International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I4P112

Volume 72 Issue 4, 121-130, April 2024

Original Article

Hybrid Deep Networks for Early Detection of Power Quality Disturbances in Smart Grids: A Resilience Enhancement Approach

Sabeena Beevi K[1*], Neethu Mohan[2], Thasneem A[1], Krishnendhu Murali[1],Visakh S.[1]

[1]*Department of Electrical and Electronics Engineering, T K M College of Engineering, Kerala, India.*
[2]*Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Kerala, India.*

[1]*Corresponding Author : sabeena3000@tkmce.ac.in*

Received: 25 January 2024 Revised: 03 April 2024 Accepted: 11 April 2024 Published: 24 April 2024

Abstract *- A smart grid is an electrical power system that uses modern digital technologies and automation to improve reliability, efficiency, and sustainability. However, the integration of these technologies can increase the risk of power quality (PQ) disturbances, which can damage electrical devices and cause significant economic losses. Conventional protection schemes in smart grids typically provide a reactive approach to detecting PQ disturbances. This is not sufficient to address the root cause of these distortions, and thus, advanced protection schemes that incorporate predictive measures are needed. Hence it will be essential to proactively detect the occurrence of power quality disturbances and implement preventive measures to mitigate their impact. Traditional forecasting methods often rely on simple models and assumptions, which can lead to inaccuracies and limitations in the predictions. This paper proposes an advanced model for the early detection of PQ disturbances by utilizing the power of artificial intelligence and machine learning. This paper utilizes a state-of-the-art encoder-decoder model for forecasting Power Quality (PQ) disturbances, accompanied by the implementation of a hybrid Convolutional Neural Network-Long Short-Term Memory model to categorize these disorders effectively. By accurately detecting the disturbances in advance, appropriate mitigation measures can be considered to minimize their effect on the system. Several experiments are conducted to find the optimum model with proper network configurations that detect the PQ disorders. The effectiveness of the proposed model is confirmed through testing with over 18 different classes of simple and mixed distortions. The study further explores the potential of a unified model capable of detecting and classifying multiple disturbances based on forecasted data points.*

Keywords *- Power grid, Power quality disturbances, Forecasting, Convolutional Neural Network, Long-Short Term Memory, Early detection.*

1. Introduction

The shift from conventional power grids to contemporary decentralized smart grids, accompanied by smart technologies, brings additional hurdles in ensuring a reliable power supply. Due to the distributed nature and intermittent power generation of renewable energy sources incorporated into the grid poses additional challenges to the electric grid. Moreover, nonlinear loads, load fluctuations, the presence of power electronic devices, and external environmental features contribute to various power quality (PQ) issues [1]. Enhancing the power supply quality offers numerous benefits, including improved energy efficiency, minimized energy wastages, reduced power disruptions, minimized equipment failures, and increased end-user satisfaction. Taking proactive measures to forecast and monitor power quality events will significantly enhance the reliability of the power grid. This will help to reduce the risk of power outages and thus improve the life of end-user equipment. Furthermore, by mitigating the PQ distortions, the stability and safety of the entire system can be highlighted. Therefore, early detection of power quality distortions plays a paramount role in improving the efficacy, stability, security, and sustainability of the electrical grid. Detection of PQ disturbances involves classifying them into distinct categories, which include voltage sags, swells, transients, flickers, interruptions, etc.

This classification is done based on the electrical attributes of PQ distortions and their impact. By categorizing power quality events, it becomes possible to identify their root causes and select suitable mitigation strategies to safeguard sensitive equipment from their effects. The integration of power quality monitoring systems with precise detection methods can contribute to diminishing the frequency and duration of PQ incidents. Statistical analysis techniques like Principal Component Analysis (PCA) are widely used to reduce the dimensionality of complex power quality signals while preserving their essential information [3]. Even though the statistical models improve the detection accuracy, they often work by holding a linear relationship among the data. However, this assumption might not always align with the complexities and non-linear nature of PQ events. Another disadvantage of statistical models is the requirement of a huge amount of data to recognize the underlying patterns of PQ signals effectively.

A mathematical approach designed to address the uncertainty in the data is fuzzy logic [4]. In the context of PQ classification, fuzzy logic can be effectively employed to tackle the intricate and non-linear attributes inherent in power data. However, the inherent complexity of fuzzy models increases the associated computational cost. Support Vector Machine (SVM) based methodologies are extensively employed in the classification of PQ events according to their electrical attributes. This is achieved by mapping the data into a feature space of higher dimensions [5]. The selection of appropriate feature parameters is important to achieve an accurate classification using SVM models. The time and computational complexity of SVM models is high compared with other non-linear models for PQ detection.

Rule-based methods are extensively utilized in PQ detection. These methods leverage the expertise of domain specialists and practical data to establish a predefined set of rules that aid in categorizing different PQ events [6]. Appropriate feature extraction and decision logic are also involved in rule-based PQ classification models. Artificial Neural Networks (ANNs) have emerged as prevailing techniques for pattern recognition and classification in various domains, including PQ. ANNs are capable of learning complex relationships within data and can be used for the accurate categorization of PQ distortions [7].

