitude	**Start Time**	**End Time**	**Duration (s)**
Swell	1.8	0.05	0.2	0.15
Interruption	0	0.25	0.35	0.1
Sag	0.5	0.4	0.55	0.15

Table 4.12 magnitude, start and end time, duration, and accuracy for all the PQ events without harmonics waveform: a) CNN-SVM model and b) LSTM model.

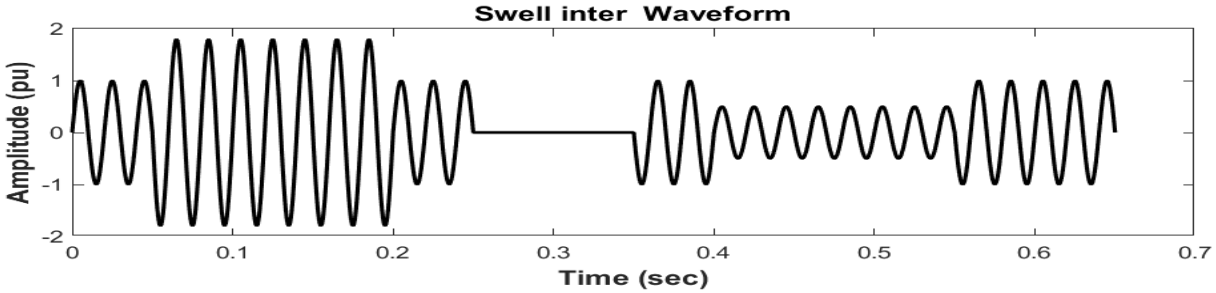
a) CNN-SVM model

	Magnitude (pu)	**Start Time**	**End Time**	**Duration (s)**	**Accuracy**
Swell	1.8	0.0519	0.2075	0.1557	97.68%
Interruption	0	0.2550	0.3543	0.0993	
Sag	0.5	0.4056	0.5562	0.1506	

b) LSTM model

	Magnitude (pu)	**Start Time**	**End Time**	**Duration (s)**	**Accuracy**
Swell	1.7944	0.0521	0.2178	0.1657	96.17%
Interruption	0.0003	0.2592	0.3467	0.0875	
Sag	0.5011	0.4162	0.5537	0.1375	

a)



b)

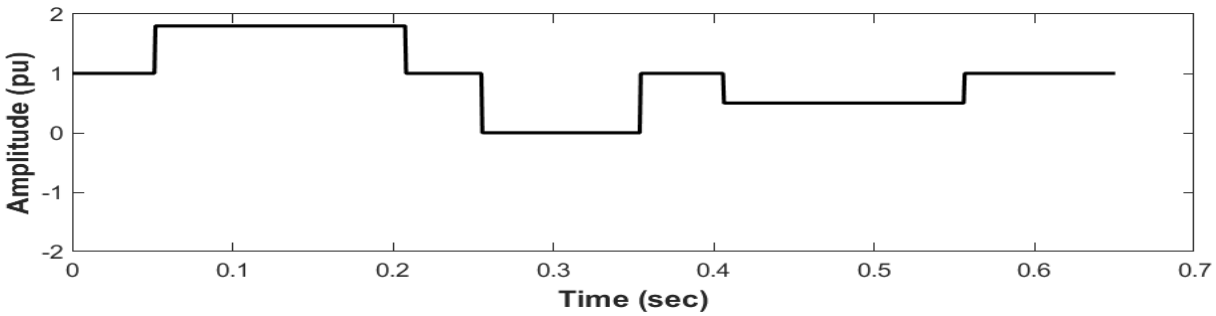
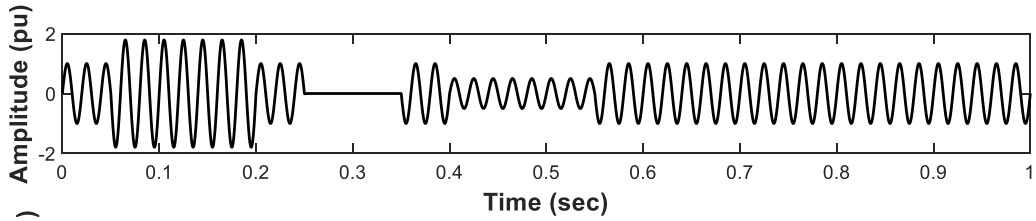
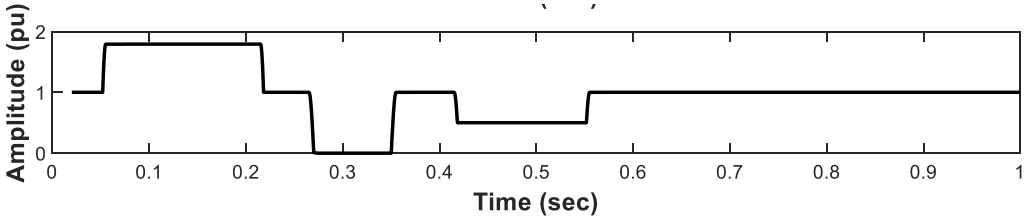


Fig 4.21 All PQ Events without Harmonics: (a) Waveform and (b) CNN-SVM Output.

