

A novel hyperparameter tuned deep learning model for power quality disturbance prediction in microgrids with attention based feature learning mechanism

R. Dineshkumar^a, Anna Alphy^b, C. Kalaivanan^c, K. Bashkaran^d, Balachandra Pattanaik^e, T. Logeswaran^f, K. Saranya^g, Ganeshkumar Deivasikamani^h and A. Johny Renoald^{i,*}

^aDepartment of ECE, Saveetha School of Engineering, SIMATS, Sriperumbudur, Thandalam, Tamilnadu, India

^bDepartment of Computer Science and Engineering, SRM IST Delhi NCR Campus Ghaziabad, India

^cDepartment of EEE, Sona College of Technology, Salem, Tamilnadu, India

^dDepartment of Biomedical Engineering, Kongunadu College of Engineering and Technology, Thottiam, Trichirapalli, Tamilnadu, India

^eSchool of Electrical and Computer Engineering, Wollega University, Nekemte, Ethiopia, Africa

^fDepartment of Electrical and Electronics Engineering, Kongu Engineering College, Perundurai, Erode, Tamilnadu, India

^gDepartment of Computer Science and Engineering, Bannari Amman institute of technology, Sathyamangalam, Tamilnadu, India

^hDepartment of Biomedical Engineering, KPR Institute of Engineering and Technology, Arasur, Coimbatore, Tamilnadu, India

ⁱDepartment of Electrical and Electronics Engineering, Erode Sengunthar Engineering College, Perundurai, Erode, Tamilnadu, India

Abstract. Microgrids (MGs) have become a reliable power source for supplying energy to rural areas in a secure, consistent, and low-carbon emission manner. Power quality disturbance (PQD) is a common issue that reduces the MGs networks' reliability and restricts its usage on a small scale. The performance, reliability and lifetime of the various power devices can be affected due to the problem of PQD in the network. Researchers have proposed numerous PQD monitoring techniques based on artificial intelligence. However, they are limited to low margins and accuracy. So, this paper suggests a novel hyperparameter-tuned or optimized deep learning model with an attention-based feature learning mechanism for PQD prediction. The critical stages of the proposed work, such as data collection, feature extraction, and PQD prediction, are as follows. The PQD signals are first produced using the IEEE 1159 standard. Following that, the original time-domain features are directly recovered from the dataset, and the frequency-domain features using discrete wavelet transform (DWT). The extracted features were fed into visual geometry group 16 with multi-head attention and optimal hyperparameter-based bidirectional long short-term memory (V16MHA-OHBM) to perform spatial and temporal feature extraction. These extracted features are concatenated

*Corresponding author. A. Johny Renoald, Department of Electrical and Electronics Engineering, Erode Sengunthar Engineering

College, Perundurai, Erode-638057, Tamilnadu, India. E-mail: jorenosee@esec.ac.in.

and given to the fully connected layer to forecast the PQD. The results showed that the suggested approach surpasses the prior state-of-the-art algorithms when trained and tested using 16 different types of synthetic noise PQD data produced using mathematical models in line with IEEE 1159.

Keywords: Micro-grids, power quality disturbance, PQD prediction, data acquisition, IEEE 1159

1. Introduction

MGs are composed of loads with distributed sources of energy that may function together inside well-defined electrical boundaries to produce a single, controlled entity which can be linked to or detached from the primary grid [1, 2]. Significant changes in frequency and voltage could threaten MG stability. Smart functions in MG include generation control, weather forecasting, data transmission, and monitoring methods [3]. Additionally, the advent of MG into the power sector is a soothing strategy for resolving issues. It has numerous deployable advantages over the prior grid, including improved quality by preventing loss during transmission and distribution, increased network stability, decreased rising temperatures and pollution by utilizing technology with low carbon uptake, and a consistent electricity supply. However, MGs' power quality (PQ) has become a significant issue [4, 5]. PQ is the collection of electrical characteristics in an electrical system at a given point that is assessed against a set of predetermined reference values [6]. PQ disturbances (PQDs) are the aberrations of the reference variables in electrical characteristics, which can be observed by the producers and consumers of electrical power [7, 8]. PQD issues are divided into two categories: (i) simple PQDs, like sag, swell, surge, interruption, flicker, and harmonic distortion; and (ii) sophisticated PQDs, such as swell plus or sag plus distortions [9, 31–34].

These PQDs are a result of system flaws, a variety of loads, or environmental variables. They can also result in significant risks, including faulty relay protection and damage to delicate equipment. As a result, it is essential to identify the PQD type and implement preventative measures [10]. Therefore, it becomes essential for MG to predict PQ disruptions. Researchers have recently developed several detection and classification strategies for complex PQDs. However, the three-step method [11, 12] is the most dependable and practical for hardware integration. The primary components are signal preprocessing, eigenvalue extraction, and signal categorization [13,

35]. Most researchers use a three-step process along with cutting-edge technology to increase computational accuracy, such as a strategy based on machine learning (ML) [14, 15]. The utilization of machine learning (ML) approaches to uncover hidden data patterns, trends, and correlations has shown encouraging results [16, 36–39]. Some of the popular ML techniques for predicting the PQD include support vector machines (SVM), artificial neural networks (ANN), decision trees (DT) and random forests (RF) [17]. However, these techniques have a small margin for improvement and low accuracy. The practicality of the algorithm is also decreased by the fact that ML techniques need extra storage capacity to keep the training set data.

Recently, some authors have focused on DL-based methods to forecast PQD [18]. DL may extract temporal and spatial information from the input without signal processing. In addition to increasing accuracy, DL in the PQD problem streamlines the procedure by doing away with manual feature extraction [19]. However, DL models' hyperparameters significantly impact how well it performs [20] because the training time and storage space requirements are increased when the hyperparameters are selected randomly in the network. So, this paper proposes an optimal deep-learning method for PQD prediction with an attention-based feature-learning mechanism. The following is the list of the contributions made by the current study:

- To represent the input signal in the frequency domain using the DWT approach, which improves the performance of the classification system and its computational efficiency.
- To utilize the V16MHA mechanism to extract spatial features from time and frequency domains representation of the signals to produce the rich feature representation of the input signals.
- To extract sequential data from V16MHA using an OHBM method, the network's hyperparameters are adjusted using OLCFFHO to reduce prediction loss.

- To find the effectiveness of the suggested model, a comparative analysis of the proposed and existing frameworks for PQD prediction is done using some performance metrics.
- To find DWT analysis of normal micro grid performance signal and harmonics.

The remainder of the paper is structured as follows: The literature review, which assessed the works relevant to the proposed work, is covered in section 2. Section. 3 describes the detailed explanations of the proposed research model. The performance of the suggested methodology with various existing methodologies is examined in Section. 4 in terms of some evaluation measures. Finally, section. 5 provides conclusions and future directions to wrap up the suggested work.

2. Literature survey

PQ issues have given much interest, and numerous studies on the automatic classification of single and multiple PQDs have been undertaken.

PQDs in MG networks were detected and classified by Suganthi S.T. et al. [21] using a probabilistic-based intelligent classifier. The disrupted voltage signal from the MG systems' common bus was first used for additional processing. The features could be retrieved by applying the DWT approach to the input three-phase time-varying voltage signal. Finally, these collected features were fed into intelligent classifiers, including naive bayes (NB), SVM, and multi-layer perceptron (MLP). The testing findings demonstrated the system's usefulness by revealing a lower root mean squared error (RMSE) for the SVM, MLP, and NB classifiers, ranging from 0.2 to 0.3. A classification of PQD in solar photovoltaic (SPV) integrated MG using a hybrid deep learning method was suggested by Belkis Eristi and Huseyin Eristi [22]. Initially, an IEEE 13-bus system was used to generate the PQD data. The generated signal was segmented into fourteen cycles and scaled relative to another signal. After that, features were retrieved using continuous wavelet transforms (CWT) and a convolutional neural network (CNN), and the significant features were chosen using the neighbor component analysis (NCA) approach. The selected features were given to the hybrid deep learning model called SVM to classify the PQD in an input signal, in which the hyperparameters of the SVM were tuned by applying the grid search approach.

