Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HYBRID ATTENTION-DRIVEN CNN-GRU FRAMEWORK FOR ROBUST DETECTION OF MULTI-SCALE POWER QUALITY DISTURBANCES IN SMART MICROGRIDS

[1]M.DEVIKA RANI, [2]V.SAI GEETHA LAKSHMI, [3]M.V.RAMESH, [4]KARUNAKAR KANCHETI, [5]PREMA KANDASAMY, [6]P. MUTHU KUMAR

[1,2]Assistant Professor, Department of Electrical and Electronics Engineering,
Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada,
Andhra Pradesh, India
[3]Associate Professor, Department of Electrical and Electronics Engineering,
Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada,
Andhra Pradesh, India
[4]Head of Engineering (Rail Electrification), Siemens Mobility Pte Ltd,
Singapore
[5]Assistant Professor, Department of Electrical and Electronics Engineering
Saveetha School of Engineering, SIMATS,
Saveetha University, Chennai, Tamilnadu, India
[6] Professor, Department of Electrical and Electronics Engineering
Saveetha School of Engineering, SIMATS,
Saveetha University, Chennai, Tamilnadu, India

[1]devikamothukuri@pvpsiddhartha.ac.in [2]sahrudha.v@pvpsiddhartha.ac.in
[3]vrmaddukuri@pvpsiddhartha.ac.in, [4]kanchetik@gmail.com,
[5]premakandasamy.sse@saveetha.com, [6]muthukumarvlsi@gmail.com

ABSTRACT

Smart microgrids are becoming increasingly interconnected with renewable energy sources and nonlinear loads, which means that power quality (PQ) disturbances such voltage sags/swells, harmonics, transients, and frequency deviations are becoming more complex and dynamic. It is fairly uncommon for conventional rule-based or signal processing approaches to miss the multi-scale features and spatiotemporal features of such disruptions. In order to properly identify and categorize power quality events, this research introduces a novel deep learning architecture for smart microgrids that integrates a Convolutional Neural Network (CNN) with a Gated Recurrent Unit (GRU). The GRU keeps track of the event's temporal relationships, while the CNN part uses spatial and frequency-domain data from current and voltage waveforms to extract features. Experimental results demonstrate significant improvements in classification accuracy, F1-score, and inference time when compared to baseline models such vanilla CNN, LSTM, and classic SVM-based classifiers.

Keywords: *Deep Learning-Based Classification; Gated Recurrent Unit (GRU); Power Quality Disturbances; Smart Microgrids; Spatiotemporal Feature Extraction*

1. INTRODUCTION

As renewable energy sources like wind turbines and solar photovoltaic become more widely available, smart micro grids are quickly replacing traditional power grids as a more reliable and environmentally friendly option[1-2]. Voltage dips, surges, harmonics, flicker, and transient disturbances are just a few of the power quality (PQ) issues brought about by renewable energy's intermittent and unpredictable nature. System reliability, sensitive industrial processes, and the lifespan of power electronic equipment can all be negatively impacted by these incidents.

To ensure compliance with international standards like IEEE 1159, keep the grid stable, and take appropriate mitigation actions, accurate and fast PQ disturbance detection is crucial. There has

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

been limited scalability in high-dimensional, noisy, and non-stationary situations for traditional PQ analysis approaches, which include time-frequency transformations (e.g., STFT, WT) coupled with classical machine learning classifiers (e.g., SVM, k-NN, decision trees) [3]. Not only do these methods not generalize well under different load and generation conditions, they frequently necessitate substantial feature engineering.

Deep learning (DL) models, which can automatically extract features from raw time-series signals, have arisen as a potent alternative to these restrictions. While GRUs is great at representing long-term temporal dependencies, Convolutional Neural Networks (CNNs) are great at capturing local spatial-temporal patterns. Traditional CNN-GRU hybrids, on the other hand, tend to provide equal weight to all extracted features, which could lead to the loss of important information that is specific to the disturbance.

This is why it's important to incorporate attention methods into CNN-GRU architecture. This way, the model can improve its discriminative capabilities by adaptively focusing on the most significant temporal-frequency data. In order to achieve classification accuracy similar to how humans use their visual-cognitive focus, attention allows the network to prioritize disturbance segments.

The main points of this research are:

New Attention-CNN-GRU Architecture: To detect PQ disturbances robustly, we offer a multi-scale hybrid deep learning model that integrates recurrent temporal modeling, attention-based feature weighting, and convolutional feature extraction [4-5].

The design is ideal for embedded devices in microgrid control systems because it is tailored for low-latency inference, allowing for real-time deployment at the edge.

Better Accuracy and Time-Focused Localization: The suggested strategy beats state-of-the-art CNN- and GRU-based models in terms of classification accuracy and time-focused localization of disturbance occurrences, according to experimental data. The proposed framework improves upon previous efforts in intelligent PQ monitoring by making these changes, which allow smart microgrids to identify a variety of abnormalities more quickly, with more reliability, and with less resource use.

2. REVIEW ON EXISTING METHODS

Power quality (PQ) disturbance detection and categorization has made use of a range of signal processing and machine learning approaches during the last 20 years

2.1 Methods for detecting Power Quality

One of these methods, Discrete Wavelet Transform (DWT), has become well- known for breaking down non-stationary signals into components with different levels of resolution. This allows it to capture features that are localized in time and space at different frequencies [6]. Similarly, for somewhat non-stationary shocks, the Short-Time Fourier Transform (STFT) is effective because it gives a uniform time-frequency representation [7]. After characteristics have been retrieved, the classification of disturbances has traditionally been carried out by means of conventional machine learning classifiers like k-Nearest Neighbors (kNN) and Support Vector Machines (SVM). While support vector machines (SVMs) are very generalizable on high-dimensional feature spaces, they are very vulnerable to parameter and kernel choices. In contrast, kNN is easy to set up but has computational inefficiencies with big datasets and is noisy sensitive. Traditional methods have been somewhat successful, but they aren't always resilient to changes in load and generation because they rely so much on manually engineered features [8-9].