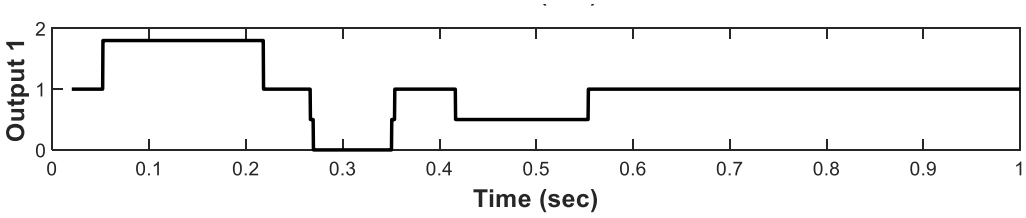
a)



b)



c)



d)

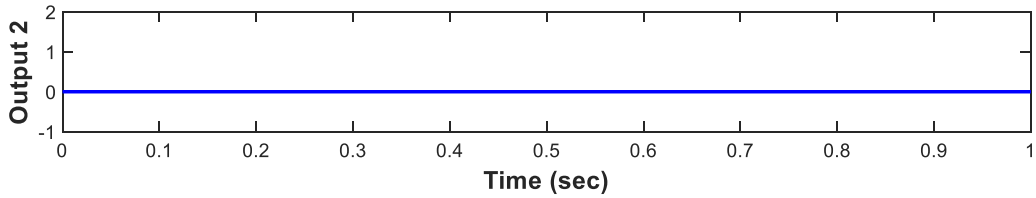


Fig 4.22 All PQ Events without Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.

4.15 All PQ events with harmonics

CNN-SVM: All the power quality events and their CNN-SVM output are shown in Fig 4.22 (a) and (b), respectively. The model classifies each disturbance according to its given label.

LSTM: Fig 4.23(a), (b), (c), and (d) illustrates the waveform, amplitude, and LSTM outputs 1 and 2 of voltage for all PQ events with harmonics waveforms in the same waveform, respectively. The LSTM output 1 equals the amplitude of each event in its interval. The LSTM output 2 equals 1, which means that this waveform is distorted.

Table 4.13 shows the set values for the start, end time, and duration of each disturbance in both models. The predicted magnitude, start and end time, duration, and accuracy for all the PQ events without harmonics waveform of each model are shown in Table 4.14.

Table 4.13 The Set Values for Each Disturbance in Both Models.

	Magnitude	Start Time	End Time	Duration (s)
Swell	1.8	0.05	0.2	0.15
Interruption	0	0.25	0.35	0.1
Sag	0.5	0.4	0.55	0.15

Table 4.14 magnitude, start and end time, duration, and accuracy for all the PQ events with harmonics waveform: a) CNN-SVM model and b) LSTM model.

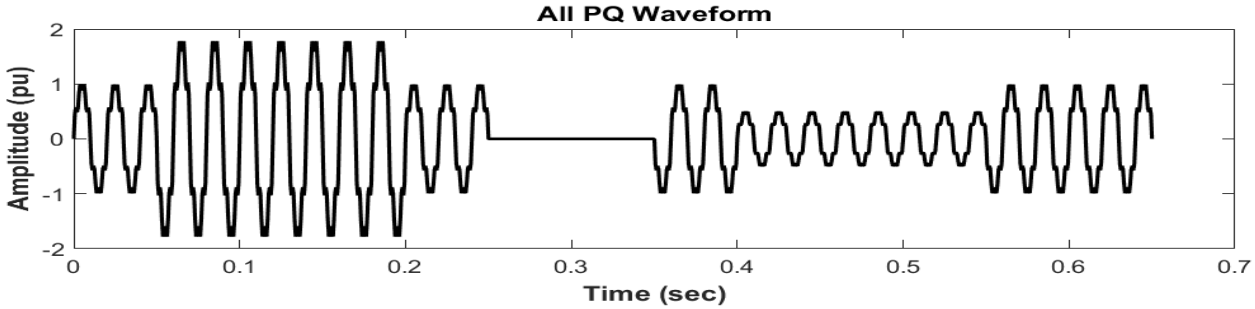
a) CNN-SVM model

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
Swell	1.8	0.0531	0.2069	0.1538	
Interruption	0	0.2531	0.3538	0.1006	96.98%
Sag	0.5	0.4081	0.5556	0.1475	

b) LSTM model

	Magnitude (pu)	Start Time	End Time	Duration (s)	Accuracy
Swell	1.8100	0.0528	0.2175	0.1646	
Interruption	0.0504	0.2600	0.3534	0.0930	95.63%
Sag	0.5043	0.4165	0.5565	0.1400	

a)



b)

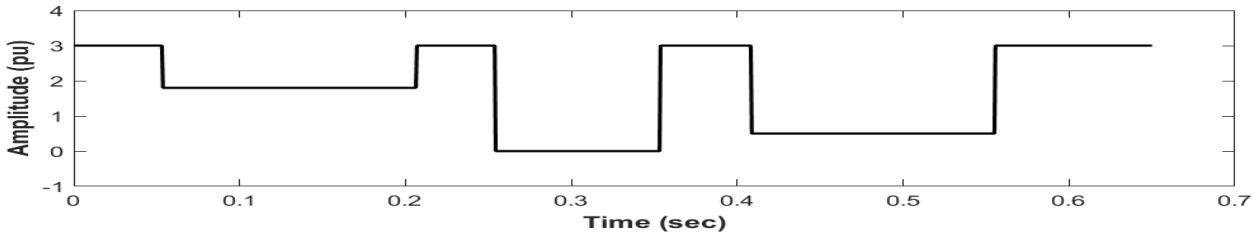


Fig 4.23 All PQ Events with Harmonics: (a) Waveform and (b) CNN-SVM Output.