The system attained a computational time of 0.184s, which showed the system's superiority over others. A novel hybrid CNN was proposed by Hatem Sindi et al. [23] for the classification of PQD. First, the original form of 13 distinct PQD signals was analyzed using a one-dimensional (1D) CNN network. Then, a two-dimensional (2D) CNN was used to process these signals once they had been transformed into images. The feature vectors produced by the 1D and 2D CNN were merged, and then this merged vector was finally categorized by a fully connected layer. The system attained a sensitivity and specificity of 99.95% and 99.99%, better than the previous related schemes. A two-dimensional Deep CNN-based PQD classifier for MG was introduced by Cheng-I Chen et al. [24]. The signal synchronization of the fundamental frequency was initially carried out using an IEC 61000-4-7 synchronizer to control the image matrix. The collected signals from the divided cycles were then synchronized, converted into submatrices, and combined to create a controlled matrix. The categorization of PQD was completed using the 2D deep CNN methods. Datasets containing 14 different PQD categories were used in the experiment. This result showed that the system could enhance the efficiency of PQD prediction in real time with only a 28-second training period. Based on a hybrid machine learning algorithm with strong noise immunity, Alper Ylmaz et al. [25] established an automatic categorization of PQD in a Solid Oxide Fuel Cell & Photovoltaic (SOFC&PV)-based distributed generation. First, the system generated PQD data by IEEE standards, with 50 Hz as the primary frequency and 10 kHz as the sampling frequency. The system used UWT for signal decomposition, and the appropriate features were then extracted using the pyramidal UWT. Finally, the PQD was predicted using the stochastic gradient boosting decision tree (SGBT) approach. The system attained an accuracy of 99.59% for detecting the PQDs in the MGs. Jinsong Li et al. [26] presented a PQD prediction system using S-transform and CNN: i) Disturbance signals were extracted using the S-transform to produce a time-frequency matrix containing the disturbance signals' properties. ii) The obtained high-dimensional time-frequency modulus matrix was subjected to secondary feature extraction using CNN to reduce the data dimensions and retrieve the primary characteristics of the disturbance signal. The primary features extracted were then classified using the SoftMax classifier. iii) The results of a series of simulation experiments demonstrated that the system could accurately categorize single disturbance

signals with less training and testing time of 205 s and 0.8 s, respectively.

A hybrid technique combining the Kalman filter based on maximum likelihood (KF-ML) and deep belief network (DBN) was suggested by Yanhui Xi et al. [27] for the detection types of PQDs and their time localization. First, the noise in the original distorted signal was reduced using KF-ML, and the innovation sequence produced from the KF-ML was utilized to determine the starting and finishing times of the PQDs. After that, the DBN performed feature extraction and correctly identified the type of PQDs based on the extracted features. The system was tested using noise interferences with twenty kinds of PQDs, and the findings revealed that the prediction time was closer to the target time of the system. The absolute error achieved was less than 0.3 ms. A PQD recognition system based on a grasshopper-optimized SVM and an adaptive chirp mode pursuit (ACMP) was suggested by Shayan Z.T. et al. [28]. In the beginning, data generation in the MATLAB environment was done via parametric equations. The ACMP was then used to extract the useful features, and a graph-based technique was used to select the infinite features. In the end, the SVM technique was used to forecast the PQD, and the grasshopper optimization algorithm was used to fine-tune the SVM's hyperparameters. The trial findings showed that the system outperformed traditional methods with a better accuracy of 97.76%.

A random subspace ensemble classification system for PQD discrimination in solar photovoltaic (P.V.) M.G. power networks were proposed by Arangarajan Vinayagam et al. [29]. Initially, the signal data from various PQEs were gathered before performing the simulation of PV-integrated MG. The DWT technique was then used to extract the characteristics from the disturbance signals of various PQEs. These extracted features were fed into the SVM learners and random subspace (R.S.) ensemble for classification, which outputted ten classes of PQDs in the generated input signals. The experimentation was completed, and it was determined that the R.F. classifier provided more excellent performance and accuracy of 99% compared to the kernel-based SVM classifier. An ensemble technique was proposed by Arangarajan Vinayagam et al. [30] for PQD prediction in P.V. integrated M.G. network. Initially, the system was simulated using MATLAB/Simulink with several PQDs, and then the current and voltage data of the various PQDs were used for different processes. Then, the DWT approach extracted the essential fea-

tures from the current and voltage data. Finally, the extracted features from the DWT were given to the intelligent classifiers, namely multi-layer perceptron (MLP), Bayesian Net, and J48-based decision tree (JDT) for PQD prediction. The experimental findings demonstrated that the system achieved superior results than traditional approaches.

2.1. Problem statement

Based on a literature review, all efforts have shown increased performance in MGs for PQD prediction. However, there are certain drawbacks to these, which are as follows: Some authors [21, 25, 28, and 29] utilized ML methodology to predict PQDs in MGs. The PQEs were categorized using the most popular approaches, including DT, RF, and SVM. SVM is the most potent and successful classifier for both linear and non-linear data [21, 28, and 29]. Compared to other conventional classifiers, it also exhibits more excellent generalization performance and can competently handle a large, dimensional input vector. As a result, most of the publications described above have used the SVM technique to identify complex PQD in MG power systems, taking into account the benefits of the SVM classifier. The SVM algorithm is unsuitable for massive data sets since it contains some essential parameters that must be set precisely to get better classification outcomes for any particular scenario. In addition, because the approach necessitates the solution of a quadratic optimization problem, SVMs can be computationally expensive for big datasets. In order to classify PQD in MGs, authors have adopted a DL technique [21–24, 26, 27, and 30]. The DL approaches demonstrate their effectiveness in various disciplines, including computer vision and natural language processing. These algorithms have hierarchical architecture and several non-linear layers to extract essential features from large amounts of trained data. However, all DL algorithms struggle to learn long-term dependencies that span more than a few time steps. The information about past events is exponentially fading with more significant time steps. Additionally, the random selection of hyperparameters in DL lengthens training time and complicates the calculation. This work proposes a unique hyperparameter-tuned DL approach with an attention-based feature learning mechanism for PQD prediction in MGs to tackle the abovementioned challenges. Micro-grid with distribution network as shown in Fig. 1(a).

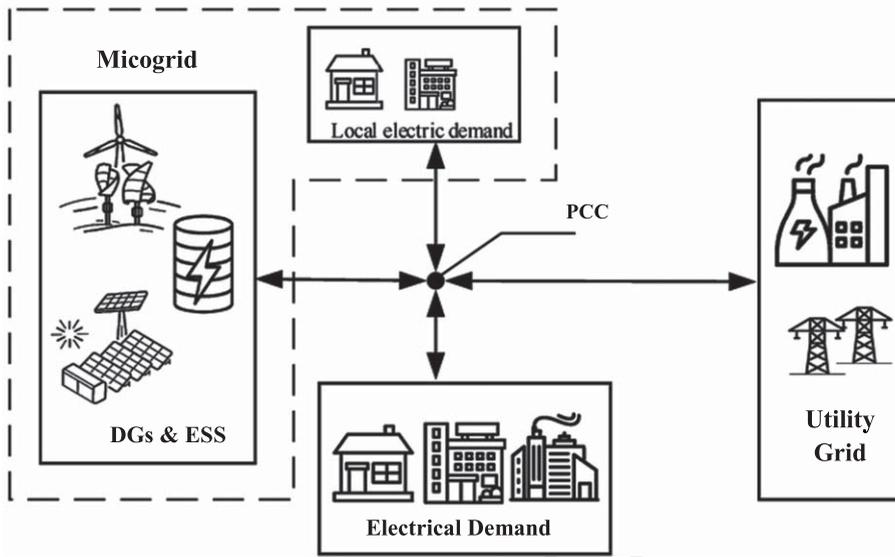


Fig. 1. (a) A micro-grid with distribution network [3].

3. Proposed methodology

This paper proposes a novel hyperparameter-tuned deep learning model with an attention-based feature learning mechanism for PQD prediction in MGs. The proposed system involved ‘3’ stages: data acquisition, feature extraction, and PQD prediction. Initially, the PQD signals are acquired from IEEE 1159. Using DWT, frequency-domain and time-domain features are extracted from the generated signals. After that, the system uses the V16MHA method to extract spatial features and the OHBM approach to extract temporal features from the extracted time and frequency domain features. Finally, the extracted spatial and temporal features are concatenated and given to the fully connected layer to yield the final output. Figure 1(b) shows the working diagram of the proposed work.