The accuracy and effectiveness of ANN models heavily depend on the quality of the selected features. Thus the selection of the right feature is a critical step in building a robust and accurate ANN model. An Extreme Learning Machine (ELM) model combined with optimization techniques is presented for PQ detection in [8]. In this hybrid approach, optimization theory is used to enhance the performance of the ELM model. Selecting the appropriate optimization technique and its parameters is a tough task for real-time PQ detection. Deep-learning-based techniques have gained significant attention in the field of PQ analysis [9, 10]. Convolution Neural Networks (CNN) are well-suited for PQ distortion analysis due to their ability to extract relevant features from the data [11]. CNN model combined with wavelet features are exploited for PQ events detection in [12]. The CNN model utilises the extracted wavelet transform coefficients as input to improve classification accuracy.

However, categorizing the PQ disturbance is insufficient for its rectification. To enhance the rectification process, it is crucial to include how the PQ signal may change in the near future. The potential influence of load fluctuations results in the amplification of voltage waveforms and this can be avoided by predicting the power signal amplitude. Forecasting enables the identification of potential power loads that could give rise to specific power quality issues prior to their connection to the electric power grid. This information is important for power system operators to make proactive decisions and take preventive actions to maintain the quality supply of power. By forecasting the potential disturbances, utilities can take preventive measures to mitigate their impact.

This proactive approach can prevent PQ events from occurring in the first place and helps to avoid equipment failures. Another advantage associated with disturbance forecasting is the better allocation of resources. Power system operators can allocate manpower and equipment more efficiently, especially during periods when disturbances are predicted.

Forecasting can aid in the optimal placement of PQ meters and also help to identify the fluctuations in power generation due to the integration of renewable energy sources like solar and wind. Both power quality forecasting and detection are integral to the development and implementation of smart grid technologies. They empower utilities with the information and tools needed to maintain a high level of power quality optimise operations, and overall performance.

Traditional methods often rely on simple models and assumptions, which can lead to inaccuracies and limitations in the predictions and classifications of PQ distortions [13,14]. To address the limitations of traditional methods, advanced techniques based on artificial intelligence and machine learning are being used to improve the accuracy and efficiency of PQ forecasting and detection tasks. Motivated by all the facts mentioned above, this work proposes a hybrid deep-learning-based network model for early detection of PQ disturbances in smart grid scenarios. The present work makes significant contributions in the following areas:

- Developing an encoder-decoder model for forecasting simple and mixed-mode power quality disturbances.
- Creating a hybrid model that incorporates Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to classify PQ disturbances.
- Proposing a unified detection model that enables early detection of disturbances.

The evaluation of the hybrid architectures proposed in this study encompasses both simple and multiple PQ distortions. The paper's structure is outlined as follows: Section II offers an in-depth exploration of the methodology proposed for this project. At the same time, Section III delves into the key outcomes of the study. Finally, in Section IV, the paper concludes by highlighting the future prospects and potential directions for further research.

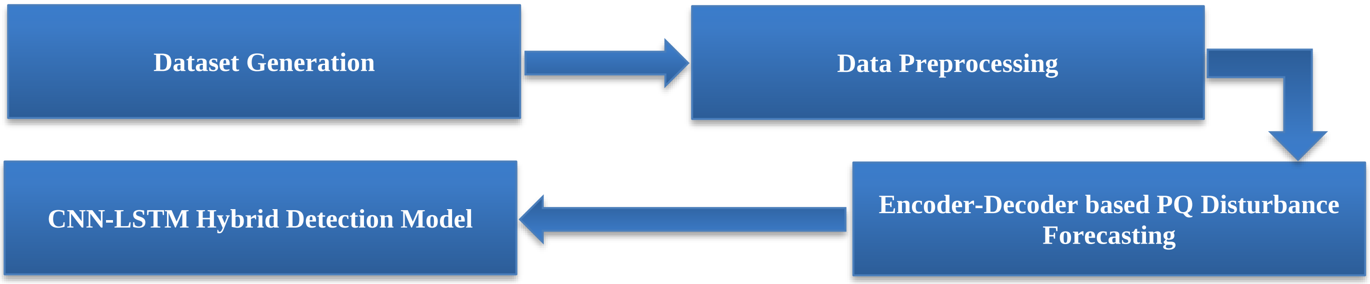

Fig. 1 A visual representation of the proposed methodology

2. Proposed Methodology

The block diagram in Figure 1 provides a high-level overview of the proposed methodology. In the upcoming section, a detailed description of each block will be provided, outlining their individual roles and functionalities within the overall system.