a)

b)

c)

d)

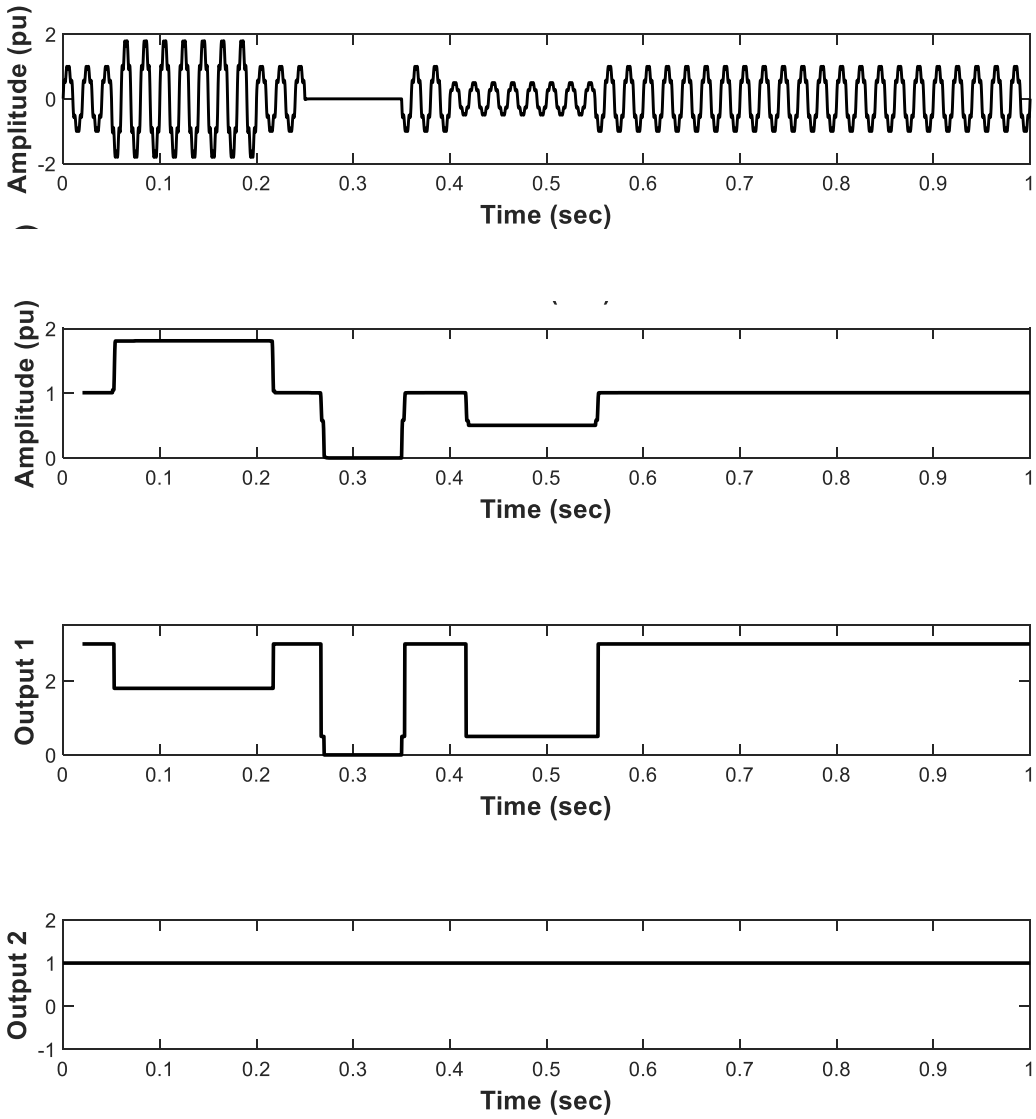
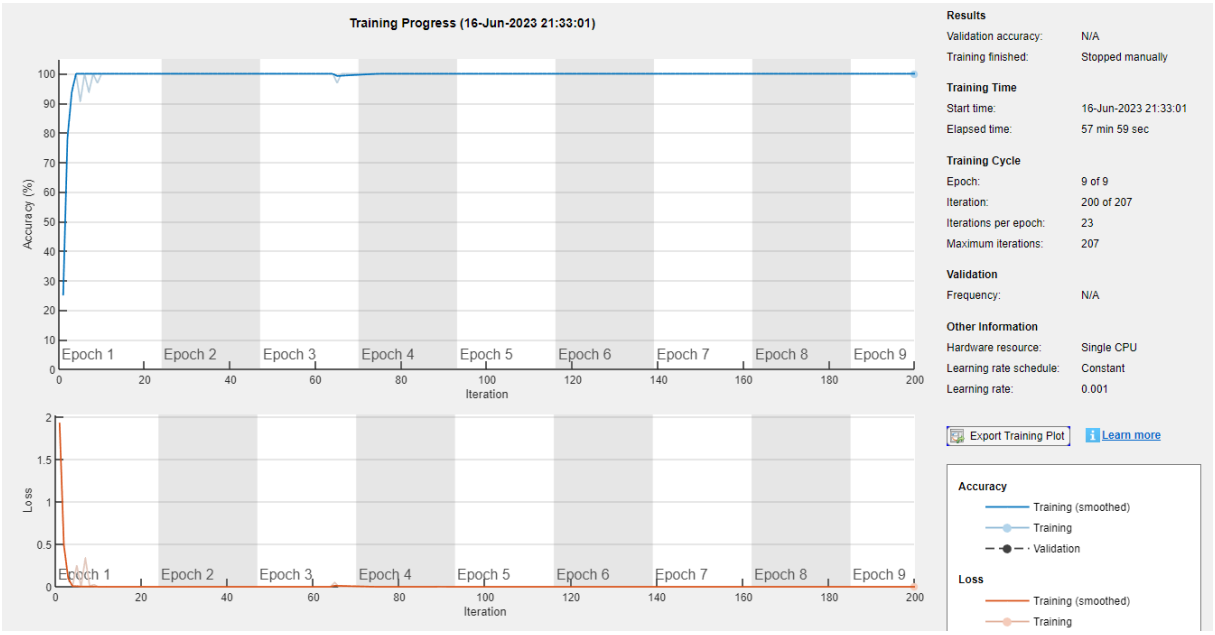