3.1. Data acquisition

Initially, the PQD signals were generated using the IEEE 1159 system, which includes different types of PQDs and their parameter variations. The threshold limits of different PQDs have been considered when generating them in the MG network according to the IEEE 1159 standard. This study produced 16 distinct types of synthetic noise PQDs data by mathematical models in line with IEEE 1159. This kind of data generation has some benefits, such as producing various signals of the same class by alter-

ing the values of parameters, which can help the classifier function more effectively. The acquired 16 types of PQD are labeled as follows: Normal (CL1), Sag (CL2), Swell (CL3), Interruption (CL4), harmonics (CL5), Flicker (CL6), Impulsive transient (CL7), Oscillatory transient (CL8), Periodic notch (CL9), Spike (CL10), Sag with harmonics (CL11), Swell with harmonics (CL12), Interruption with harmonics (CL13), Flicker with harmonics (CL14), Flicker with Sag (CL15), and Flicker with Swell (CL16). Random noises are added to each generated signal which ranges between 0 and 40 dB. The mathematical models of the PQDs are shown in Table 1 and 2 that includes 10 single type (P1-P10) and 6 multiple type signals (P11-P16). In the mathematical equations, G is a constant value (generally equal to 1) that denotes the amplitude of the waveform, α denotes the several intensity disturbances in distinct events, the duration of the disturbances is controlled by a step function $y(t)$, the disturbances occurred during t_1 to t_2 is indicated as $y(t-t_1)-y(t-t_2)$ and the real valued step function is denoted as sgn , which is computed as,

$$sgn(r) = \begin{cases} 1 & x < 0 \\ 0 & r = 0 \\ -1 & r > 0 \end{cases}$$

For every type of disturbance, a thousand waveforms of voltage signals are produced by randomly

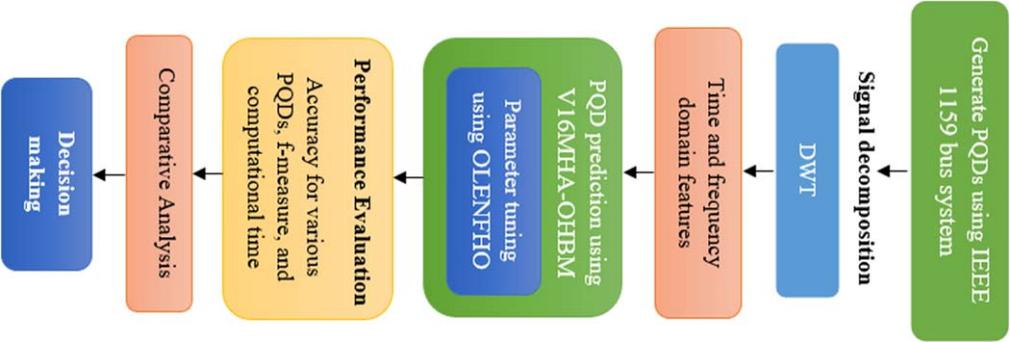


Fig. 1. (b) Proposed flow diagram.

varying the parameter values. In each PQDs waveform, the 10 cycles signal is recorded. The sampling rate and fundamental frequency are set to 50 Hz and 10 KHz, respectively.

Each waveform contains 2000 sample points and records data every 0.2 seconds under the aforementioned circumstances. This kind of data generation has some benefits, such as producing various signals of the same class by altering the values of parameters, which can help the classifier function more effectively. A total of 16,000 PQDs are divided into training, validation, and testing sets in the proportions of 6:2:2. The actual PQDs data, however, came from sensors that frequently produce noise. In order to increase the accuracy of our suggested model, different ranges of Gaussian white noise are added to the training and validation set. The signal-to-noise ratios

Table 1
Mathematical model for single PQDs

Code	Power quality disturbance	Mathematical model	Parameters
P1	Normal	$K(t) = G[1 \pm \alpha(y(t-t1) - y(t-t2))] \sin(\Omega t)$	$\alpha < 0.1, T \leq t2 - t1 \leq 9T$
P2	Sag	$K(t) = G[1 - \alpha(y(t-t1) - y(t-t2))] \sin(\Omega t)$	$0.1 \leq \alpha \leq 0.9, T \leq t2 - t1 \leq 9T$
P3	Swell	$K(t) = G[1 + \alpha(y(t-t1) - y(t-t2))] \sin(\Omega t)$	$0.1 \leq \alpha \leq 0.8, T \leq t2 - t1 \leq 9T$
P4	Interruption	$K(t) = G[1 - \alpha(y(t-t1) - y(t-t2))] \sin(\Omega t)$	$0.9 \leq \alpha \leq 1, T \leq t2 - t1 \leq 9T$
P5	Harmonics	$K(t) = G[\alpha1 \sin(\Omega t) + \alpha3 \sin(3\Omega t) + \alpha5 \sin(5\Omega t) + \alpha7 \sin(7\Omega t)]$	$0.5 \leq \alpha3, \alpha5, \alpha7 \leq 0.15, \Sigma(\alpha^2) = 1$
P6	Flicker	$K(t) = G[1 + \alpha f \sin(\beta \Omega t)] \sin(\Omega t)$	$0.5 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20 \text{ Hz}$
P7	Impulsive transient	$K(t) = G \left[\sin(\Omega t) + \alpha \exp * \frac{t-1}{\mathcal{L}} * ((y(t-t1) - y(t-t2))) \right]$	$0.1 \leq \alpha \leq 0.8, \frac{T}{20} \leq t2 - t1 \leq \frac{T}{10} * 8 \leq \mathcal{L} \leq 40$
P8	Oscillatory transient	$K(t) = G \left[\sin(\Omega t) + \alpha \exp * \frac{t-1}{\mathcal{L}} * \sin \Omega n(t-t1) * (y(t-t1) - y(t-t2)) \right]$	$0.1 \leq \alpha \leq 0.8, 0.5T \leq t2 - t1 \leq 3T * 8 \leq \mathcal{L} \leq 40, 300 \leq f_n \leq 900 \text{ Hz}$
P9	Periodic notch	$K(t) = \sin(\Omega t) - \text{sgn}(\sin(\Omega t)) * \left\{ \sum_0^9 a [y(t - (t1 - 0.02n)) - y(t - (t1 - 0.02n))] \right\}$	$0.1 \leq t1, t2 \leq 0.5T, 0.01T \leq t2 - t1 \leq 0.05T, 0.1 \leq a \leq 0.4$
P10	Spike	$K(t) = \sin(\Omega t) + \text{sgn}(\sin(\Omega t)) * \left\{ \sum_0^9 a [y(t - (t1 - 0.02n)) - y(t - (t1 - 0.02n))] \right\}$	$0.1 \leq t1, t2 \leq 0.5T, 0.01T \leq t2 - t1 \leq 0.05T, 0.1 \leq a \leq 0.4$

Table 2
Mathematical model for multiple PQDs

Code	Power quality disturbance	Mathematical model	Parameters
P11	Sag with harmonics	$K(t) = G[1 - \alpha(y(t-t_1) - y(t-t_2))] [\alpha_1 \sin(\Omega t) + \alpha_3 \sin(3\Omega t) + \alpha_5 \sin(5\Omega t) + \alpha_7 \sin(7\Omega t)]$	$0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T, 0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
P12	Swell with harmonics	$K(t) = G[1 + \alpha(y(t-t_1) - y(t-t_2))] [\alpha_1 \sin(\Omega t) + \alpha_3 \sin(3\Omega t) + \alpha_5 \sin(5\Omega t) + \alpha_7 \sin(7\Omega t)]$	$0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T, 0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
P13	Interruption with harmonics	$K(t) = G[1 - \alpha(y(t-t_1) - y(t-t_2))] [\alpha_1 \sin(\Omega t) + \alpha_3 \sin(3\Omega t) + \alpha_5 \sin(5\Omega t) + \alpha_7 \sin(7\Omega t)]$	$0.9 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T, 0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
P14	Flicker with harmonics	$K(t) = G[1 + \alpha f \sin(\beta \Omega t)] [\alpha_1 \sin(\Omega t) + \alpha_3 \sin(3\Omega t) + \alpha_5 \sin(5\Omega t) + \alpha_7 \sin(7\Omega t)]$	$0.1 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20 \text{ Hz}, 0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
P15	Flicker with Sag	$K(t) = G[1 + \alpha f \sin(\beta \Omega t)] [1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\Omega t)$	$0.1 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20 \text{ Hz}, 0.01 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$
P16	Flicker with Swell	$K(t) = G[1 + \alpha f \sin(\beta \Omega t)] [1 + \alpha(u(t-t_1) - u(t-t_2))] \sin(\Omega t)$	$0.1 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20 \text{ Hz}, 0.01 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$

of 20, 30, and 40 dB are added, correspondingly, to the test set.