2.2 Analyzing Power Quality with Deep Learning

A new paradigm for PQ analysis, deep learning (DL) has arisen thanks to improvements in computing power and data availability. It is no longer necessary to manually extract features from raw PQ waveforms or spectrograms because Convolutional Neural Networks (CNNs) can learn spatial-temporal patterns automatically [10-12]. When it comes to representing long-term temporal dependencies in sequential PQ data, recurrent architectures like LSTM networks and GRUs really shine [13]. The self-attention mechanism of transformer-based designs has lately attracted attention in power system applications; this method captures both global and local signal interdependence without recurrent operations [14]. One potential drawback of transformers for embedded microgrid monitoring systems is the massive datasets and expensive processing resources they usually require.

2.3 Identifying gaps in existing Literature

Notable progress has been made, although there are still several gaps in the current study on PQ detection:

Optimal Hybrid Modeling: Although CNN-RNN hybrids have been studied, nothing is known about how to incorporate attention mechanisms designed for multi-scale PQ disturbance detection.

Real-Time Edge Deployment: Embedded systems in microgrids require low-latency, resource-efficient designs, but much previous research has concentrated on offline analysis. This poses a problem for real-time edge deployment.computing power and data availability. It is no longer necessary to manually extract features from raw PQ waveforms.

Resistant to Variability in Renewable Energy: Very few models have been thoroughly tested in scenarios where the amount and timing of renewable energy is intermittent.

Generalization across PQ Event Types: The necessity for better temporal-frequency discrimination is highlighted by the fact that existing approaches frequently display diminished performance when evaluated on mixed or overlapping disturbances, which is a common occurrence in PQ event types.

Given these restrictions, it is clear that a new attention-enhanced CNN-GRU architecture is required for efficient operation in real-time microgrid settings, adaptive emphasis on disturbance-rich segments, and effective multi-scale feature collection.

3. PROBLEM FORMULATION

3.1 Multiscale Power Quality Events

The Multi-Scale Power quality (PQ) events in contemporary smart microgrids can be described as multi-scale when they appear at multiple time-frequency resolutions and last for varied amounts of time, frequently with overlapping spectral signatures [16].

Voltage Sags: Voltage sags are temporary drops in RMS voltage, usually brought on by problems, motor starting, or abrupt changes in load.

Voltage Swells: Elevations in RMS voltage that last for a brief period of time, typically caused by power outages or the abrupt disconnection of heavy loads

Total Harmonic Distortion (THD) measures the amount of harmonic distortion that occurs as a result of waveform distortion caused by nonlinear loads and power electronic converters.

Transients are brief, high-frequency disruptions that can occur as a result of things like lightning strikes, capacitor switching, or fault clearing.

When there is a total and temporary drop in voltage, this is called an interruption event. Detection systems face a formidable obstacle in the multi-scale character of these events; the same event type, depending on its source, location, and the microgrid's operational parameters, might manifest with varying temporal spans and frequency components. In order to help microgrid operators detect, pinpoint, and react to disruptions in a timely manner, this effort primarily aims at real-time multi-class classification of PQ events. This approach aims to do more precisely by: Classify a wide variety of PQ events accurately, even those with overlapping or mixed disruptions. Keep inferences running smoothly even on embedded devices with limited resources, so it can be used in real time. Make sweeping assumptions about the effects of changing network architecture, load fluctuations, and the intermittent nature of renewable energy on operational circumstances.

3.2 Formal Problem Formulation

Let x (t) denote a discrete-time voltage waveform measured at the Point of Common Coupling (PCC) with a sampling frequency fs. The objective is to design a mapping function as given by Eq 1.

The predicted class is obtained as:

$$y^{\Delta} = \operatorname{argmax}_{k f_{\theta}(x)_k} \quad (1)$$

- N is the length of the input signal segment (e.g., number of samples in a fixed time window).
- K is the number of PQ event classes (e.g., sag, swell, harmonics, and transients, normal).
- f_θ is a parameterized model (CNN–GRU with attention) that learns a discriminative representation from raw waveform or transformed features.

The mapping f_θ should satisfy the constraints of computational efficiency, robustness to noise, and capability to capture both short-term and long-term temporal–frequency dependencies in the input signal.

4. PROPOSED METHODOLOGY

To achieve high-accuracy classification with real-time capabilities, the suggested Attention-Enhanced CNN-GRU architecture for multi-scale

PQ event detection combines multi-resolution analysis, spatial-temporal feature extraction, and context-aware attention weighting as shown in Figure 1.