Fig 4.24 All PQ Events with Harmonics: (a) Waveform, (b) Amplitude, (b) LSTM output 1, and (d) LSTM output2.

4.16. CNN-SVM and LSTM Training

The training status and effect of the CNN-SVM and LSTM networks is shown in **Fig 4.25** (a) and (b).

a)



b)

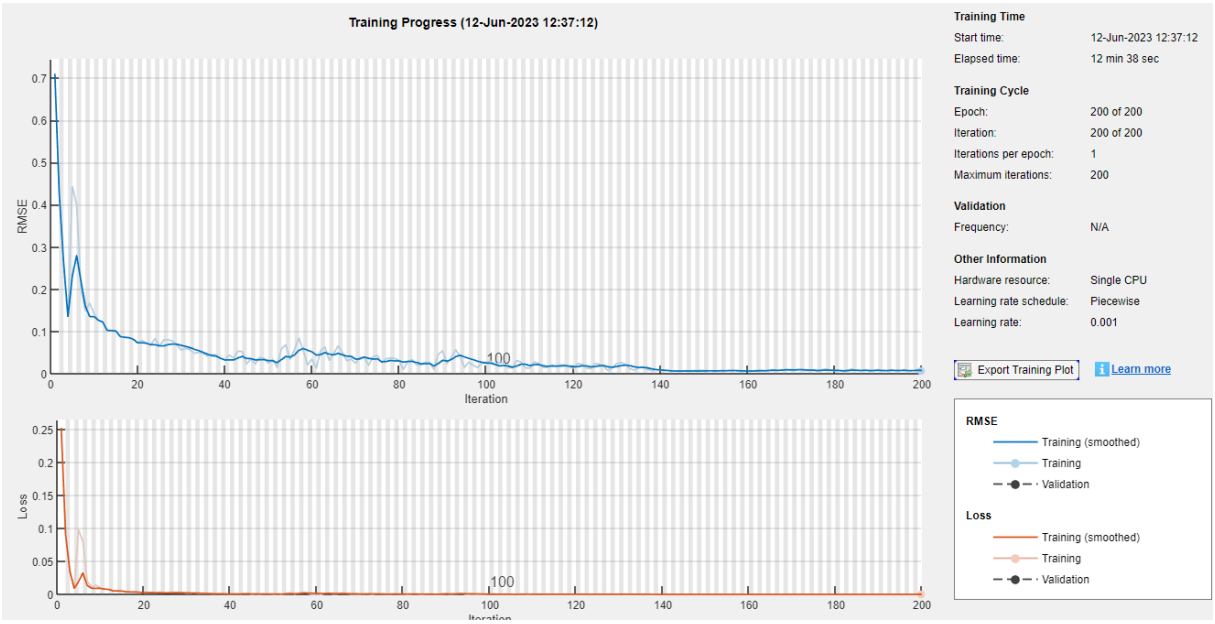


Fig 4.25 CNN-SVM and LSTM Training: a) CNN-SVM Training b) LSTM Training.

CNN-SVM: In order to evaluate the training effect of the network, the cross-entropy loss function and the classification accuracy rate are plotted with the number of iterations (each epoch represents 25 iterations. The classification accuracy curve gradually increases as the number of iterations increases and finally converges to a higher classification accuracy that is close to1.

LSTM: Regression is done by creating a layer array containing a sequence input layer, an LSTM layer, a fully connected layer, and a regression output layer. The size of the sequence input layer is set according to the number of features of the input data. The size of the fully connected layer to the number of responses. The cross-entropy loss function and the Root Mean Squared Error are plotted with the number of iterations (each epoch represents one iteration). As the number of iterations increases, the RMSE curve gradually converges. This decreasing trend indicates that the network is learning and reducing the error between predicted and actual values.

4.17 Discussion

To evaluate the performance of the hybrid CNN-SVM method, a comparison was made with the LSTM model. The hybrid and LSTM classification models were compared by analyzing the accuracy of the two algorithms under different conditions. From Tables 4 - 14, it can be seen that the CNN-SVM model has a better capability of detecting the start time, end time, and duration than the LSTM model; which has a small-time delay. Fig 4.25 shows that both models have excellent classification accuracies of nearly 100%.