3.2. Feature extraction

After generating the PQD signal, feature extraction is done. The performance of classification and computational power can both be enhanced by a successful feature extraction mechanism. The collected signals contain both time and frequency domain features, in which the frequency domain features are extracted from the signals using discrete wavelet transform (DWT), and the time domain features are extracted directly from the signals. A DWT is a transform that divides a given signal into several sets, where each set is a time series of coefficients that reflects the signal's periodic evolution in the associated frequency band. It represents the signal with many fewer wavelets and is an extremely fast time-frequency estimator. Equation (1) below represents the signal in a wavelet domain using DWT.

$$DWT(i, j) = \frac{1}{\sqrt{l_0^i}} \sum_{\delta} \hat{d}\hat{s}(\delta) \phi^* \left(\frac{j - \delta \epsilon_0 l_0^i}{l_0^i} \right) \quad (1)$$

Where, i and j refers to the dilation and scaling parameters, replaced by l_0^i and $\delta \epsilon_0 l_0^i$, respectively, $\hat{d}\hat{s}(\delta)$ denotes the discrete point sequence, δ indicates the integer, and ϕ^* represents the mother wavelet function. After that, compute the approximation $(\overline{\overline{AC}})$ and detailed coefficients $(\overline{\overline{DC}})$ of the DWT. At each level, the length of the detailed and approximation coefficients (low- and high-frequency coefficients) is equal to that of the original signal. It is expressed as follows:

$$(\overline{\overline{AC}})_x(\delta) = \sum_{i=-\infty}^{\infty} (\overline{\overline{AC}})_{x+1}(i) \xi_1(i - 2\delta) \quad (2)$$

$$(\overline{\overline{DC}})_x(\delta) = \sum_{i=-\infty}^{\infty} (\overline{\overline{DC}})_{x+1}(i) \xi_2(i - 2\delta) \quad (3)$$

Where, ξ_1 and ξ_2 refers to the low-pass and high-pass filters, respectively. Using these processes, the frequency-domain features are extracted from the dataset. Finally, the extracted time and frequency domain features are fed into the V16MHA-OHBM network for PQDs prediction.

3.3. PQD prediction

After feature extraction, the prediction of PQDs is made using the V16MHA-OHBM. The V16MHA-OHBM networks are neural networks that combine two networks, namely V16MHA (mainly convolution, pooling, and multi-head attention layer) and OHBM, for presenting the signals with spatial and temporal rich representation. This framework processes frequency domain and time domain signals separately and simultaneously. Then obtained features from V16MHA-OHBM are combined and inputted into the fully connected layer to obtain the final prediction. These are briefly explained as follows:

3.3.1. Spatial feature extraction using V16MHA

The spatial features from the time and frequency domain features are extracted using visual geometry group 16 with multi-head attention (V16MHA) mechanism. Visual geometry group 16 (V16) is a high-speed and efficient CNN network. The network includes 13 convolution layers, five max-pooling layers and three fully connected layers to perform its operations, in which the convolution and fully connected layers have tunable parameters. So, the network is named VGG-16 due to its total tunable parameters in the convolution (13 layers) and fully connected layers (3 layers). The convolution and pooling layers are frequently combined to reduce the size of the input feature maps and the computation cost. However, too many pooling layers in the network cause the feature maps to lose information about the small targets in the input data. So, this paper includes a multi-head attention (MHA) model to the V16 network to give the global feature-rich representation of the input data. In addition, the V16 network suffers from a vanishing gradient problem, affecting the network's performance in PQD prediction. So, the suggested system uses the swish activation function to solve the gradient vanishing problem of V16. These incorporations (MHA and swish activation) in conventional V16 are termed V16MHA. The V16MHA framework includes convolution, pooling, and MHA layers which are briefly explained as follows:

a) Convolution layer: The primary layer in V16MHA is convolution, which transfers low-level features of the input data into high-level features. The convolutional operator is combined with the kernel function, and the kernel size of (3×3) is applied to convolutional layers to generate the feature maps. A padding size of 1 is utilized to maintain the output

of the convolution layer, and each contains 64 filters. The output of the convolution layer is expressed using the following equation:

$$\ddot{T}_{ly} = SAF_n (\varpi^q * \ddot{N} + \Omega^q) \quad (4)$$

Where, \ddot{T}_{ly} indicates the output of the convolution layer, ϖ^q and Ω^q refers to the weight and bias of the q^{nddashth} layer, \ddot{N} represents input frequency and time domain features extracted from the signals, and $SAF_n(\cdot)$ refers to the swish activation function. It is the alternative form of the ReLU activation function. ReLU is the most widely used activation function, and it efficiently addresses the gradient disappearance issue since it is easy to compute, simple to execute, and has a quick convergence speed. However, the ReLU function reduces some of the neuron output to zero, which results in the output with migration phenomena. So, to solve these issues, the proposed work uses the swish activation function. It effectively solves the gradient vanishing problem and a smaller number of negative weights to be propagated through, where the ReLU assigns zero to all negative weights. So compared to ReLU, swish works better in deep networks for classification tasks. Equation (5) expresses the swish activation function carried out on the convolution layer of the network.

$$SAF_n(\ddot{N}) = \ddot{N} * \sigma(\ddot{N}) = \frac{\ddot{N}}{1 + e^{-\ddot{N}}} \quad (5)$$

b) Pooling layer: The output of the convolution layer is given to the max pooling layer to minimize the data dimension and retrieve important information from the data. It could mitigate the impact of data volatility, and it is superior to average pooling for retrieving the information from the PQD waveform. The resolution of the data is reduced by half using a max-pooling size of 2×2 with strides of 2 after each block. The max-pooling formulation is expressed in Equation (6).

$$J''_{ly} = \max(\ddot{T}_{ly}) \quad (6)$$

c) MHA layer: The feature maps obtained from convolution and pooling layers are passed to the MHA layer for global feature extraction. The MHA mechanism allows the network to learn richer representation between the piece of input data sequences, which leads to the performance improvement of the machine learning models. In the MHA, the 'Query', 'Key', and 'Value' are packed into matrices $\ddot{Q}\ddot{R}$, $\ddot{K}\ddot{Y}$, and $\ddot{V}\ddot{L}$, which are used in the same set of queries for the operation of the MHA function. The output matrix

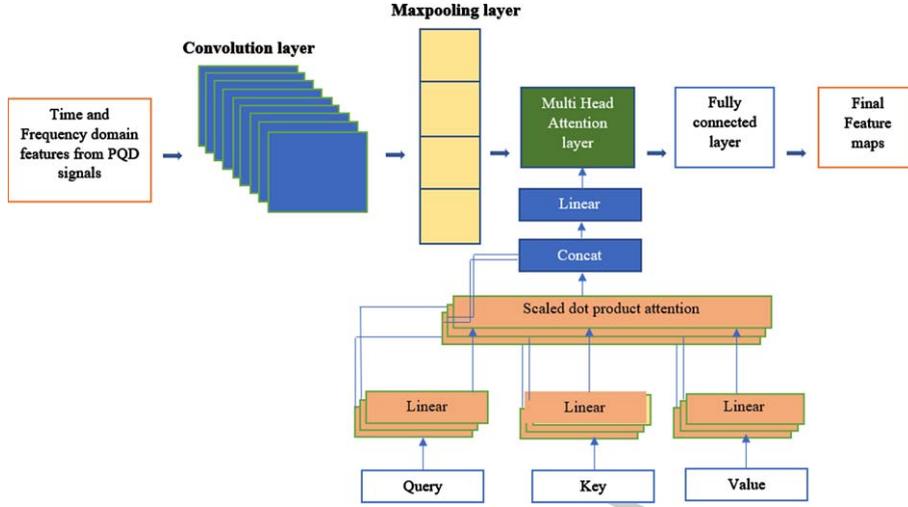


Fig. 2. (a) Layout of V16MHA.

of MHA can be defined as:

$$\overleftrightarrow{A}_{ms}(\ddot{Q}\ddot{R}, \ddot{K}\ddot{Y}, \ddot{V}\ddot{L}) = softmax\left(\frac{\ddot{Q}\ddot{R} \cdot \ddot{K}\ddot{Y}^T}{\sqrt{r}}\right) \ddot{V}\ddot{L} \quad (7)$$

Where, r —indicates the dimension of $\ddot{K}\ddot{Y}$ and $\ddot{V}\ddot{L}$. The MHA linearly processes $\ddot{Q}\ddot{R}$, $\ddot{K}\ddot{Y}$, and $\ddot{V}\ddot{L}$ multiple times via different weight matrices $\omega_d^{\ddot{Q}\ddot{R}}$, $\omega_d^{\ddot{K}\ddot{Y}}$, and $\omega_d^{\ddot{V}\ddot{L}}$. The results from the linear transformation are fed to the scaled dot product attention (SDA) and are indicated by \overleftrightarrow{I}_d , as shown in Equation (8).

$$\overleftrightarrow{I}_d = SDA\left(\ddot{Q}\ddot{R}\omega_d^{\ddot{Q}\ddot{R}}, \ddot{K}\ddot{Y}\omega_d^{\ddot{K}\ddot{Y}}, \ddot{V}\ddot{L}\omega_d^{\ddot{V}\ddot{L}}\right) \quad (8)$$

Finally, the output of the MHA mechanism is obtained by performing the following mathematical operations.

$$\begin{aligned} Head &= MULTI_{Head}(\ddot{Q}\ddot{R}, \ddot{K}\ddot{Y}, \ddot{V}\ddot{L}) \\ &= Concat\left(\overleftrightarrow{I}_1, \overleftrightarrow{I}_2, \dots, \overleftrightarrow{I}_d\right) \psi \end{aligned} \quad (9)$$

Where, d —refers to the head number of the MHA layer. The extracted spatial features are passed to the OHBM for sequential data extraction.