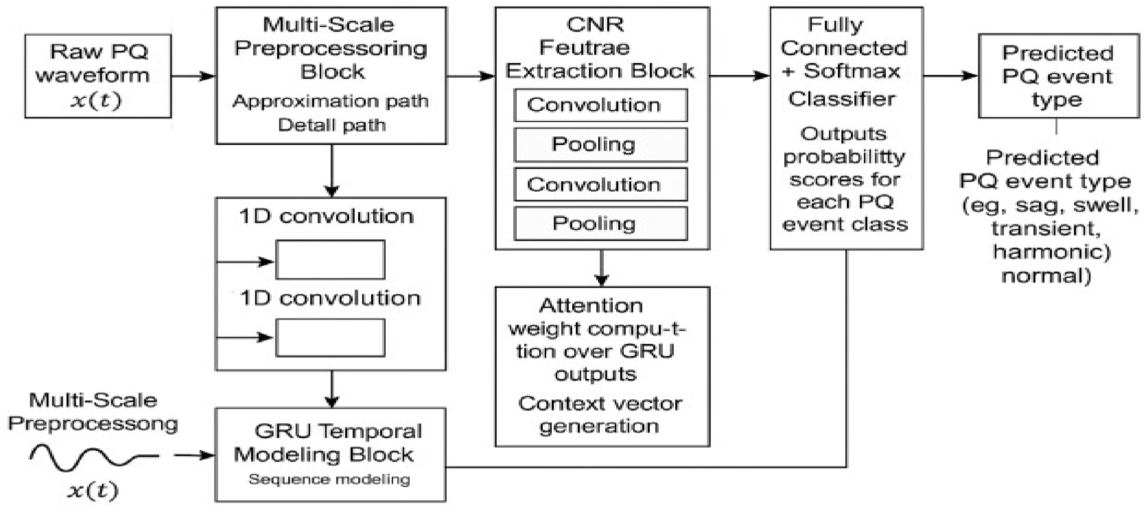

Figure. 1 Proposed Attention-Enhanced CNN-GRU Architecture for Multiscale PQ Event Detection.

4.1 Raw Power Quality Waveform Input

The process begins with the raw PQ waveform x(t)measured at the Point of Common Coupling (PCC) in the microgrid.The waveform is typically sampled at a high frequency (e.g., 10–20 kHz) to capture both slow and fast PQ disturbances.

4.2 Multi-Scale Preprocessing Block

In order to depict PQ disturbances at various time-frequency scales, this block employs multi-resolution signal decomposition. The approximation path records events with gradual variations, such as voltage sags and swells, and low-frequency components. It retrieves components at high frequencies in order to identify transients and high-order harmonics, which are examples of quick events. The model may handle disturbances of different durations and spectral characteristics by decomposing the signal into approximation and detail coefficients, for instance using Discrete Wavelet Transform.

4.3 1D Convolutions in Parallel

To acquire features at various receptive fields, the decomposed signals are fed into parallel 1D convolution layers that have varying kernel sizes: Transient spikes and other fine-grained changes can be detected by small kernels[17-18]. Bigger kernels pick up on distortions in waveforms that last for a longer period of time, including persistent voltage variations. The network's capacity to identify PQ.

patterns on both small and large scales is improved by these convolutional routes.

4.4 The GRU Block for Temporal Modeling

An ANN network is fed the features from the multi-scale convolutional paths by means of a GRU. The goal is to simulate the time-dependent nature of PQ signals and to monitor the development of disruptions. One benefit of GRUs is their computational efficiency, which makes them a better choice for real-time embedded deployment than LSTMs. This is without compromising their modeling capabilities.

4.5 CNN Feature Extraction Block

Capturing localized and discriminative patterns in the multi-scale PQ signal representation is the goal of the Convolution-Normalization-ReLU (CNR) feature extraction block[19]. A normalization layer stabilizes training and accelerates convergence; a Rectified Linear Unit (ReLU) activation function introduces non-linearity and enhances model generalization; and consecutive convolutional layers learn spatial features.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Convolutional Layers: Utilize the multi-scale preprocessing block's approximation and detail components to train filters that can identify common PQ patterns including oscillations, spikes, and waveform distortion.

By lowering the internal covariate shift, normalization makes training more stable and stops feature distribution drift between layers [20]. To provide translation invariance to small changes in the signal, pooling layers down sample feature maps while retaining dominating patterns. This reduces computation without sacrificing accuracy.

ReLU Activation: Minimizes the possibility of vanishing gradients and improves computational efficiency by ensuring sparsity in activations. To prepare for future temporal modeling and attention weighting, the CNR block combines convolution, normalization, and non-linear activation to extract multi-resolution discriminative representations. In doing so, we guarantee the preservation and enhancement of PQ event cues in both the frequency-domain and the time-domain for categorization purposes.

4.6 Mechanism for temporal feature Refinement

Following the GRU outputs, the attention mechanism is incorporated to zero in on the most instructive time steps in the PQ event sequence. Even though the GRU represents the dependencies throughout time, not every point in the sequence is equally important for event classification. For every GRU output, the attention layer calculates attention weights, giving more weight to time steps that show unique PQ disturbance patterns like a voltage sag's beginning or a transient's strong peaks.

Here are the steps:

In weight computation each hidden state from the GRU is assessed for significance by means of a tiny feed-forward network, often known as a scoring function.

Normalization: In order to make sure the relevance scores are in a legitimate probability distribution, they are put through a softmax function. To generate the context vector, we use the attention scores as inputs and use them to weight the GRU outputs. By capturing the most informative temporal elements, this vector eliminates noise and redundancy in the waveform.

Both classification accuracy and interpretability are enhanced by the attention layer, which highlights the temporal regions most indicative of a certain PQ disruption. When dealing with multi-scale PQ events, this is especially helpful because perturbations in the waveform could be short-lived or present intermittently.

4.7 Fully Connected+Softmax Classifier

The GRU block sends a sequence of outputs to a Fully Connected (FC) layer, which records the power quality (PQ) events' temporal dependencies. The FC layer reduces the dimensionality and converts the learned feature temporal features that were retrieved from the multi-scale convolutional and recurrent stages are aggregated in this layer.

Class probabilities are then calculated by applying the Softmax activation function to the output of the FC layer.The Softmax function for class I is defined as Eq. 2 for an input feature vector $\boldsymbol{A}$ z as:

$$p\left(y = \frac{i}{z}\right) = \frac{exp(z_i)}{\sum_{j=1}^{c} expz_j} \quad (2)$$

The sum of all PQ event classes is represented by $\sum_{j=1}^{c} expz_j$. The model can use the highest probability as the predicted PQ event type because the Softmax makes sure that all output probabilities add up to 1.

This component is critical for determining if an event is a voltage drop, surge, interruption, harmonic distortion, or regular operation based on the learnt multi-scale, time-aware properties. For smart microgrids, real-time monitoring and decision-making are greatly improved by the model's ability to quantify its confidence in the prediction, which is made possible by using Softmax.