Additionally, a comparison of elapsed training time between the two models revealed that the CNN-SVM model requires a relatively longer duration compared to the LSTM model. This can be attributed to the CNN-SVM model's intricate structure and large number of neurons, resulting in higher computational complexity when compared to the LSTM model. From the comprehensive analysis results, it can be seen that the time detection accuracy of the LSTM method (95.63%) is slightly lower than that of CNN-SVM (96.98%), proving that the CNN-SVM method is superior.

4.18 Conclusion

In this chapter, various types of disturbances were simulated to test and evaluate the performance of the CNN-SVM and LSTM detection and classification methods. The two techniques were compared, and results were derived from the simulations, indicating that the CNN-SVM technique exhibits higher accuracy compared to the LSTM technique. The training time was also compared, revealing that the CNN-SVM takes more time than LSTM due to its complex architecture.

General Conclusion

This work investigates the CNN-SVM and LSTM architectures for the detection and classification of power quality disturbances. The primary objective was to understand the functioning of these algorithms in order to identify distinctive features for each disturbance type and achieve precise classification. MATLAB was employed to generate diverse disturbances, including sag, swell, interruption, and harmonics.

The hybrid CNN-SVM model highlighted its potential by capitalizing on the strengths of Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs). This combination enabled efficient feature extraction through the CNN layers, followed by accurate classification using the SVM classifier. The CNN-SVM model proved to be a robust framework for power quality analysis. The LSTM architecture, on the other hand, demonstrated its unique advantages in handling time-series data with temporal dependencies. Its ability to capture long-term memory through hidden layers, serving as memory units, makes it well suited for power quality analysis. Despite the slightly lower time detection accuracy compared to CNN-SVM, the LSTM model remains a valuable tool in effectively analyzing and characterizing power quality disturbances.

Overall, this project shed light on the effectiveness and capabilities of CNN-SVM and LSTM architectures for PQDs detection and classification. The findings provided valuable insights into the distinctive features of different disturbances and contributed to enhancing the accuracy of power quality analysis. The results also highlight the importance of selecting the appropriate architecture based on the specific requirements and characteristics of the analysis task.

Future Work

Based on the project results, future work can focus on the following points:

- Identifying the source of power quality disturbances within the signal. This could involve developing advanced algorithms and techniques to pinpoint the specific devices or systems responsible for the observed disturbances. By accurately identifying the source, it would be possible to implement targeted mitigation strategies and improve the overall power quality.
- Obtaining real power quality data from diverse sources, such as utility companies or power system operators, by collecting and utilizing real data instead of generated data, this step would enhance the models' ability to handle the intricacies and variations present in actual power systems, leading to improved detection and classification accuracy.
- Integrating the developed algorithms directly into smart meters would enable real-time detection and classification of power quality disturbances. Evaluating the performance and feasibility of these algorithms on smart meters would contribute to the development of intelligent power quality monitoring and mitigation systems within smart grid environments.

References

[1] S. Chattopadhyay, M. Mitra, and S. Sengupta, *Electric Power Quality*. in Power Systems. Dordrecht: Springer Netherlands, 2011. doi: 10.1007/978-94-007-0635-4.

[2] D. Thomas, "Conducted emissions in distribution systems (1 kHz–1 MHz)," *IEEE Electromagn. Compat. Mag.*, vol. 2, no. 2, pp. 101–104, 2013,doi:10.1109/MEMC.2013.6550941.

[3] "Power Quality: Problems and Mitigation Techniques: Singh, Bhim, Chandra, Ambrish, Al-Haddad, Kamal: 9781118922057: Amazon.com: Books." https://www.amazon.com/Power-Quality-Problems-Mitigation-Techniques/dp/1118922050 (accessed Jun. 07, 2023).

[4] "Transients and Its Classification | Power System | Electricity." https://www.engineeringenotes.com/electrical-engineering/transients/transients-and-its-classification-power-system-electricity/32361 (accessed Jun. 07, 2023).

[5] "Electrical Power Systems Quality: Surya Santoso, H. Wayne Beaty, Roger C. Dugan, Mark F. McGranaghan: 9780071386227: Amazon.com: Books." https://www.amazon.com/Electrical-Power-Systems-Quality-Santoso/dp/007138622X (accessed Jun. 07, 2023).

[6] R. K. Jena, "Electrical Power Quality," *Dep. Electr. Eng. CET BBSR*, vol. 66.)

[7] "Electrical power equipment maintenance and testing." https://www.industrial-electronics.com/epemt_12a.html (accessed Jun. 07, 2023).

[8] "Multiple Chaos Synchronization System for Power Quality Classification in a Power System - PMC." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3934410/ (accessed Jun. 07, 2023).

[9] "(PDF) Performance Comparison of Voltage Sag/Swell Detection Methods Implemented in Custom Power Devices." https://www.researchgate.net/publication/318316470_Performance_Comparison_of_Voltage_SagSwell_Detection_Methods_Implemented_in_Custom_Power_Devices (accessed Jun. 07, 2023)

[10] "Voltage: Variation, Imbalance and Fluctuations | Electrical Engineering." https://www.engineeringenotes.com/electrical-engineering/voltage/voltage-variation-imbalance-and-fluctuations-electrical-engineering/32357 (accessed Jun. 07, 2023).

[11] “Voltage Imbalance (Unbalance) Study | Omazaki Engineering,” *Omazaki Group*, Apr. 05, 2021. https://www.omazaki.co.id/en/voltage-imbalance-unbalance-study/ (accessed Jun. 07, 2023).