2.1. Dataset Generation

Power Quality (PQ) disturbances refer to irregular fluctuations that occur when the waveform deviates from its expected range. These variations can lead to issues for sensitive equipment and have the potential to impact the ability of the power system to maintain its equilibrium under normal and disturbed conditions. Based on the characteristics, these disturbances can be categorized as linear, nonlinear, or mixed distortions. Linear PQ disturbances [15, 16] affect the magnitude and frequency of the voltage and current waveforms but do not cause significant changes in the waveform shape. Voltage swell, sag, interruptions, etc, are examples of linear disturbances. Non-linear PQ disturbances cause significant changes in the waveform shape. Unlike linear disturbances, non-linear disturbances can cause harmonic distortion, inter-harmonics, and voltage fluctuations that can have a significant impact on sensitive electronic devices and equipment [17, 18]. Non-linear disturbances include power spikes, sub and inter-harmonics, transients, notches, and noises. These interruptions can cause machine failure and coil combustion when they penetrate the electrical system. It is important to properly design and manage power systems to minimize the impact of non-linear PQ disturbances. Mixed-mode PQ disturbance refers to a combination of two or more types of distortions. These disturbances can reduce the overall efficiency of the system. It is important to identify the cause of mixed-mode disturbances and implement mitigation strategies to minimize their impact.

Massive amounts of data are needed for Deep Neural Network training. PQD mathematical modeling provides the sampled values of 9000 power signals. Using parameter equations, 500 samples from each of the 18 power quality disturbances are generated. These disturbances range from linear to non-linear and simple to mixed mode. As per the specified IEEE standards outlined in Table I, the parameter range undergoes variations. For instance, disturbances that cause high-frequency fluctuations in voltage or current can exhibit a frequency interval of 300Hz and 900Hz. To address this, signals are generated using 81 sample points at a frequency of 1 kHz. However, it is worth noting that the maximum frequency supported by the suggested model is 450Hz. In each equation, parameter A describes the RMS voltage within a band of 315 to 320, and parameter α denotes the extent of a voltage swell, sag, or interruption. On a pure wave, the duration of the disturbance is determined by the step function u (t). The frequency and magnitude of the disturbance's fluctuation are specified by the parameters nω and αn. The harmonics consist of 3rd, 5th, and 7th harmonic components.

All basic mode disturbances arise from the overlay of one waveform onto the original waveform (C1). The magnitude of the superimposing waveform for swell has an amplitude higher than the nominal value for 'β' less than 1. Table I displays the waveforms of the six basic mode disturbances obtained by varying the appropriate parameters in a similar manner. By amalgamating two basic modes, the rest eleven mixed mode disturbances are created. For the production of the dataset, different combinations are taken into account. The simulated waveforms for normal, sag, and swell disturbances are shown in Figure 2 and waveforms for harmonics, harmonics with sag and swell combinations are shown in Figure 3.

2.2. Data Preprocessing

PQ signals are generated at a sampling frequency of 1 kHz. Using IEEE standard equations, 9000 signals are produced, each sampled at 81 points. Initially, the dataset is partitioned into a training set, referred to as train1, and a test set, known as test1, with an 80:20 ratio. Prior to feeding the encoder-decoder model, the PCA module transformed the dataset, originally comprising 56 dimensions, into primary components with five dimensions. It aims to find the principal directions in the data that capture most of the variance and represent the underlying structure of the data. This is done by transforming the data from the original feature space to a new space and capturing the maximum variations in the data. This quickens processing while improving the performance of the deep learning method.