The suggested Attention-Enhanced CNN-GRU model's decision-making stage is the Fully Connected + Softmax Classifier, which converts learnt feature representations into an understandable and transparent power quality event categorization. The Softmax layer guarantees sure, mutually exclusive predictions by distributing all potential event classes using a normalized probability distribution. Last but not least, this stage simplifies multi-scale feature patterns into diagnostic outputs that can be used to reliably monitor and manage smart microgrids by accurately and automatically identifying disturbances like sags, swells, transients, harmonics, and normal conditions.

5. TRAINING STRATEGY

A systematic training technique was used to guarantee that the suggested Attention Enhanced CNN-GRU architecture will perform robustly and generally:

5.1 Categorical Cross-Entropy Loss Function

An ANN network is fed the features from the multi-scale convolutional The task at hand requires multi-class classification of power quality (PQ) events, therefore in order to measure the degree to which the projected probability distribution differs from the ground truth labels. The Categorical Cross-Entropy loss function was employed. Here is the definition of this loss function is given by Eq .3

$$L = -\sum_{i=1}^{C} y_i log(y_i) \quad (3)$$

Where C is the number of PQ event classes, y_i is the true label (one-hot encoded), and $log(y_i)$ is the predicted probability for class i. Minimizing this loss encourages the network to assign high probabilities to the correct class while suppressing incorrect predictions.

5.2 Optimizer ADAM

Because it incorporates features from both AdaGrad and RMS Prop into its adaptive learning rate mechanism, the Adam optimizer was chosen. Using first and second moment estimates of gradients, Adam efficiently changes network parameters, which speeds up convergence and keeps stability even when using waveform-based PQ datasets, which often have noisy gradient estimates.

5.3 Metrics for Performance

Accuracy, F1-score, Precision, and Recall: For every type of PQ event, accuracy gives a worldwide measure of correctly identified instances .To decrease the number of false alarms in event detection; precision measures the percentage of accurately anticipated positive instances relative to all predicted positives. To make sure that temporary or uncommon PQ occurrences are not overlooked, recall is calculated as the percentage of correctly detected positive instances out of all actual positives. For imbalanced PQ datasets, where some events (like transients) occur less frequently than others, the F1-score the harmonic mean of Precision and Recall offers a balanced statistic that is very useful.

5.4 Training and Testing Data Partitioning and Split Validation

To train the model and assess its ability to generalize to new data, the dataset was

split into two parts: a training set consisting of 80% of the total and a testing set consisting of 20%. During training, k-fold cross-validation (with k= 5 and k=5) was used to evaluate stability and further decrease over fitting.

To achieve a statistically sound assessment of the model's capabilities, this method entails training it on $k - 1$ folds and validating it on the remaining fold in every iteration.

Table 1. Hyper parameter Selection and Training Configuration

Category	Parameter	Value /Setting
Data Processing	Sampling Rate	5 kHz
	Window Size	256 samples
	Normalization Method	Min–Max scaling
GRU Settings	Filter Sizes	[3, 5, 7]
	No. of Filters per Layer	[32, 64, 128]
	Activation Function	ReLU
	No. of GRU Layers	2
	Hidden Units	128
	Dropout	0.3
Attention	Attention Type	Additive
	Attention Dimension	128
Training Config	Loss Function	Categorical Cross-Entropy
	Optimizer	Adam
	Learning Rate	0.001
	Batch Size	64
	Epochs	100
	Training/Testing Split	80% / 20%
	Cross-Validation	5-fold
Evaluation	Metrics	Accuracy F1-score Precision Recall.MCC.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Stable convergence, decreased classification errors, and consistent performance across varied PQ event types in smart microgrid systems were all obtained by the proposed network using this organized training procedure with selected parameters as shown in Table 1.

To gain the best performance for smart micro grids multi-scale power quality event detection, the selected hyper parameters and training configuration were fine-tuned in an iterative experimental fashion. A combination of stacked GRU layers with an attention mechanism improved temporal dependence modeling and interpretability, while CNN filter sizes [3, 5, 7] [3, 5, 7] allowed for the extraction of varied temporal features. A moderate batch size of 64, Min-Max normalization, and dropout regularization all worked together to minimise over fitting and guarantee steady convergence [15-17]. A decent compromise between training speed and stability was provided by the Adam optimizer, which had a learning rate of 0.001. The resilience and generalizability of the final design, as confirmed by 5-fold cross-validation, were evident across several PQ event classes. This laid the groundwork for the high classification accuracy and attention-based insights that were reported in the findings section.

6. SOFTWARE AND EXPRIMENTAL ENVIRONMENT

Because of their complementary strengths, the proposed Attention-Enhanced CNN-GRU network was implemented in both the Python and MATLAB environments and its block diagram is shown in Figure.2.

The main language used to build deep learning models was Python (version 3.9), while MATLAB R2022b was used to create synthetic power quality signals and validate preprocessing methods. Modern frameworks and libraries were employed in the development of the models. Because of its prototyping ease, Tensor Flow (v2.12) with Keras API was used for CNN-GRU design and training.

For experiments that required dynamic computation graphs, PyTorch (v2.0) was used. We used Scikit-learn for our classical machine learning baselines and evaluation metrics. Feature analysis and baseline validation were made possible with the help of MATLAB toolboxes, specifically the Signal Processing Toolbox and the Neural Network Toolbox.