[12] “Electrical Bus: DC offset.” http://electric-bus.blogspot.com/2009/02/dc-offset.html (accessed Jun. 07, 2023).

[13] “What Is Harmonic? Definition From WhatIs.” https://www.techtarget.com/whatis/definition/harmonic (accessed Jun. 07, 2023).

[14] M. B. Marz, “Interharmonics: What They Are, Where They Come From and What They Do”.

[15] “Voltage Notching – Voltage Disturbance.” https://voltage-disturbance.com/voltage-quality/voltage-notching/ (accessed Jun. 07, 2023).

[16] “What Is Electrical Noise?” https://www.riello-ups.com/questions/45-what-is-electrical-noise (accessed Jun. 07, 2023).

[17] “Flicker in electrical systems - DEWETRON.” https://www.dewetron.com/2022/03/flicker-in-electrical-systems/ (accessed Jun. 07, 2023).

[18] “Voltage Fluctuation - an overview | ScienceDirect Topics.” https://www.sciencedirect.com/topics/engineering/voltage-fluctuation (accessed Jun. 07, 2023).

[19] O. Mahela, A. 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, Jan. 2015, doi: 10.1016/j.rser.2014.08.070.

[20] “Continuous and Discrete Wavelet Transforms - MATLAB & Simulink.” https://www.mathworks.com/help/wavelet/gs/continuous-and-discrete-wavelet-transforms.html (accessed Jun. 07, 2023)

[21] “The fast continuous wavelet transformation (fCWT) for real-time, high-quality, noise-resistant time–frequency analysis | Nature Computational Science.” https://www.nature.com/articles/s43588-021-00183-z (accessed Jun. 07, 2023).

[22] C. F. Drummond and D. Sutanto, “Classification of Power Quality disturbances using the iterative Hilbert Huang Transform,” in *Proceedings of 14th International Conference on*

Harmonics and Quality of Power - ICHQP 2010, Bergamo, Italy: IEEE, Sep. 2010, pp. 1–7. doi: 10.1109/ICHQP.2010.5625326.

[23]"Hilbert-Huang Transform - an overview | ScienceDirect Topics." https://www.sciencedirect.com/topics/engineering/hilbert-huang-transform?fbclid=IwAR0KdWoOa7ZFbq00KrtIOQDKFDG8mNtYdivA2LIqi-1Q2hINGtVskWy5cMc (accessed Jun. 07, 2023).

[24]"Improved Hilbert-Huang Transform Applied in Power Quality Detection in Ship Power System | Scientific.Net." https://www.scientific.net/AMM.88-89.365?fbclid=IwAR0FLn9A8Y1Z7iYNWnhgdt_YGXEVmsGt-nz4WtEyIjIJaide4a2LSJ6dCY4 (accessed Jun. 07, 2023).

[25] "Frontiers | Classification of Power Quality Disturbance Based on S-Transform and Convolution Neural Network." https://www.frontiersin.org/articles/10.3389/fenrg.2021.708131/full?fbclid=IwAR1fb6jBfOAr-BNb5V33d-Nlc5xoZ9NscBRiCCnhsI0liruRxiNJF74fUu4 (accessed Jun. 07, 2023).

[26] M. Dehghani, "Comparison of S-transform and Wavelet Transform in Power Quality Analysis," *World Academy of Science, Engineering and Technology*, vol. 38, Feb. 2009.

[27] C. Beuter and M. Oleskovicz, "S Transform:from main concepts to some power quality applications," *IET Signal Processing*, vol. 14, May 2020, doi: 10.1049/iet-spr.2019.0042.

[28] "Introduction to AI: A Modern Approach." http://people.eecs.berkeley.edu/~russell/intro.html (accessed Jun. 07, 2023).

[29] E. Yiğit, U. Özkaya, Ş. Öztürk, D. Singh, and H. Gritli, "Automatic Detection of Power Quality Disturbance Using Convolutional Neural Network Structure with Gated Recurrent Unit," *Mobile Information Systems*, vol. 2021, p. e7917500, Aug. 2021, doi: 10.1155/2021/7917500

[30] "Machine Learning and Better Decision Making.," *Rewise*, Feb. 10, 2023. https://www.rewiseglobal.com/opinion-mining-through-text-analysis/ (accessed Jun. 07, 2023).

[31] "What is Supervised Learning? | Machine Learning Tasks [Updated 2023] | SuperAnnotate." https://www.superannotate.com/blog/supervised-learning-and-other-machine-learning-tasks (accessed Jun. 07, 2023).

[32] “The most insightful stories about Classification Models - Medium.” https://medium.com/tag/classification-models (accessed Jun. 07, 2023).

[33] “Building Smarter AI: A Comprehensive Overview of Data Mining Algorithms and Techniques.” https://ts2.space/en/building-smarter-ai-a-comprehensive-overview-of-data-mining-algorithms-and-techniques/ (accessed Jun. 07, 2023).

[34] “Classification vs Regression in Machine Learning,” *GeeksforGeeks*, Jan. 08, 2019. https://www.geeksforgeeks.org/ml-classification-vs-regression/ (accessed Jun. 07, 2023).

[35] “Unsupervised learning,” *Wikipedia*. Jun. 02, 2023. Accessed: Jun. 07, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Unsupervised_learning&oldid=1158188297

[36] “Semi-Supervised Learning: Techniques & Examples [2023].” https://www.v7labs.com/blog/semi-supervised-learning-guide, https://www.v7labs.com/blog/semi-supervised-learning-guide (accessed Jun. 07, 2023).