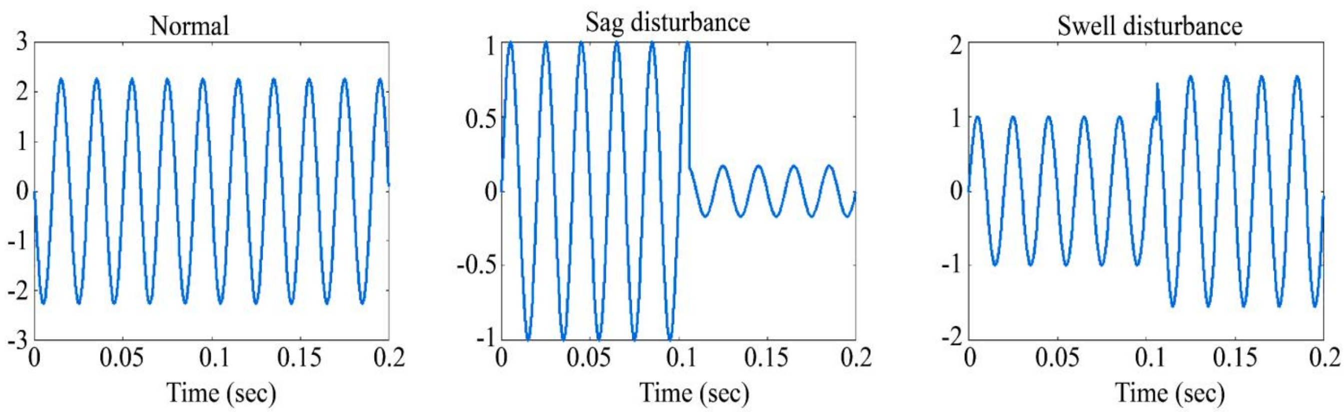

Fig. 2 Representation of normal, sag and swell disturbances

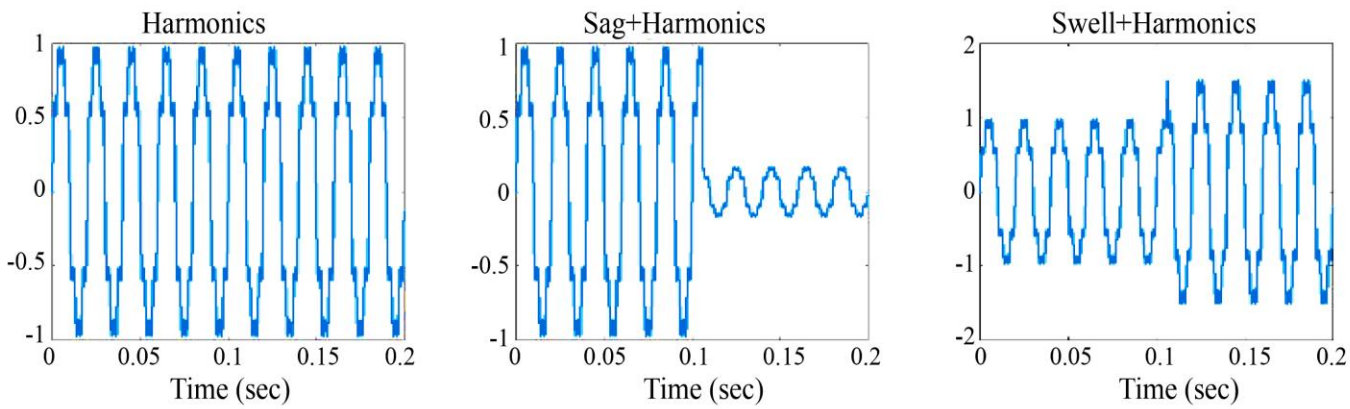

Fig. 3 Representation of harmonics, harmonics with sag and swell disturbances

Table 1. Data generation

Power Quality Disturbance	Label	Equation	Parameters
Pure wave	C 1	$A \sin(\omega t)$	2*π*50 radians per second
Sag	C 2	$A(1 - \alpha(u(t - t1) - u(t - t2))) \sin(\omega t)$	$0.01 \le \alpha \le 0.09$, T ≤ (t2 - t1) ≤ 9T
Oscillatory Transient	C 3	$\sin(\omega t) + \beta e^{-\frac{(t-t1)}{c}} * \sin(\omega_n(t - t_1)) * (u(t - t_2) - u(t - t_1))$	$0.1 \le \beta \le 0.8$, $0.5T \le (t_2 - t_1) \le 3T$, $8ms \le r30ms$ and $300Hz \le fn \le 900Hz$
Swell	C 4	$A(1 + \beta(u(t - t1) - u(t - t2))) \sin(\omega t)$	$0.1 \le \beta \le 0.8$, T ≤ (t2 - t1) ≤ 9T
Interruption	C 5	$A(1 - \rho(u(t - t1) - u(t - t2))) \sin(\omega t)$	$0.9 \le \rho \le 1$, T ≤ (t2 - t1) ≤ 9T
Flicker	C 6	$(1 + \lambda \sin(\kappa\omega t)) * \sin(\omega t))$	$0.1 \le \lambda \le 0.2$, $5 \le \kappa \le 50Hz$
Harmonics	C 7	$\sin(\omega t) + \sum_{n=3} \alpha n \sin(n\omega t)$	$0.05 \le a3, a5, a7 \le 0.1$ 5 and $\sum_n a^2 = 1$

2.3. Encoder Decoder-Based PQ Disturbance Forecasting

Encoder-decoder models are a type of neural network architecture mainly explored for machine translation and language processing applications. The encoder component converts the input sequence into a vector of fixed length, while the decoder component transforms this vector back into an output sequence. The encoder-decoder blocks are linked through a context vector. The encoder processes each token in the input sequence. A simple illustration of the encoder-decoder block is shown in Figure 4. The context vector represents a hidden state depiction of the input sequence generated by the encoder block. It is then passed into the decoder block, which generates the output sequence. The context vector effectively summarises the information within the input sequence, which the decoder utilises to generate output predictions.

The encoder component is implemented using an LSTM network, a kind of Recurrent Neural Network (RNN) that usefully apprehends long-term dependencies within sequential data by integrating memory cells and gating mechanisms. During the encoding process, the LSTM scans through the input sequence token by token, updating its internal state at each step to capture the semantic meaning of the tokens encountered. Likewise, the decoder component typically employs an LSTM as well. The decoder takes the encoded representation generated by the encoder and sequentially generates the output sequence token by token. During each step, the decoder makes predictions for the subsequent token based on the previously generated tokens and the encoded representation provided by the encoder.

The discarded outputs from the encoder at each time step ensure that solely the ultimate hidden state and cell state are used to form the context vector. This approach guarantees that the context vector encapsulates the complete significance of the input sequence. The context vector is often initialized with a fixed length vector, such as a zero vector or a learned parameter, and is updated as the encoder processes the input sequence. An LSTM layer with 100 cells is fed with the training set, train1. The network is exhibited one sample at a time from the input sequence of 56 samples. It gains knowledge of how the steps are related to one another and forms an internal representation of these connections.