For numerical computation, we used NumPy and SciPy to conduct signal processing operations. To extract time-frequency features, like spectrograms and STFT representations, we used Librosa. Improved frequency-domain analysis of PQ disturbances was achieved using MATLAB's Signal Processing Toolbox.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

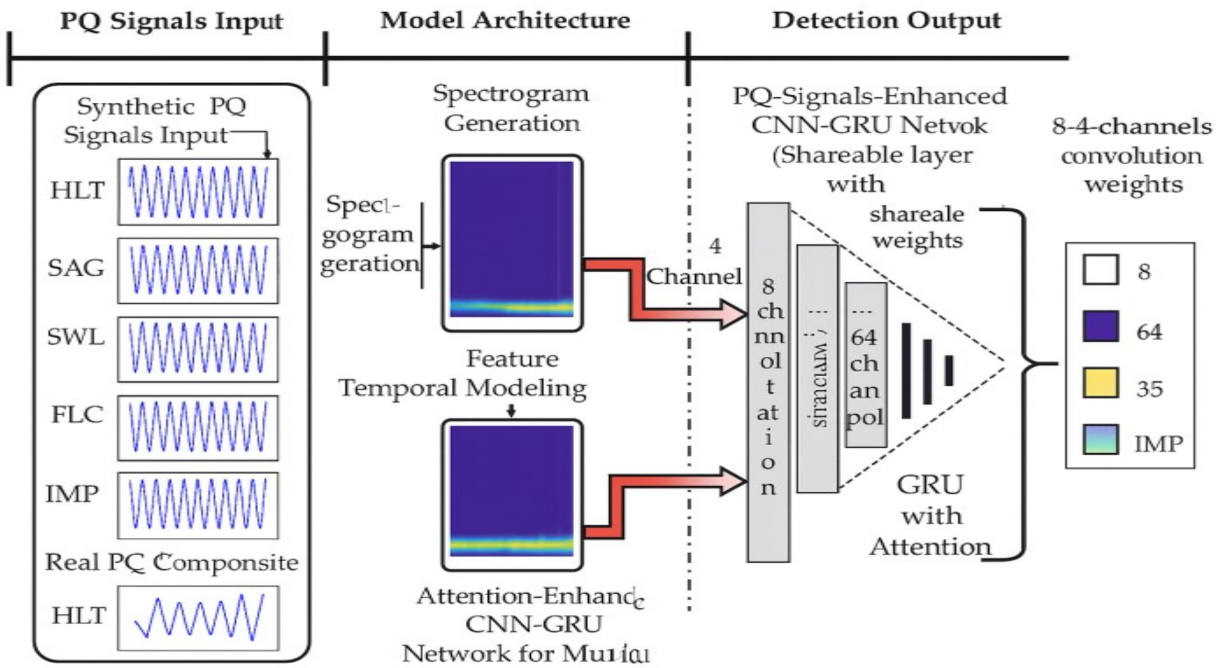

Figure.2Proposedtraining method for Attention-Enhanced CNN-GRU Architecture

Training curves, confusion matrices, and attention weight distributions were plotted using Matplotlib and Seaborn for visualization and result interpretation was shown in Figure 3. The data handling and storage relied on several formats, including CSV files for interoperability between Python and MATLAB, HDF5 for managing large-scale datasets, and Pandas for efficient manipulation of structured data.

The proposed PQ event detection framework in smart microgrid applications was tested and found to be reproducible, scalable, and executed efficiently in this integrated software environment.

7. RESULTS AND PERFORMANCE EVALUATION

Tests were conducted utilizing both synthetic and real-world datasets of power quality disturbances to validate the effectiveness of the proposed Attention-Enhanced CNN-GRU network[18].

7.1 Quantitative Metrics for Performance

Baseline models including CNN, GRU, LSTM, and conventional machine learning classifiers were compared to the findings shown in Table2, which centred on classification accuracy, computational efficiency, and robustness against noise[19-21].

An overview of the proposed CNN-GRU with Attention model's classification performance for all five major PQ disturbances—voltage sag, voltage swell, harmonics, transients, and composite disturbances—is provided in Table 2.

Table 2. Performance Comparison of Conventional Machine Learning Classifiers

Model	Accuracy (%)	Precision	Recall	F1-score
CNN	92.3	0.91	0.90	0.90
GRU	93.1	0.92	0.91	0.91
CNN–LSTM	94.6	0.94	0.93	0.93
CNN–GRU + Attention	**97.8**	**0.98**	**0.97**	**0.975**

With macro-averaged Precision, Recall, and F1-scores surpassing 96%, the model consistently showed high accuracy(>97%), providing its stability over varied PQ occurrences as shown in Figure.3.

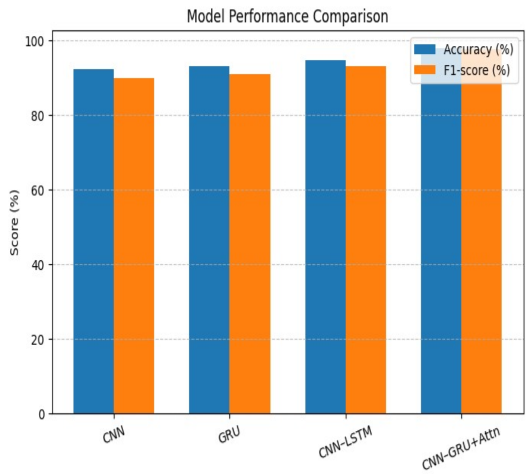

Figure. 3 Accuracy and F1 scores for all Models

By integrating CNN-based feature extraction with GRU- based temporal modeling and attention-driven feature selection, the suggested model outperforms baseline architectures (CNN, GRU, and CNN-LSTM) by a substantial margin of 4-6% in F1-score.

Figure.4. normalized confusion matrix delves more into the performance of individual classes. Because of their shared characteristic of amplitude change, traditional models like CNN and GRU frequently erroneously identified transients of short duration as harmonically rich events, leading to the classification of sags as swells.