3.3.2. Temporal feature extraction using OHBM

The proposed system uses optimal hyperparameter-tuned bidirectional long short-term memory (OHBM) to extract the temporal features from the V16MHA network. Bidirectional long short-term memory (BM) is an extension of

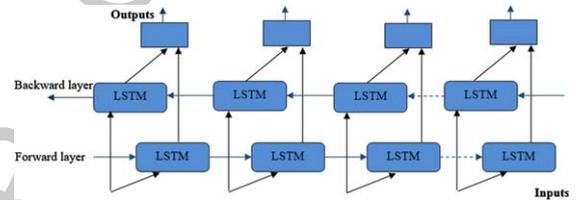


Fig. 2. (b) Structure of the BM.

long short-term memory (LSTM), which learns the dependencies between the input data in both forward and backward directions. Proposed layout of V16MHA depicted as Fig. 2(a). The structure of BM is given in Fig. 2(b). The network comprises the following layers: an input layer, a reverse LSTM layer, a forward LSTM layer, and an output layer. The memory cells of the network are controlled using the input, output and forget gates. However, the random initialization of the network’s hyperparameter takes longer to converge and does not yield good results in classification. Selecting the optimal set of hyperparameters in the network is essential because it impacts the classification performance. So, the proposed system uses oppositional learning and exponential nonlinear parameter-based fire hawks’ optimization (OLENFHO) algorithm to select the hyperparameters of the BM optimally. This optimal parameter selection in conventional BM is named OHBM.

The foraging habits of whistling kites, black kites, and brown falcons inspire the development of the metaheuristic algorithm known as fire hawk’s optimization (FHO). These birds are called fire hawks

because they use fire to capture their prey in nature. Its benefits include fewer parameters, straightforward configuration, simplicity, and excellent computation accuracy. However, the conventional FHO faces the following limitation while working on higher dimensional optimization problems: the poor balance between the exploration and exploitation capabilities, slower convergence speed and being trapped into local optimal solutions. So, the proposed system includes oppositional learning (OL) strategy as a population initialization model, which enhances the population diversity of the algorithm and avoids the premature convergence problem. In addition, the proposed system uses an exponential nonlinear (EN) parameter in the position updating stage of the algorithm, which increases the global search ability of the algorithm. These two incorporations enhance the algorithm’s performance by adequately balancing its exploration and exploitation abilities, and the conventional FHO is named OLENFHO.

The populations of fire hawks are initialized using an OL strategy that provides complementary candidates from a set of initialized solutions. Since any random solution and its opposite are preferably better than two independent solutions, OL’s central notion is to take both estimates and counter-estimates to find the best solutions. For any solution $\{h_1, h_1, \dots, h_y\}$ in a y -th dimensional space, its oppositional solution can be obtained using Equation (10).

$$\ddot{H}_{xy} = Rand \times (\overline{UB}_y - \overline{LB}_y) + \overline{LB}_y - H \quad (10)$$

Where, \ddot{H}_{xy} indicates the x -th agent at y -th dimension, \overline{UB}_y and \overline{LB}_y upper and lower boundaries of the algorithm at y -th dimension, and $Rand \in [0,1]$ refers to a random value. After that, compute the fitness of the agent, which is used to assess each \ddot{H}_x performance.

$$Fitness_{Cal} = Max (\ddot{Y}_{Accuracy}) \quad (11)$$

$$\ddot{Y}_{Accuracy} = \frac{trp + trn}{tnv} \quad (12)$$

Where, $\ddot{Y}_{Accuracy}$ indicates the accuracy metric, trp indicates the true positive, trn refers the true negative, and tnv indicates the total number of samples. The agents with higher accuracy are selected as the best at each iteration. Here, the global optimal solution is considered as the central fire that is first applied by the fire hawks and the schematic representation of the

aspect is given as follows:

$$\ddot{P}\ddot{R} = \ddot{P}\ddot{R}_1, \ddot{P}\ddot{R}_2, \dots, \ddot{P}\ddot{R}_m, \dots, \ddot{P}\ddot{R}_n \quad (13)$$

$$\ddot{F}\ddot{H} = \ddot{F}\ddot{H}_1, \ddot{F}\ddot{H}_2, \dots, \ddot{F}\ddot{H}_u, \dots, \ddot{F}\ddot{H}_v \quad (14)$$

Where, $\ddot{P}\ddot{R}_m$ indicates the m -th prey in the searching space w.r.t. the overall amount of n preys and $\ddot{F}\ddot{H}_u$ denotes the u -th fire hawk considered an overall amount of v fire hawks in the searching region. Then, the overall distance between the prey and the fire hawks is evaluated using Equation (15).

$$\overline{DT}_m^u = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}, \quad \begin{cases} u = 1, 2, 3, \dots, v \\ m = 1, 2, 3, \dots, n \end{cases} \quad (15)$$

Where, \overline{DT}_m^u indicates the overall distance betwixt the u -th fire hawk and the m -th prey, v represents the overall amount of fire hawks in the searching domain, n indicates the overall amount of prey in the searching domain, (a_1, b_1) and (a_2, b_2) characterizes the coordinate of Fire Hawks and prey in the searching domain. After that, compute the territory of $\ddot{F}\ddot{H}_u$ according to the following formula:

$$\ddot{F}\ddot{H}_u(\tau + 1) = \ddot{F}\ddot{H}_u(\tau) + \left(\overleftrightarrow{R}_1 \times B_{sol}^* - \overleftrightarrow{R}_2 \times \ddot{F}\ddot{H}_e(\tau) \right) \times \ddot{\psi} \quad (16)$$

Where, B_{sol}^* indicates the best solution, $\ddot{F}\ddot{H}_e(\tau)$ refers to one of the other fire hawks at iteration τ , \overleftrightarrow{R}_1 and \overleftrightarrow{R}_2 represents random values ranges between $[0, 1]$, and $\ddot{\psi}$ refers to the EN parameter that controls and stabilizes the abilities of exploitation and exploration and prevents the algorithm from being trapped into local optimal solutions. As a result, the global search ability of the algorithm is enhanced and it is written as follows:

$$\ddot{\Psi} = \beta \left(\beta_{min_{max}} \times e^{2\tau / max_{itr}} \right)_{min} \quad (17)$$

Where, β_{min} and β_{max} indicates the minimum and maximum of $\ddot{\Psi}$, τ refers to the current iteration, and max_{itr} denotes the maximum counts of iterations. The major feature of animal activities for updating the location is assumed using the motion of prey inside the fire hawk’s territory.

$$\begin{aligned} &\ddot{P}\ddot{R}_z(\tau + 1) \\ &= \ddot{P}\ddot{R}_z(\tau) + \left(\overleftrightarrow{R}_3 \times \ddot{F}\ddot{H}_u - \overleftrightarrow{R}_4 \times \ddot{S}_u(\tau) \right) \end{aligned} \quad (18)$$

Where, $\vec{P}\vec{R}_z(\tau)$ indicates the z -th prey's new position vector at iteration τ , and \vec{R}_3 and \vec{R}_4 represents random values between 0 and 1. The term \vec{S}_u refers to the safest position of an individual under an area of the u -th fire hawk, and it is computed as follows:

$$\vec{S}_u = \left(\frac{\sum_{z=1}^{\vec{R}} \vec{P}\vec{R}_z}{\vec{R}} \right), z = 1, 2, 3, \dots, \vec{R} \quad (19)$$

After that, the position of the prey is updated using the below formula (20):

$$\begin{aligned} \vec{P}\vec{R}_z(\tau + 1) &= \vec{P}\vec{R}_z(\tau) \\ &+ \left(\vec{R}_5 \times \vec{F}\vec{H}_{alter} - \vec{R}_6 \times \vec{S} \right) \times \vec{\Psi} \end{aligned} \quad (20)$$

Where, $\vec{F}\vec{H}_{alter}$ indicates one of the other agents in the search space, \vec{R}_5 and \vec{R}_6 indicates a random number, and \vec{S} denote the safest area outside the u -th fire hawk's territory, and it is computed as follows:

$$\vec{S} = \left(\frac{\sum_{m=1}^n \vec{P}\vec{R}_m}{\vec{R}} \right), m = 1, 2, 3, \dots, n \quad (21)$$