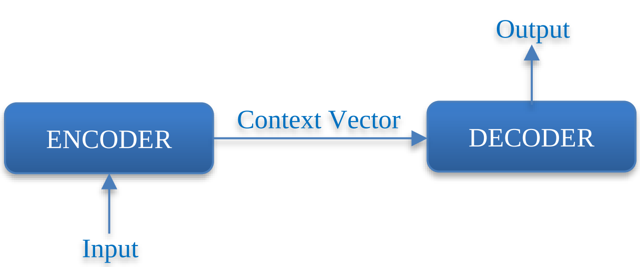

Fig. 4 Representation of encoder-decoder model

The number of memory cells determines the size of the fixed-size vector produced by the model, which, in this case, has a length of 100. The output makes use of a single dense layer. The output sequence generated in each time step is given the same weights using a time-distributed wrapper.

2.4. Detection of PQD using CNN-LSTM Network

The purpose of the detection model is to determine the type of disturbance to which the waveform predicted by the prediction model belongs. To achieve this, the training dataset is selected to include the last 25 samples predicted by the model. This dataset is then divided into two sets, train2 and test2, employing a stratified split method to ensure an equal proportion of each disturbance category within both sets. The dimensionality of the data is reduced from 25 to 15 by PCA. By doing so, it enables faster computations and produces better results for the subsequent deep learning algorithm.

The proposed model utilizes two specific architectures for disturbance identification:

- 1D Convolutional Neural Network (CNN)
- Hybrid Convolutional Neural Network-Long Short Term Memory (CNN-LSTM)

These architectures are specifically designed to process the input data and extract meaningful features for accurate disturbance classification.

2.4.1. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is primarily employed for classification tasks and recognition tasks. It has multiple layers that extract highly complex features from input data and reduce its dimensionality [2]. The convolution layer extracts the features from the input data matrix and generates the corresponding feature maps. Stride and padding are the hyperparameters used to generate the suitable output feature map. Stride is the step size by which the filter moves across the input, and stride decides the spatial resolution of the feature map. Padding is the addition of the zeros around the input before convolution. Padding helps to preserve the spatial dimensions of the input and controls the output feature map size.

The choice of stride and padding values can impact the performance of the model and can influence on what the model learns. The pooling layers down-sample the feature map and lower the number of parameters while retaining the data. The function of the fully connected layer is to map the high-level, abstract features learned by the previous layers of the network to the final output. This layer performs a linear combination of the input activations, followed by a non-linear activation function to produce the required output. It is used to make final predictions based on the features learned by the previous layers. The weights and bias are the learnable parameters used during the training stage to reduce final prediction error. A standard CNN structure comprises several convolution layers, which are then succeeded by one or more pooling layers and fully connected layers. The output after convolution can be represented as,

$$lt = tanh(xt * kt * +bt) \quad (1)$$

Where xt is the input vector, tanh is the activation function, kt is the weight, and bt is the bias of the convolution kernel, respectively. One-dimensional data on power quality disturbances are produced using mathematical models. As a result, the data is trained using a one-dimensional convolutional neural network. In each of the two convolutional layers, there are 32 kernels of size two. Convolution uses feature maps produced by various kernels to extract spatial characteristics from the data. The information is maintained after convolution using the same padding. Each convolutional layer undergoes an activation procedure after the parameters are given weights. This describes the non-linearity of the CNN model and hence aids in understanding the nonlinear property of the provided data.

The activation function is chosen for its capability to expedite the training process. Another benefit of ReLU is that the gradient vanishing problem is not present, thereby comprehending the non-linear nature inherent in the provided data. The addition of the pooling layer reduces the network's dimension and consequently its parameters. The overfitting issue is resolved, and computation time is decreased. The output dense layer employs a softmax function to determine the signal's class membership.

The training process involves two primary phases: forward propagation and backward propagation. To facilitate the forward propagation phase, the 7200 samples extracted from the train2 dataset, along with their corresponding labels, are partitioned into 20 batches. Each batch consists of 360 samples. These batches are sequentially fed into the two CNN layers and pooling layers, enabling the processing of the input data in a batch-wise fashion. The output is then flattened and passed on to the output layer. Initially, the parameters are assigned with arbitrary values.

The loss function, in this case, is the cross-entropy, which is computed within the output layer. In the backward propagation phase, the main goal is to reduce the loss function by modifying parameter values. This involves computing derivatives of weights (w) and biases (b) concerning the loss function, as determined in the forward propagation step. By iterative computing and updating the gradients of the weights, the model's parameters are adjusted in a way that gradually reduces the loss and enhances the overall performance of the network. By employing the Adam optimization technique throughout the iteration phase, the computational workload is reduced.

2.4.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network explicitly engineered to address the challenge of the vanishing gradient problem. The performance of traditional RNNs decreases as the gap length increases, whereas LSTM tackles this problem as it can retain the information for a long time. LSTM network has memory cells that are responsible for holding information for a long period of time. LSTM consists of gate mechanisms, namely input, forget, and output gates.