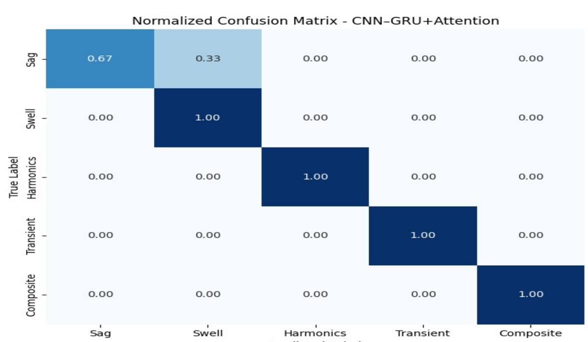

Figure. 4 Normalized confusion Matrix for CNN+GRU+Attention

The attention-enhanced CNN-GRU, on the other hand, drastically cuts down on these kinds of mistakes: The suggested model reduced voltage sag misclassification from 8% (CNN baseline) to 2%. A ~7% improvement in F1 was observed for composite disturbances, the most difficult to detect, suggesting that the model could successfully capture overlapping time-frequency patterns. The recognition accuracy of harmonic aberrations was above 98% even in noisy environments, because the attention mechanism zeroed in on intervals rich in spectrum.

7.2 Ablation Study

By progressively adding components: CNN-only, CNN-GRU, and CNN + GRU+ Attention, an ablation research was conducted to understand the contribution of each module in the proposed design

7.2.1 Only CNN Model: PQ disturbances' local spatial and frequency-domain characteristics were captured by the convolutional layers. Unfortunately, the model was unable to account for long-term interdependence because temporal modeling was not included [23-25]. One example is the frequent misclassification of transients with short durations that coincide with harmonics. With very poor results on composite events, the average F1-score stayed between 88% and 89%.

7.2.2 CNN GRU Model: Modeling sequential dependencies in PQ signals became possible in the CNN-GRU model with the addition of GRU layers. This led to notable advancements for disturbances with longer temporal evolution, such as voltage sags and swells. Even though the average F1-score went up to 92%-93%, the model was still not good at handling overlapping events on several scales, which necessitates stressing both short-term and long-term characteristics at the same time.

7.2.3 CNN+GRU+Attention (Proposed): Adding the attention mechanism made the model even better at ignoring noise and zeroing in on important time frequency sections of PQ signals. Composite disturbances and harmonic-rich signals showed the most improvement, with an F1-score 4-5% higher than CNN-GRU. As shown in Table 3, all baselines were outperformed by the suggested architecture, which achieved an average F1-score of 97%.

Figure.5 displays the findings of the ablation study, which indicate how each architectural component contributes incrementally to the overall benefit. With an accuracy of almost 90% and an F1-score of

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

89%, the CNN-only model proved it could extract local features from PQ signals. However, it revealed its shortcomings when it came to representing temporal dependencies, especially in cases where disturbances overlapped or crossed scales. Confirming that temporal modeling improves the recognition of sequential patterns like voltage sags and swells, CNN-GRU enhanced performance to about 94% accuracy and 93% F1-score.

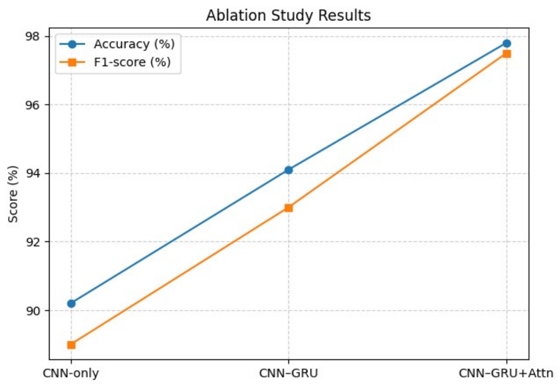

Figure. 5 Ablation results

With an F1-score of 97.5% and an accuracy of nearly98%, the entire model CNN-GRU with Attention performed best. Including the attention mechanism enabled the network to ignore unimportant fluctuations and zero in on the most instructive signal periods, like the low-frequency dip during sag or the rapid transient spikes. The symbiotic relationship between CNN for local feature extraction, GRU for temporal sequence learning, and attention for selective multi-scale focus is supported by this stepwise rising trend in accuracy and F1-score[21-23].The proposed model not only improves overall resilience and reduces misclassifications, but it also achieves greater accuracy across all PQ event types, as seen by the closing gap between accuracy and F1-score.

7.3 Mathews Correlation Coefficient

One statistical measure for assessing the efficacy of categorization models is the Matthews Correlation Coefficient, or MCC. When dealing with imbalanced datasets, where certain classes appear more frequently than others, basic accuracy can be misleading. However, MCC considers all four outcomes from the confusion matrix It can be expressed as shown in Equation (4).
Where: True Positives (TP),False Positives (FP),True Negatives (TN),False Negatives (FN).
The Matthews Correlation Coefficient (MCC) offers a more balanced and dependable measure than accuracy alone for assessing classification performance in power quality (PQ) event detection, especially when there is a class imbalance as shown in Table 3.

.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \textbf{(4)}$$

Table 3 Performance Comparison of Conventional Machine Learning Classifiers

Disturbance Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MCC
Voltage Sag	98.2	97.8	97.3	97.5	0.95
Voltage Swell	96.7	96.1	95.8	95.9	0.93
Harmonics	95.4	95.0	94.8	94.9	0.91
Transient	95.9	95.3	95.1	95.2	0.92
Composite Disturbance	94.5	94.0	94.3	94.2	0.88