The solutions are continuously updated until the stopping condition is met. The agents offer optimal solutions at all iterations and are considered optimal hyperparameters for the BM network. Once getting the optimal parameters, the working progress of BM takes place in which the mathematical formulations of forward LSTM (\vec{h}_α) are defined. The BM comprises an input gate, forget gate and an output gate. The forget gate controls the effects and time dependence of previous inputs and decides which states to be forgotten or remembered, and it is computed using Equation (22).

$$\vec{L}\vec{F}_\alpha = SAF_n \left(\vec{W}_{\vec{L}\vec{F}} \cdot \left[\vec{h}_{\alpha-1}, \vec{E}\vec{S}_\alpha \right] + \vec{O}\vec{B}_{\vec{L}\vec{F}} \right) \quad (22)$$

Where, $\vec{W}_{\vec{L}\vec{F}}$ and $\vec{O}\vec{B}_{\vec{L}\vec{F}}$ represents the optimal weight and biases of forget gate $\vec{L}\vec{F}$, $\vec{E}\vec{S}_\alpha$ indicates the input extracted feature set at the time step α , $\vec{h}_{\alpha-1}$ refers to the forward LSTM cell's output, and SAF_n denotes the swish activation function, which is formulated using Equation (5). Then the input gate is responsible for selecting the current moment's degree

of consideration.

$$\vec{L}\vec{I}_\alpha = SAF_n \left(\vec{W}_{\vec{L}\vec{I}} \cdot \left[\vec{h}_{\alpha-1}, \vec{E}\vec{S}_\alpha \right] + \vec{O}\vec{B}_{\vec{L}\vec{I}} \right) \quad (23)$$

After computing the input gate, the cell state ($\vec{L}\vec{C}_\alpha$) is determined, and the resulting value updates the cell state, and is expressed by Equation (24) and the new memory state ($\vec{L}\vec{C}_\alpha$) is updated using Equation (25).

$$\vec{L}\vec{C}_\alpha = SAF_n \left(\vec{W}_{\vec{L}\vec{C}} \cdot \left[\vec{h}_{\alpha-1}, \vec{E}\vec{S}_\alpha \right] + \vec{O}\vec{B}_{\vec{L}\vec{C}} \right) \quad (24)$$

$$\vec{L}\vec{C}_\alpha = \vec{L}\vec{F}_\alpha \vec{L}\vec{C}_{\alpha-1} + \vec{L}\vec{C}_\alpha \vec{L}\vec{I}_\alpha \quad (25)$$

Finally, the output gate proffers the final output information and it is computed as follows:

$$\vec{L}\vec{O}_\alpha = SAF_n \left(\vec{W}_{\vec{L}\vec{O}} \cdot \left[\vec{h}_{\alpha-1}, \vec{E}\vec{S}_\alpha \right] + \vec{O}\vec{B}_{\vec{L}\vec{O}} \right) \quad (26)$$

$$\vec{h}_{\alpha-1} = \vec{L}\vec{O}_\alpha SAF_n (\vec{L}\vec{C}_\alpha) \quad (27)$$

These above processes are repeated to compute the backward LSTM (\overleftarrow{h}_α). Finally, we can connect the two hidden states to get the temporal features of the BM, which is formulated as follows:

$$\vec{h}_\alpha = LSTM \left(\vec{E}\vec{S}_\alpha, \vec{h}_{\alpha-1} \right) \quad (28)$$

$$\overleftarrow{h}_\alpha = LSTM \left(\vec{E}\vec{S}_\alpha, \overleftarrow{h}_{\alpha-1} \right) \quad (29)$$

$$\vec{F}\vec{P}_{op}'' = \left[\vec{h}_\alpha \oplus \overleftarrow{h}_\alpha \right] \quad (30)$$

Where \oplus is the summation of the forward and backward components and ($\vec{F}\vec{P}_{op}''$) refers to the extracted temporal features from the extracted time-domain and frequency-domain features, which also fed into the FCL layer.

3.3.3. FCL layer

The last layer is the FCL layer, which predicts the final output. It provides the final PQD prediction results. The output from the V16MHA (spatial

Table 3
Classification accuracy of the proposed approach

Disturbance types	Classification Accuracy			
	0 dB	20 dB	30 dB	40 dB
CL1	99.85	93.87	99.85	100
CL2	99.68	98.74	98.97	99.94
CL3	99.69	99.74	99.04	99.03
CL4	99.74	99.97	100	99.82
CL5	100	100	100	100
CL6	100	100	100	100
CL7	100	100	100	100
CL8	100	99.61	100	100
CL9	99.86	90.41	99.74	99.68
CL10	99.79	99.34	99.82	99.54
CL11	99.12	99.85	99.21	99.08
CL12	100	99.93	100	100
CL13	99.91	99.88	100	99.58
CL14	100	100	100	100
CL15	98.64	98.28	98.96	98.84
CL16	99.31	96.78	99.86	99.63
Average	99.72	98.53	99.72	99.70

features) and OHBM (temporal features) are concatenated and given as input to the FCL layer.

4. Results and discussion

In this section, the outcomes of the suggested methodology, a novel hyperparameter-tuned deep learning model with attention-based feature learning for PQD prediction in MGs, are investigated with state-of-the-art existing frameworks regarding some evaluation metrics. To assess the generalizability of the suggested approach, tenfold cross-validation was performed. The training was done in two stages, with the OLENFHO, dynamic learning rate, and Early-Stop. The idea is to prevent fluctuating around the optimum by approaching it as quickly as feasible with a high learning rate and then lowering the learning rate while improving the solution. The loss value is tracked to ensure that it oscillates about the optimum. The method starts with a learning rate of 0.001 and progresses to the second stage if the loss value drops 10 consecutive times. The learning rate is reduced to 0.0001 in the second phase, which ends when the loss value worsens 20 consecutive times. The suggested approach is trained and tested using an Intel Core Xeon CPU running at 2.3 GHz, 16 GB of DDR4 RAM, and an NVIDIA Tesla P100 GPU in a TensorFlow framework environment based on Python 3.7. The system was tested using the signal acquired from the standard IEEE 1159, which is briefly explained

in section 3.1, and section 4.1 examines the performance of the proposed and existing models for PQD prediction.

4.1. Performance analysis

Table 3 shows the attained accuracy of the suggested V16MHA-OHM approach at different SNR levels (0 dB to 40 dB) for various classes of PQDs (CL1 to CL16). SNR measures the strength of a desired signal relative to background noise. For 0 dB SNR, the disturbances CL5 to CL8, CL12 and CL14 attain 100% accuracy. When the SNR increases from 0 dB to 20 dB, the proposed work achieves maximum accuracy for all classes of disturbances. Similarly, varying SNR to 30 dB and 40 dB, the results are also good. Overall, in the below table, the accuracy of CL5-CL7 always gets 100% accuracy in varying all the SNR values, and the model attains better outcomes for various levels of SNR. The proposed approach achieves an average accuracy of 99.72% for 0 dB, 98.53% for 20 dB, 99.72% for 30 dB, and 99.70% for 40 dB. When varying the SNR 0 dB to 40 dB, the proposed one achieves maximum accuracy of 99.72%. Thus, it clearly shows that the proposed one achieves satisfactory performance.

The typical types of PQD voltage waveforms from the simulated system are shown in Fig. 3 and 4. The three waveforms using different colors in Fig. 3 (a-d) represent the three-phase fault waveforms. Figure 4 (a-d) shows the single-phase disturbance waveform.

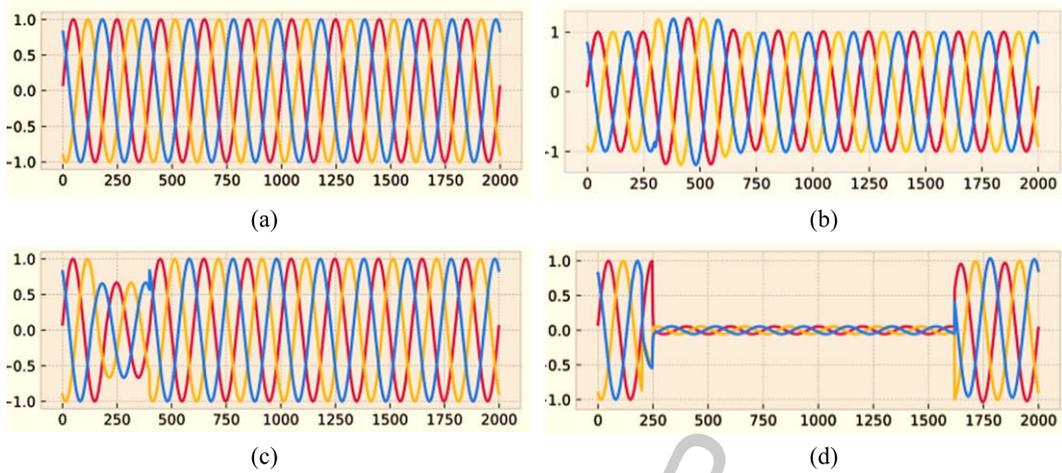


Fig. 3. Three-phase voltage waveform: (a) normal, (b) swell, (c) sag, and (d) interruption.