The gate mechanisms in LSTM help to control the flow of information into and out of memory cells. The input gate, forget gate and output gate are typically implemented as sigmoid activation functions, while the memory cell uses a tanh activation function. The forget gate determines the extent to which previous information is to be remembered. When the forget gate output, represented as ft, is equal to 0, it signifies that the specific information is disregarded or omitted. Conversely, when ft is equal to 1, it indicates that the information is retained or preserved for further processing, equation (2).

$$f_t = \sigma(W_f \,.\, [h_t - 1, x_t] + b_f) \quad (2)$$

The input cell determines the information that needs to be added to the memory cells.

$$it = \sigma\ ([ht - 1, xt] + bi) \quad (3)$$

$$ct = tanh\ (W[ht - 1, xt) + bc) \quad (4)$$

$$ct = ct-1 * ft + it * ct \quad (5)$$

The output gate decides the information allowed to pass out of the memory cells. It also influences the value of the next hidden state, which is used as input for the next time step. A sigmoid function receives the current input as well as the prior hidden state. This is multiplied with the tanh of the cell state to obtain a new hidden state, equation (7).

$$ot = \sigma\ ([ht - 1, xt] + bo) \quad (6)$$

$$ht = \ * tanh(ct) \quad (7)$$

2.5. Proposed Hybrid CNN-LSTM Model

The hybrid model combines the power of CNN and LSTM. The proposed approach comprises several key components; maximum pooling is employed with a pool size and strides both set to 2, allowing for effective downsampling of the input data. This integration enables the model to leverage the strengths of both CNN and LSTM for improved performance. The first LSTM layer is configured with 100 units, while the second LSTM layer utilizes 50 units. As for the loss function, the combination of Adam optimization and cross-entropy is employed. The model is trained using 20 batches of 360 samples each.

2.5.1. Hyper Parameter Tuning

The exceptional predictive performance of the models relies on the careful selection of optimal parameters, which is accomplished through a process known as hyperparameter tuning. In hybrid networks, crucial parameters such as the learning rate and the number of memory blocks or filters must be chosen optimally to attain the highest performance of each network.

2.6. Proposed Unified Early Detection Model

To create a comprehensive approach, an encoder-decoder model is combined with a hybrid model. The detection process centers on the predicted sampling points from the earlier prediction model. The second model is then trained using the most recent sampling points, treating the model's output as the label. The combined model is designed to predict the next 25 samples based on the given 56 samples, followed by the classification of the signal using the predicted samples. In order to assess the performance, the 25 sampling points predicted by the prediction model are fed into the classification models. The labels predicted by these models are then compared with the actual labels, enabling an evaluation of the classification accuracy.

3. Results and Discussions

3.1. Encoder Decoder-Based PQ Disturbance Forecasting

Performance was measured using the Root Mean Squared Error (RMSE), Eq. 8, which calculates the root of the squared discrepancy between the actual and predicted values. RMSE was selected as it addresses the issue of negative term cancellation by squaring the differences. Taking the root of the squared differences ensures that the unit of the loss is consistent with the projected values.

$$\text{RMSE} = \sqrt{\left(\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)\right)} \tag{8}$$

In this equation, 'n' represents the total sample count, $y_j$signifies the original signal, and $\hat{y}_j$denotes the decoded signal.

The subsequent 25 points are predicted using the initial 56 points from the test1 dataset using the encoder-decoder model. Test1 has 1800 samples, whereas Train 1 has 7200. When comparing the computation time without PCA to one with PCA of five components, the time was much longer for the dataset without PCA. Further, the RMSE for the model excluding PCA was nearly two times higher than the other models. When RMSE for each disturbance was compared, C8 had the smallest error, and C5 had the biggest. Figure 5 shows the comparison of actual and predicted values of mixed mode disturbance C8, swell, and harmonics.

3.2. Detection of PQ Disturbances

Two detection models were constructed to classify the disturbances. The effectiveness of these models was assessed using the test2 dataset, following their training on the train2 dataset. The evaluation was based on several performance metrics, including Accuracy, Recall, Precision, and F1-Score.

Accuracy, as defined by equation (9), calculates the proportion of accurately classified points in relation to the total number of points. Recall evaluates the model's capability to accurately recognize true positives. Precision quantifies the proportion of true positives in comparison to the overall number of positive predictions. Lastly, the F1 score provides a balanced measure of accuracy and recall by taking their harmonic mean.

$$\text{Accuracy} = \frac{\text{Total number of predictions}}{\text{True Positive+True negative}} \tag{9}$$

The convolutional layers in the hybrid CNN-LSTM model utilize 64 kernels of size 3. The model is trained using 20 batches, each containing 360 samples. This hybrid model combines the strengths of Convolutional Neural Networks, enabling effective feature extraction from the data, and LSTM, which captures interdependencies and also automatically detects relevant patterns. By integrating these two components, the model leverages their respective advantages to improve performance and adaptability to the data at hand.