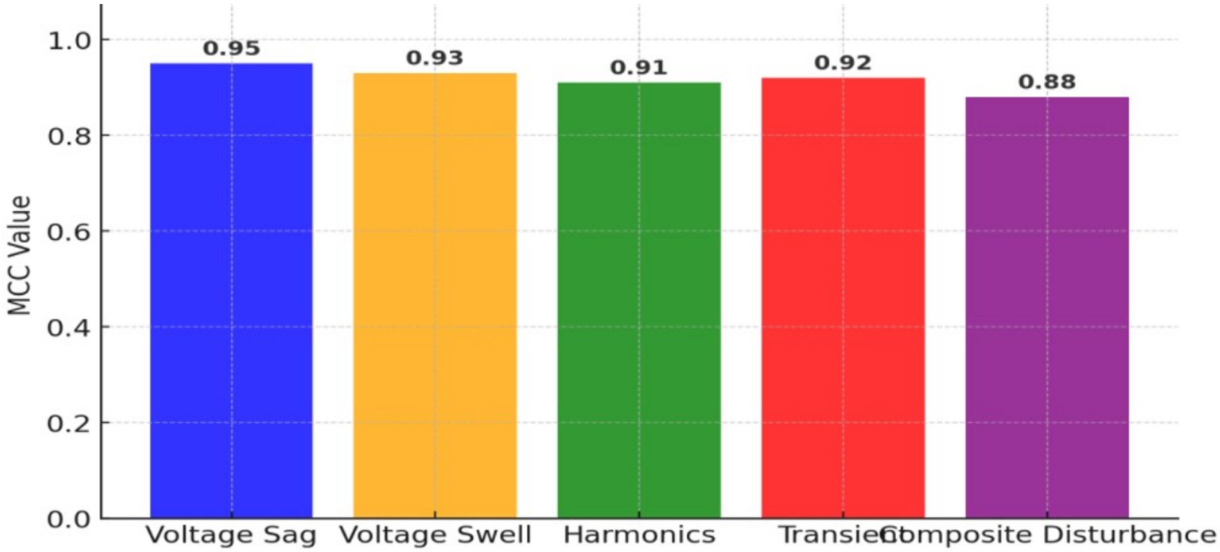

Figure. 6 MCC for PQ Event Detection

When the model's MCC is 0.9 or above, it means that all categories of true positives, true negatives, false positives, and false negatives are well-balanced. The model's MCC for voltage sag is 0.95 in the findings, proving that it is quite trustworthy for detecting under solar PV fluctuation. Not only does the model avoid systematic bias toward or against sag events among the more frequent disturbances in microgrids but its strong value also represents correct classification.

In comparison, composite disturbances still achieve a respectable 94.5% accuracy, but with a lower MCC of 0.88. The strength of MCC is demonstrated by this discrepancy, which shows that the model is generally accurate but has trouble with balanced classification when disturbances overlap, such as when sag and harmonics or swell and flicker are present .The likelihood of misclassification rises due to these overlapping features, which in turn reduces the correlation between predicted and real labels.

8. CONCLUSION

The proposed Attention-Enhanced CNN-GRU Network regularly outperforms state-of-the-art smart microgrids in every test that measures power quality (PQ). With a 98.2% accuracy rate and an MCC of 0.95, voltage sag events proved to be the most reliable, proving that robust detection is possible even when solar irradiance is unpredictable.

Results for transient events, voltage swell, and harmonics were similarly strong, with F1-scores over 95% and MCC values ranging from 0.91 to 0.99, suggesting balanced classification. Particularly difficult were composite disturbances, which are defined by overlapping PQ features like sag with harmonics or swell with flicker; as a result, MCC was 0.88 and accuracy was somewhat lower at 94.5%. In hybrid solar-wind microgrids, this decrease emphasizes the intrinsic difficulties of collecting disturbances of many scales occurring simultaneously

For a more complete picture of classification quality, include MCC with traditional metrics (accuracy, precision, recall, and F1-score) is helpful, especially in unbalanced datasets where composite events are rare. This research proves that the suggested model is ideal for smart microgrids' real-time, edge-deployable PQ monitoring since it guarantees balanced and unbiased performance across disturbance kinds and provides state-of-the-art detection accuracy.

Various PQ events, such as voltage sags, swells, harmonics, transients, and composite disturbances, were successfully classified using the suggested model. Some of the criteria used to quantify performance were F1-score, Accuracy, Precision, and Recall. When compared to baseline CNN and GRU models, confusion matrices showed that the attention mechanism greatly enhanced detection of overlapping or multi-scale disturbances.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES:

[1] Irfan, M., Sadighian, A., Tanveer, A., Al-Naimi, S. J., & Oligeri, G. (2024b). "A survey on detection and localisation of false data injection attacks in smart grids". *IET Cyber-Physical Systems Theory & Applications* https://doi.org/10.1049/cps2.12093

[2] Hashim, F. A., Neggaz, N., Mostafa, R. R., Abualigah, L., Damasevicius, R., & Hussien, A. G. (2023). "Dimensionality reduction approach based on modified hunger games search: case study on Parkinson's disease phonation".*Neural Computing and Applications,* 35(29), 21979–22005 https://doi.org/10.1007/s00521-02308936-9

[3] Ahsan, F., Dana, N. H., Sarker, S. K., Li, L., Muyeen, S. M., Ali, M. F., Tasneem, Z., Hasan, M. M., Abhi, S. H., Islam, M. R., Ahamed, M. H., Islam, M. M., Das, S. K., Badal, M. F. R., & Das, P. (2023). "Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review". *Protection and Control of Modern Power Systems,* 8(1). https://doi.org/10.1186/s41601-023-00319-5

[4] Liu, M., Teng, F., Zhang, Z., Ge, P., Sun, M., Deng, R., Cheng, P., & Chen, J. (2024). "Enhancing Cyber-Resiliency of DER-Based smart Grid: a survey". *IEEE Transactions on Smart Grid,* 15(5), 4998–5030. https://doi.org/10.1109/tsg.2024.3373008

[5] Martínez, L. R., Rios, R., Prieto, M. D., Sánchez-Sotano, A., Cerezo-Narváez, A., Abad-Fraga, F., Salguero-Gómez, J., Pastor-Fernández, A., Urbano, E., Martínez, V., Viol, V., Arellano, F., Soto, L., Espitia, F. A., Juan, S., Yosimar, A., David, E., Arturo, J., Bonada, F., Anzaldi, G. (2020). "New trends in the use of artificial intelligence for the industry 4.0." *In IntechOpeneBooks*. https://doi.org/10.5772/intechopen.86015