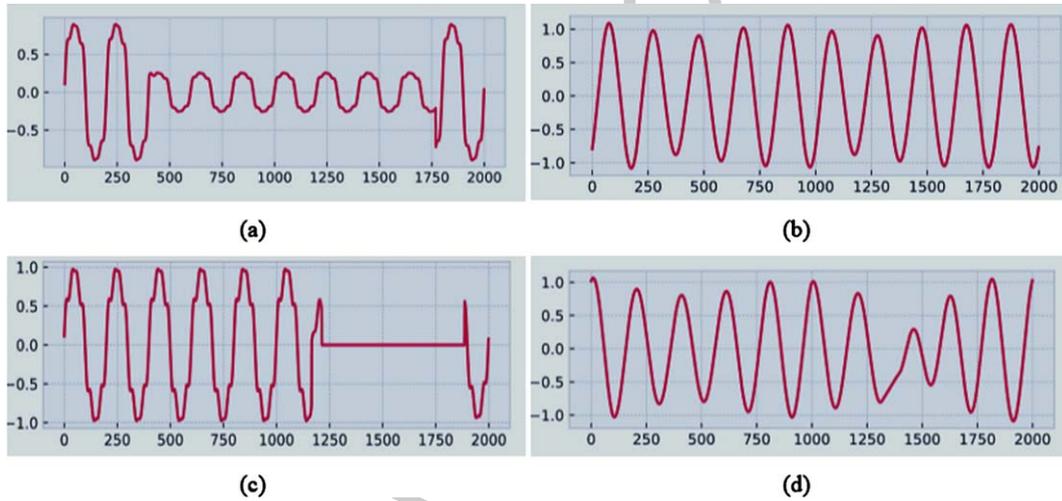


Fig. 4. Three-phase voltage waveform: one-phase voltage waveform: (a) sag with harmonics, (b) flicker, (c) interruption with harmonics, and (d) flicker with sags.

4.2. Comparative analysis

Here, the outcomes of the proposed V16MHA-OHM are compared against the existing BiLSTM, RF, recurrent neural network (RNN), SVM and CNN. The evaluation metrics were used to compare the outcomes of the suggested approach regarding the accuracy, f-measure, and computational time, respectively. These are shown in the following figure and table.

Figure 5 shows the results of the proposed V16MHA-OHM and existing approaches regarding (a) accuracy and (b) f-measure. First, concerning Fig. 5 (a), for 0 dB, the proposed one achieves maximum accuracy of 99.72%, but the existing

BiLSTM, CNN, RNN, RF, and SVM produce less accuracy of 98.85%, 95.12%, 93.64%, 91.12%, and 89.14%, respectively. For 20 dB, the existing BiLSTM, CNN, RNN, RF, and SVM attains offers accuracy of 97.74%, 93.47%, 91.17%, 89.64%, and 86.98%, respectively, which is lower when compared to the proposed one, because the proposed one achieves 98.53% accuracy. Similarly, when varying the SNR from 30 dB and 40 dB, the proposed one achieves maximum accuracy of 99.72% and 99.71%. Figure 5 (b) indicates the performance of the suggested method with existing methods in terms of the f-measure metric. The highest possible value of f-measure indicates the perfect system to detect the PQD in MG. Herein also, the proposed one achieves

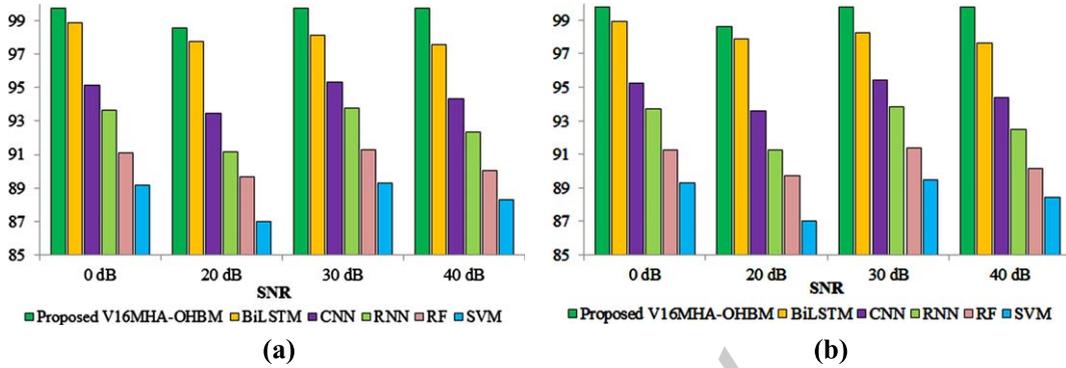


Fig. 5. (a) Accuracy and (b) F-measure analysis.

Table 4
Computational time (s) analysis

Techniques	0 dB	20 dB	30 dB	40 dB
Proposed V16MHA-OHM	0.0312	0.0591	0.0486	0.0386
BiLSTM	0.0393	0.0682	0.0573	0.0462
CNN	0.0478	0.0764	0.0648	0.0587
RNN	0.0612	0.0948	0.0874	0.0712
RF	0.0874	0.1047	0.1048	0.0987
SVM	0.0956	0.1237	0.1247	0.1143

better f-measure value than the existing methods. For example, for 0 dB, the proposed one attains a high f-measure of 99.81%, but the existing BiLSTM, CNN, RNN, RF, and SVM have f-measure of 98.93%, 95.23%, 93.72%, 91.26%, and 89.27%, which is comparatively lower than the proposed one. Likewise, the proposed system attains better results than the existing systems when varying the SNR values from 20 dB to 40 dB.

Thus, it confirmed that the proposed one outperformed the conventional methods in PQD prediction. Next, Table 4 illustrates the proposed approach performance with the existing methods concerning the computational time. Figure 6 (a & b) depicted as three phase voltage signal is recorded during the different PQD in proposed MG's, which is analysis the performance of DWT method to extract the features like normal MG signal and harmonics. The proposed system decomposition of the original voltage signal using coefficient of "D1 to D5" and approximation "a5" signals. The above results show the results of the proposed system with V16MHA based effective feature learning mechanism, in which the comparison of the proposed system with and without including swish activation function in V16MHA is given in Table 5. Thus, the overall experimental analysis shows that the proposed one attains superior performance than the existing methods. The

reason is that the proposed work initially performs feature extraction using DWT approach, which eliminates the irrelevant features and selects the important features from the dataset which increases the prediction accuracy and decreases the computational time of the classifier. In addition, the proposed system uses V16MHA and OHBM approaches to extract the spatial and temporal features from the time and frequency domain data extracted from the PQD signals, which makes the prediction results more accurate by learning global and sequential information from the signals. The hyperparameters of the deep learning model used for PQD prediction was done optimally using the OLENFHO algorithm which minimizes the classification loss and improves the prediction performance of the classifier.

In Table 4, the proposed work outcomes are investigating against the existing approaches with respect to the computational time metric. If the system took less time to predict the output, the system would be regarded as a good system. Here also, varying the SNR values, the proposed one takes less time to predict the output than the conventional methods. For example, for 40 dB, the proposed one takes only 0.0386 s to predict the output, but the existing approaches like BiLSTM, CNN, RNN, RF, and SVM take computational time of 0.0462 s, 0.0587 s, 0.0712 s, 0.0987 s, and 0.1143 s, which is higher

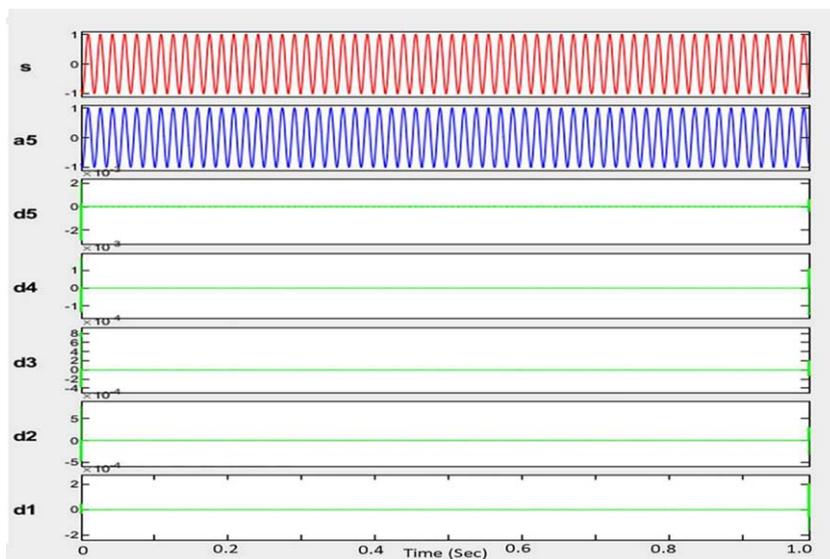


Fig. 6. (a) DWT method of proposed MG's normal signals.