[37] “Reinforcement learning,” *Wikipedia*. May 18, 2023. Accessed: Jun. 07, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Reinforcement_learning&oldid=1155454334

[38] “Reinforcement learning - GeeksforGeeks.” https://www.geeksforgeeks.org/what-is-reinforcement-learning/ (accessed Jun. 07, 2023)

[39] Cui Y, Li A Liu Z, "A classification method for complex power quality disturbances using EEMD and rank wavelet SVM," *IEEE Transations on Smart Grid*, vol. 6, no. 4, pp. (pp. 1678-85), 2015 Jul.

[40] Romero-Troncoso RJ, Osornio-Rios RA, Garcia-Perez A, Cabal-Yepez E Granados-Lieberman D, "Techniques and methodologies for power quality analysis and disturbances classification in power systems: a review," *IET Generation Transmission & Distribution*, vol. 5, no. 4, 2011 Apr 1.

[41] “K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint,” *www.javatpoint.com*. https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning (accessed Jun. 07, 2023).

[42] “What is a Decision Tree | IBM.” https://www.ibm.com/topics/decision-trees (accessed Jun. 07, 2023).

[43] “Decision Tree - GeeksforGeeks.” https://www.geeksforgeeks.org/decision-tree/ (accessed Jun. 07, 2023).

[44] “Machine Learning Random Forest Algorithm - Javatpoint.” https://www.javatpoint.com/machine-learning-random-forest-algorithm (accessed Jun. 07, 2023).

[45] “Random forest - Wikipedia.” https://en.wikipedia.org/wiki/Random_forest (accessed Jun. 07, 2023).

[46] “Random Forest Classifier Tutorial: How to Use Tree-Based Algorithms for Machine Learning.” https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/ (accessed Jun. 07, 2023).

[47] “Support Vector Machine (SVM) Python Example - Data Analytics.” https://vitalflux.com/classification-model-svm-classifier-python-example/

[48] R. A. de Oliveira and M. H. J. Bollen, “Deep learning for power quality,” *Electric Power Systems Research*, vol. 214, p. 108887, Jan. 2023, doi: 10.1016/j.epsr.2022.108887.

[49] “What are Neural Networks? | IBM.” https://www.ibm.com/topics/neural-networks (accessed Jun. 07, 2023).

[50] “What are Recurrent Neural Networks? | IBM.” https://www.ibm.com/topics/recurrent-neural-networks?fbclid=IwAR10B8ce18uNd9YQQp4DFKSiBuyDx-7Jnwth9ejk74b-Eyc9RxbzFWO-7Kk (accessed Jun. 07, 2023).

[51] P. Binsha, S. Kumar S, S. Athira, and S. Kp, “Power quality signal classification using convolutional neural network,” vol. 9, pp. 6405–6414, Jan. 2016.

[52] J. Brownlee, “A Gentle Introduction to Long Short-Term Memory Networks by the Experts,” *MachineLearningMastery.com*, May 23, 2017. https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/ (accessed Jun. 07, 2023).

[53] “What is LSTM - Introduction to Long Short Term Memory.” https://intellipaat.com/blog/what-is-lstm/?US (accessed Jun. 08, 2023).

[54] “Long short-term memory - Wikipedia.” https://en.wikipedia.org/wiki/Long_short-term_memory (accessed Jun. 08, 2023).

[55] “Long Short Term Memory Networks Explanation - GeeksforGeeks.” https://www.geeksforgeeks.org/long-short-term-memory-networks-explanation/ (accessed Jun. 08, 2023).

[56] “(PDF) An efficient implementation of 2D convolution in CNN.” https://www.researchgate.net/publication/312438814_An_efficient_implementation_of_2D_convolution_in_CNN (accessed Jun. 08, 2023).

[57] C.-L. Fan and Y.-J. Chung, “Design and Optimization of CNN Architecture to Identify the Types of Damage Imagery,” *Mathematics*, vol. 10, no. 19, Art. no. 19, Jan. 2022, doi: 10.3390/math10193483.

[58] “Multi-Label Classification for Power Quality Disturbances by Integrated Deep Learning | IEEE Journals & Magazine | IEEE Xplore.” https://ieeexplore.ieee.org/document/9597531 (accessed Jun. 08, 2023).

[59] A. Agarwalla, D. Dileep, P. Jyothsana, P. Unnikrishnan, and K. Thirumala, “Principal Component Analysis and CNN for Classification of Power Quality Disturbances,” in *2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Nov. 2021, pp. 1–5. doi: 10.1109/UPCON52273.2021.9667629.

[60] S. Pokhrel, “Beginners Guide to Understanding Convolutional Neural Networks,” *Medium*, Sep. 20, 2019. https://towardsdatascience.com/beginners-guide-to-understanding-convolutional-neural-networks-ae9ed58bb17d (accessed Jun. 08, 2023).

[61] “Convolutional Neural Networks (CNN): Step 3 - Flattening - Blogs - SuperDataScience | Machine Learning | AI | Data Science Career | Analytics | Success.” https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-3-flattening (accessed Jun. 08, 2023).

[62] “Fully Connected Layer vs Convolutional Layer: Explained | Built In.” https://builtin.com/machine-learning/fully-connected-layer (accessed Jun. 08, 2023).