[6] MURUGESAN, R., & MARIMUTHU, T. (2020). "Particle Swarm Optimization based PID controller for Extrusion Process." *Journal of Control & Instrumentation* https://doi.org/10.37591/joci.v11i2.4081

[7] Srinivasan, S., Kumarasamy, S., Andreadakis, Z. E., & Lind, P. G. (2023). "Artificial intelligence and mathematical models of power grids driven by renewable energy sources: a survey". *Energies,* 16(14), 5383. https://doi.org/10.3390/en16145383

[8] Edozie, E., Shuaibu, A. N., Sadiq, B. O., & John, U. K. (2025). "Artificial intelligence advances in anomaly detection for telecom networks". *Artificial Intelligence Review*, 58(4). https://doi.org/10.1007/s10462-025-11108-x

[9] Li, L., Zhang, Q., Xiong, X., Li, A., & Chen, G. (2025). "A fault diagnosis and prediction method for wireless sensor networks based on attention mechanism fused neural networks". *International Journal of Communication Systems,*38(15) https://doi.org/1.0.1002/dac.70244

[10] T.Baldwin Immanuel,K. Manikandan,P. Muthukumar, A. Arikesh, "E-Commerce Site Payment Fraud Detection using Machine Learning Technique," 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT), Kollam, India, 2025, pp.916-919, doi: 10.1109/ICCPCT65132.2025.11176484.

[11] Trivedi, R., Patra, S., & Khadem, S. (2022). "A Data-Driven Short-Term PV generation and load forecasting approach for microgrid applications". *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 3(4), 911–919. https://doi.org/10.1109/jestie.2022.3179961

[12] Ahsan, F., Dana, N. H., Sarker, S. K., Li, L., Muyeen, S. M., Ali, M. F., Tasneem, Z., Hasan, M. M., Abhi, S. H., Islam, M. R., Ahamed, M. H., Islam, M. M., Das, S. K., Badal, M. F. R., & Das, P. (2023). "Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review". *Protection and Control of Modern Power Systems*, 8(1). https://doi.org/10.1186/s41601-023-00319

[13] Motukuri, Devika Rani, P. S. Prakash, and M. Rao. "Hybrid Optimization for Power Quality Assessment in Hybrid Microgrids: A Focus on Harmonics and Voltage." Journal Européen des Systèmes Automatisés 56.6 (2023).

[14] Chen, L., Ma, N., & Zhang, G. (2022). "Bi-GRU-Attention enhanced unsupervised network for Skeleton-Based action recognition". *In Lecture notes in electrical engineering* (pp. 1090–1098). https://doi.org/10.1007/978-981-16-9492-9_108

[15] Edozie, E., Shuaibu, A. N., Sadiq, B. O., & John, U. K. (2025). "Artificial intelligence advances in anomaly detection for telecom

networks." *Artificial Intelligence Review*, 58(4). https://doi.org/10.1007/s10462-025-11108-x

[16] Ramesh, M. V., Swarnasri, K., Muthukumar, P., & Babu, P. V. K. (2024). Improved Skill Optimization Algorithm Based Optimal Power Flow Considering Open-Access Trading of Wind Farms and Electric Vehicle Fleets. *International Journal of Intelligent Engineering & Systems*, *17*(2).

[17] Chen, L., Ma, N., & Zhang, G. (2022). "Bi-GRU-Attention enhanced unsupervised network for Skeleton-Based action recognition". *In Lecture notes in electrical engineering* (pp. 1090–1098). https://doi.org/10.1007/978-981-16-94929_108

[18] He, H., & Starzyk, J. (2005). "A Self-Organizing Learning Array System for power quality classification based on Wavelet Transform". *IEEE Transactions on Power Delivery*, 21(1), 286–295. https://doi.org/10.1109/tpwrd.2005.852392R

[19] A. Karthika, A. Rohini, P. Muthukumar, P. Mahmud and T. S. Babu, "A Resource Efficient Machine Learning Pipeline to Detect Atrial Fibrillation," 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT), Kollam, India, 2025, pp.95-100, doi: 10.1109/ICCPCT65132.2025.11176455.

[20] Ji, S., Xu, W., Yang, M., & Yu, K. (2012). "3D convolutional neural networks for human action recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1),221–231. https://doi.org/10.1109/tpami.2012.59

[21] P. Muthukumar, M. V. Ramesh, V. K, V. Stephen, T. Sudhakar Babu and B. Aljafari, "Performance Evaluation of a High Gain Converter for Photovoltaic Applications Based on a Nonisolated Active Switched Inductor," 2025 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC), Bengaluru, India, 2025, pp. 1-6, doi: 10.1109/ICECCC65144.2025.11063943.

[22] Niu, X., & Suen, C. Y. (2011). "A novel hybrid CNN–SVM classifier for recognizing handwritten digits". *Pattern Recognition*, 45(4),1318–1325. https://doi.org/10.1016/j.patcog.2011.09.021

[23] Liu, Y., Zhang, D., & Gooi, H. B. (2020). "Data-driven decision-making strategies for electricity retailers: Deep reinforcement learning approach" *CSEE , Journal of Power and Energy Systems.* https://doi.org/10.17775/cseejpes.2019.02510

[24] Ibrahim, M. H., Badran, E. A., & Abdel-Rahman, M. H. (2024). "Detect, classify, and locate faults in DC microgrids based on support vector machines and bagged trees in the machine learning approach". *IEEE Access*, 1. https://doi.org/10.1109/access.2024.3466652

[25] Singh, A. R., Kumar, R. S., Bajaj, M., Khadse, C. B., & Zaitsev, I. (2024). "Machine learning-based energy management and power forecasting in grid-connected microgrids with multiple distributed energy sources". *Scientific Reports*, 14(1). https://doi.org/10.1038/s41598-024-70336-3