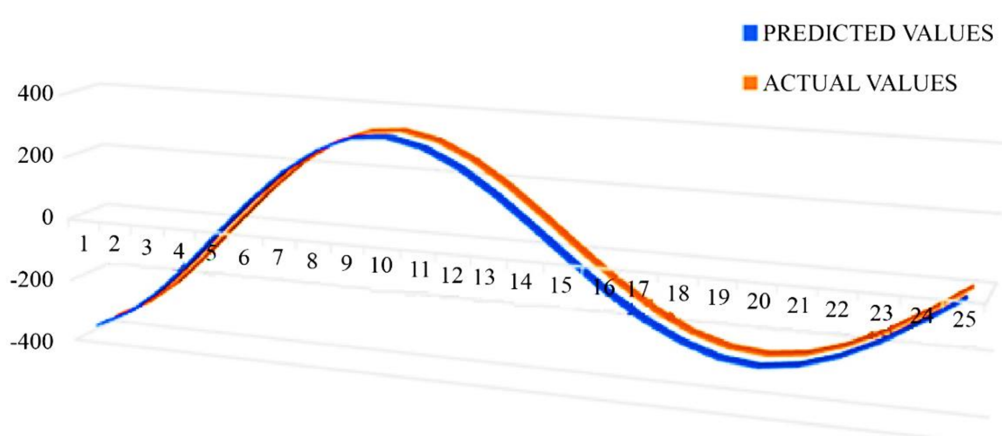

Fig. 5 Representation of encoder-decoder model

Table 2. Performance of 1D CNN model for mixed-mode disturbances

Class	Precision	Recall	F1-score
Sag+ Interruption	0.95	0.63	0.76
Swell +Interruption	0.90	1.00	0.95
Oscillation +Interruption	1.00	1.00	1.00
Sag +Oscillation	0.74	1.00	0.85
Accuracy	-	-	0.96

Table 3. Performance of CNN-LSTM model for mixed-mode disturbances

Class	Precision	Recall	F1-score
Sag+ Interruption	0.97	1.00	0.98
Swell +Interruption	0.97	0.99	0.98
Oscillation +Interruption	0.99	1.00	0.99
Sag +Oscillation	0.99	1.00	0.99
Accuracy	-	-	0.99

Table 4. Performance of combined model for mixed-mode disturbances

Class	Precision	Recall	F1-score
Swell+ Harmonics	0.95	0.95	0.95
Sagl+ Harmonics	0.57	1.00	0.73
Oscillation + Harmonics	0.98	0.98	0.98
Flicker + Harmonics	0.95	0.95	0.95
Sag +Flicker	0.90	1.00	0.95
Swell+Flicker	0.98	0.98	0.98
Oscillation +Flicker	0.95	0.95	0.95
Sag+Interruption	0.50	0.62	0.55
Swell +Interruption	0.95	0.95	0.95
Oscillation +Interruption	0.95	0.95	0.95
Sag +Oscillation	0.52	1.00	0.68
Accuracy			0.9

In order to determine the appropriate learning rate, two trials within the range of 0.01 to 0.2 are conducted. Through this, optimal performance was achieved when using a learning rate of 0.01. Beyond this point, performance gradually declined. As a result, the learning rate is set to a fixed value of 0.01. This selection ensures optimal model performance and facilitates effective learning and convergence.

The model's performance was enhanced by modifying parameters such as the count of convolutional and pooling layers, the quantity of filters and filter dimensions, and the type of padding. The model uses a single-layer LSTM network with 50 hidden units in the visible layer. This is followed by a thin, dense layer. The Adam optimizer optimized it. A batch size of 1 was used during the training, which lasted for 50 epochs. Based on the evaluation metrics of the mixed mode disturbances presented in Tables II and III, it can be deduced that the hybrid CNN-LSTM architecture outperforms the 1D-CNN approach. The accuracy of the hybrid model reached an impressive 99%, while the standard CNN model achieved an accuracy of 96%. Analysis of the confusion matrix further illustrates the superiority of the hybrid model, with only 12 misidentified signals compared to 75 in the simple model. Additionally, the simple model exhibited an increase in misclassified classes. It is worth noting that the hybrid model involved the computation of more parameters, highlighting its increased complexity but superior performance.

3.3. Unified Prediction - Classification Model

In the unified model, the 25 sampling points forecasted by the prediction model from the test1 dataset were fed to the detection models. The labels detected by these models were compared with the actual labels of test1 data. The same performance indicators used in the classification model were also used here to evaluate the performance of the unified prediction and classification model. Table IV demonstrates that the utilization of the CNN-LSTM classification model resulted in an accuracy rate of 90%. Conversely, when employing the simple CNN model, an accuracy rate of 85% was achieved. A 2% difference in the classification model turned out to be a 5% difference in the combined prediction and classification model. Also, the accuracy decreased from 98% to 90% when predicted values were used for classification. Therefore, an improved version of the prediction model could be built to obtain better classification performance for the unified model.