[63] “CNN and Softmax - Andrea Perlato.” https://www.andreaperlato.com/aipost/cnn-and-softmax/ (accessed Jun. 08, 2023).

[64] “Softmax Activation Function with Python - MachineLearningMastery.com.” https://machinelearningmastery.com/softmax-activation-function-with-python/ (accessed Jun. 08, 2023).

[65] “An Introduction to Neural Network Loss Functions - Programmathically.” https://programmathically.com/an-introduction-to-neural-network-loss-functions/ (accessed Jun. 08, 2023).

[66] “Cross Entropy Loss: Intro, Applications, Code.” https://www.v7labs.com/blog/cross-entropy-loss-guide (accessed Jun. 08, 2023).

[67] “Softmax and Cross Entropy Loss.” https://www.parasdahal.com/softmax-crossentropy (accessed Jun. 08, 2023)

[68] “Root Mean Square Error (RMSE),” *C3 AI*. https://c3.ai/glossary/data-science/root-mean-square-error-rmse/ (accessed Jun. 16, 2023).

[69] “Root Mean Squared Error (RMSE) | SAP Help Portal.” https://help.sap.com/docs/SAP_PREDICTIVE_ANALYTICS/41d1a6d4e7574e32b815f1cc87c00f42/5e5198fd4afe4ae5b48fefe0d3161810.html (accessed Jun. 16, 2023).

[70] “5 Popular CNN Architectures Clearly Explained and Visualized | by Albers Uzila | Towards Data Science.” https://towardsdatascience.com/5-most-well-known-cnn-architectures-visualized-af76f1f0065e#8f20 (accessed Jun. 08, 2023).

[71] “Introduction to Support Vector Machines (SVM) - GeeksforGeeks.” https://www.geeksforgeeks.org/introduction-to-support-vector-machines-svm/ (accessed Jun. 08, 2023).

[72] “Kernel Functions. Lately, I have been doing some reading… | by Tejumade Afonja | Towards Data Science.” https://towardsdatascience.com/kernel-function-6f1d2be6091 (accessed Jun. 08, 2023).

[73] “A New Design Based-SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition - ScienceDirect.” https://www.sciencedirect.com/science/article/pii/S1877050916309991 (accessed Jun. 08, 2023).

[74] “LSTM | Introduction to LSTM | Long Short Term Memory Algorithms.” https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-

lstm/?fbclid=IwAR3lo-xgdAFb__X4d5f5zGnPXykvl564sQAp-52m7rPGwEWYPJ_AP59ZCUk (accessed Jun. 08, 2023).

[75] “How LSTM networks solve the problem of vanishing gradients | by Nir Arbel | DataDrivenInvestor.” https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577 (accessed Jun. 08, 2023).

[76] “Deep Learning | Introduction to Long Short Term Memory - GeeksforGeeks.” https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/ (accessed Jun. 08, 2023)

[77] “Understanding LSTM Networks -- colah’s blog.” https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed Jun. 08, 2023).

[78] “Long Short-Term Memory Networks (LSTM)- simply explained! | Data Basecamp.” https://databasecamp.de/en/ml/lstms (accessed Jun. 08, 2023).

[79] “10.1. Long Short-Term Memory (LSTM) — Dive into Deep Learning 1.0.0-beta0 documentation.” https://d2l.ai/chapter_recurrent-modern/lstm.html (accessed Jun. 08, 2023).

[80] “Long short-term memory (LSTM) cell architecture. Input gate i(t): It... | Download Scientific Diagram.” https://www.researchgate.net/figure/Long-short-term-memory-LSTM-cell-architecture-Input-gate-it-It-controls-what_fig3_336256820 (accessed Jun. 08, 2023).

[81] “Introduction to LSTM Units in RNN | Pluralsight.” https://www.pluralsight.com/guides/introduction-to-lstm-units-in-rnn (accessed Jun. 08, 2023).

[82] “Engineering Proceedings | Free Full-Text | Detection and Classification of Power Quality Disturbances Using LSTM.” https://www.mdpi.com/2673-4591/29/1/2 (accessed Jun. 08, 2023).

[83] Q. Wang, X. Liang, and S. Qin, “Research on power quality disturbance analysis and identification based on LSTM,” *Energy Reports*, vol. 8, pp. 709–718, Nov. 2022, doi: 10.1016/j.egyr.2022.09.170.

[84] A. Elsayed, “What Precision, Recall, F1 Score and Accuracy Can Tell You ?,” *Analytics Vidhya*, Feb. 28, 2020. https://medium.com/analytics-vidhya/what-precision-recall-f1-score-and-accuracy-can-tell-you-fe1eab1ada5a (accessed Jun. 08, 2023).

[85] A. Ambika, “Decluttering the performance measures of classification models,” *Analytics Vidhya*, Dec. 16, 2020. https://www.analyticsvidhya.com/blog/2020/12/decluttering-the-performance-measures-of-classification-models/ (accessed Jun. 08, 2023).

[86] “Basic evaluation measures from the confusion matrix,” *Classifier evaluation with imbalanced datasets*, Jun. 03, 2015. https://classeval.wordpress.com/introduction/basic-evaluation-measures/ (accessed Jun. 08, 2023).

[87] “Detection and Classification of Power Quality Disturbances Using Wavelet Transform and Support Vector Machines: Electric Power Components and Systems: Vol 38, No 2.” https://www.tandfonline.com/doi/abs/10.1080/15325000903273387 (accessed Jun. 08, 2023).