4. Conclusion

The integration of advanced technologies and automation has led to more complex electrical systems, with a higher risk of PQ distortions, resulting in reduced efficiency, damage to equipment, and even complete failure of the electrical system. Thus, it is paramount to forecast and classify the PQ distortions accurately to take proactive measures to prevent or mitigate the impact of these disturbances. This paper proposes an early detection of PQ distortion with an effective forecasting system using an encoder-decoder model. Further, a hybrid deep learning architecture using CNN-LSTM is proposed to effectively classify the PQ distortions. Several experiments are conducted to refine a precise and optimal model, focusing on specific network parameters. PQ distortions of 18 simple and mixed classes are considered to validate the proposed model. The classification model provides an accuracy of 98% for the hybrid CNN-LSTM model and 95% for the CNN model. Based on the performance efficacy the proposed model can be recommended for real-time PQ distortions monitoring in smart grid scenarios. It also proposes a unified model for detecting the PQ disturbances from the predicted magnitude of the disturbances. Improvement of the encoder-decoder model with attention mechanism is considered as a future scope of this work to further improve the accuracy of the combined model.

References

[1] Pankaj D. Achlerkar, S.R. Samantaray, and M. Sabarimalai Manikandan, “Variational Mode Decomposition and Decision Tree Based Detection and Classification of Power Quality Disturbances in Grid-Connected Distributed Generation System,” *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3122-3132, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[2] Carlos Iturrino Garcia et al., “A Comparison of Power Quality Disturbance Detection and Classification Methods Using CNN, LSTM and CNN-LSTM,” *Applied Sciences*, vol. 10, no. 19, pp. 1-22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yue Shen et al., “Power Quality Disturbance Monitoring and Classification Based on Improved PCA and Convolution Neural Network for Wind-Grid Distribution Systems,” *Energies*, vol. 12, no. 7, pp. 1-26, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Bizjak Boris, and Peter Planinsic, “Classification of Power Disturbances Using Fuzzy Logic,” *2006 12th International Power Electronics and Motion Control Conference*, Portoroz, Slovenia, pp. 1356-1360, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jianmin Li et al., “Detection and Classification of Power Quality Disturbances Using Double Resolution S-Transform and DAG-SVMs,” *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 10, pp. 2302-2312, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[6] Mohammad E. Salem, Azah Mohamed, and Salina Abdul Samad, “Rule Based System for Power Quality Disturbance Classification Incorporating S-Transform Features,” *Expert Systems with Applications*, vol. 37, no. 4, pp. 3229-3235, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[7] Iigo Monedero et al., “Classification of Electrical Disturbances in Real Time Using Neural Networks,” *IEEE Transactions on Power Delivery*, vol. 22, no. 3, pp. 1288-1296, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[8] Umamani Subudhi, and Sambit Dash, “Detection and Classification of Power Quality Disturbances Using GWO ELM,” *Journal of Industrial Information Integration*, vol. 22, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Shouxiang Wang, and Haiwen Chen, “A Novel Deep Learning Method for the Classification of Power Quality Disturbances Using Deep Convolutional Neural Network,” *Applied Energy*, vol. 235, pp. 1126-1140, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Jian Ma et al., “Classification of Power Quality Disturbances via Deep Learning,” *IETE Technical Review*, vol. 34, no. 4, pp. 408-415, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[11] Renxi Gong, and Taoyu Ruan, “A New Convolutional Network Structure for Power Quality Disturbance Identification and Classification in Micro-Grids,” *IEEE Access*, vol. 8, pp. 88801-88814, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Sami Ekici et al., “Power Quality Event Classification Using Optimized Bayesian Convolutional Neural Networks,” *Electrical Engineering*, vol. 103, no. 1, pp. 67-77, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] P.K. Dash, B.K. Panigrahi, and G. Panda, “Power Quality Analysis Using S-Transform,” *IEEE Transactions on Power Delivery*, vol. 18, no. 2, pp. 406-411, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[14] M. Sabarimalai Manikandan, S.R. Samantaray, and Innocent Kamwa, “Detection and Classification of Power Quality Disturbances Using Sparse Signal Decomposition on Hybrid Dictionaries,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 1, pp. 27-38, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[15] Ngo Minh Khoa, and Le Van Dai, “Detection and Classification of Power Quality Disturbances in Power System Using Modified-Combination between the Stockwell Transform and Decision Tree Methods,” *Energies*, vol. 13, no. 14, pp. 1-30, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Dar Hung Chiam, King Hann Lim, and Kah Haw Law, “LSTM Power Quality Disturbance Classification with Wavelets and Attention Mechanism,” *Electrical Engineering*, vol. 105, no. 1, pp. 259-266, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] P. Kanirajan, and V. Suresh Kumar, “Power Quality Disturbance Detection and Classification Using Wavelet and RBFNN,” *Applied Soft Computing*, vol. 35, pp. 470-481, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[18] B. Perunicic et al., “Power Quality Disturbance Detection and Classification Using Wavelets and Artificial Neural Networks,” *8th International Conference on Harmonics and Quality of Power. Proceedings (Cat. No.98EX227)*, Athens, Greece, vol. 1, pp. 77-82, 1998. [CrossRef] [Google Scholar] [Publisher Link]