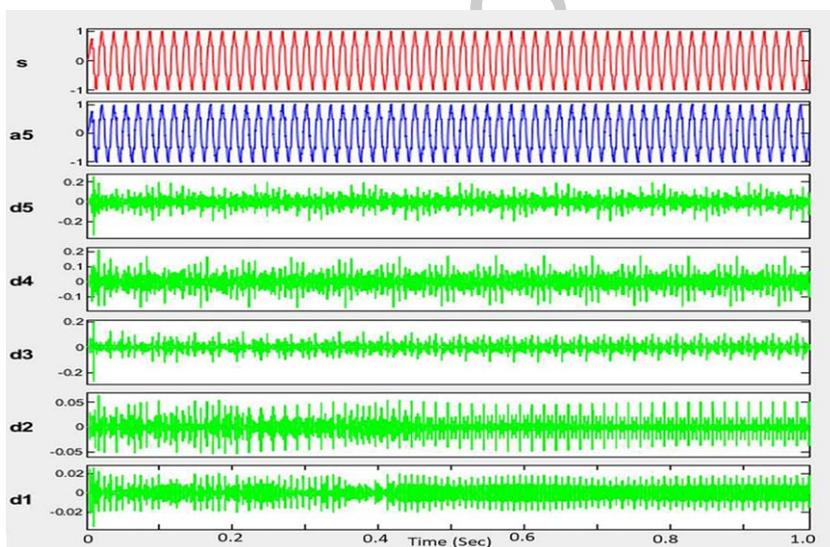


Fig. 6. (b) DWT method of proposed MG's voltage harmonics analysis.

Table 5
Analysis of the system based on proposed techniques

Technique	Accuracy (%)	F-measure (%)
Proposed system with swish in V16MHA	99.21	98.65
Proposed system without swish in V16MHA	97.87	96.12

than the proposed one. Similarly for the remaining SNR ranges (20 dB to 40 dB), the proposed one takes computational time of 0.0312 s, 0.0591 s, and 0.0486 s, this shows the suggested one takes lesser

time to predict the output. Thus, the overall experimental analysis shows that the proposed one attains superior performance than the existing methods. The reason is that the proposed work initially performs

feature extraction using DWT approach, which eliminates the irrelevant features and selects the important features from the dataset which increases the prediction accuracy and decreases the computational time of the classifier. In addition, the proposed system uses V16MHA and OHBM approaches to extract the spatial and temporal features from the time and frequency domain data extracted from the PQD signals, which makes the prediction results more accurate by learning global and sequential information from the signals. The hyperparameters of the deep learning model used for PQD prediction was done optimally using the OLENFHO algorithm which minimizes the classification loss and improves the prediction performance of the classifier.

5. Conclusion

This paper suggests a novel hyperparameter-tuned deep learning model with an attention-based feature learning mechanism for PQD prediction in MGs. The proposed system mainly comprises '3' phases: data acquisition, feature extraction, and PQD prediction. The suggested model is trained and tested using 16 types of PQD signals acquired from standard IEEE 1159. First, the proposed work's outcomes are analyzed for various SNRs (0 dB to 40 dB) in terms of accuracy metric. Here, varying 0 dB to 40 dB, some classes, say CL5-CL7 and CL14 reach 100% accuracy. Next, the proposed approach is compared against the existing BiLSTM, CNN, RNN, RF, and SVM approaches regarding the accuracy, computational time, and f-measure. First, concerning the accuracy metric, for 0 dB to 40 dB, the proposed one achieves maximum accuracy of 99.72%, 98.53%, 99.72%, and 99.71%, which is higher than the existing methods. Similarly considering f-measure and computational time metric, the proposed one attains f-measure and computational time of 99.81% and 0.0312 s for 0 dB, 98.64% and 0.0591s for 20 dB, 99.81% and 0.0486 s for 30 dB, and 99.79% and 0.0386 s for 40 dB, which were better outcomes than the conventional methods. The experimental results show that the proposed method performs better than the traditional methods, and they also demonstrate that the proposed work effectively predicts PQD in MGs. In future, this work will be extended using advanced deep learning and feature reduction mechanisms to detect the additional types of disruptions from the PQ signals with improved accuracy.

References

- [1] B.R. Lingampalli, S.R. Kotamraju, M.K. Kumar, C. Reddy, M. Pushkarna, M. Bajaj,... and S. Alphonse, Integrated Microgrid Islanding Detection with Phase Angle Difference for Reduced Nondetection Zone, *International Journal of Energy Research* **2023** (2023).
- [2] L. Tightiz, H. Yang and H. Bevrani, An Interoperable Communication Framework for Grid Frequency Regulation Support from Microgrids, *Sensors* **21**(13) (2021), 4555.
- [3] A. Ostrowska, L. Michalec, M. Skarupski, M. Jasiński, T. Sikorski, P. Kostyła,... and T. Rodziewicz, Power Quality Assessment in a Real Microgrid-Statistical Assessment of Different Long-Term Working Conditions, *Energies* **15**(21) (2022), 8089.
- [4] M. Mozaffari, K. Doshi and Y. Yilmaz, Real-Time Detection and Classification of Power Quality Disturbances, *Sensors* **22**(20) (2022), 7958.
- [5] H.F. Sindi, S. Alghamdi, M. Rawa, A.I. Omar and A.H. Elmetwaly, Robust control of adaptive power quality compensator in Multi-Microgrids for power quality enhancement using puzzle optimization algorithm, *Ain Shams Engineering Journal* **14**(8) (2023), 102047.
- [6] A. Mariscotti, Power quality phenomena, standards, and proposed metrics for dc grids, *Energies* **14**(20) (2021), 6453.
- [7] O.N. Pardo-Zamora, R.D.J. Romero-Troncoso, J.R. Millan-Almaraz, D. Morinigo-Sotelo, R.A. Osornio-Rios and J.A. Antonino-Daviu, Power Quality Disturbance Tracking Based on a Proprietary FPGA Sensor with GPS Synchronization, *Sensors* **21**(11) (2021), 3910.
- [8] J.E. Caicedo, D. Agudelo-Martínez, E. Rivas-Trujillo and J. Meyer, A systematic review of real-time detection and classification of power quality disturbances, *Protection and Control of Modern Power Systems* **8**(1) (2023), 1–37.
- [9] A.A. Abdelsalam, A.M. Hassanin and H.M. Hasanien, Categorisation of power quality problems using long short-term memory networks, *IET Generation, Transmission & Distribution* **15**(10) (2021), 1626–1639.
- [10] R.K. Beniwal, M.K. Saini, A. Nayyar, B. Qureshi and A. Aggarwal, A critical analysis of methodologies for detection and classification of power quality events in smart grid, *IEEE Access* **9** (2021), 83507–83534.
- [11] X. Xiao and K. Li, Multi-label classification for power quality disturbances by integrated deep learning, *IEEE Access* **9** (2021), 152250–152260.
- [12] V.H. Le, H. Ma, C. Ekanayake and T. Saha, On recognition of multiple power quality disturbance events in electricity networks, *Iet Generation, Transmission & Distribution* **16**(9) (2022), 1697–1713.
- [13] J. Wang, D. Zhang and Y. Zhou, Ensemble deep learning for automated classification of power quality disturbances signals, *Electric Power Systems Research* **213** (2022), 108695.
- [14] Y. Liu, T. Jin, M.A. Mohamed and Q. Wang, A novel three-step classification approach based on time-dependent spectral features for complex power quality disturbances, *IEEE Transactions on Instrumentation and Measurement* **70** (2021), 1–14.
- [15] L. Zjavka, Power quality multi-step predictions with the gradually increasing selected input parameters using machine-learning and regression, *Sustainable Energy, Grids and Networks* **26** (2021), 100442.
- [16] S.M. Miraftebadeh, M. Longo, F. Foadelli, M. Pasetti and R. Igual, Advances in the application of machine learning techniques for power system analytics: A survey, *Energies* **14**(16) (2021), 4776.

- [17] U. Subudhi and S. Dash, Detection and classification of power quality disturbances using GWO ELM, *Journal of Industrial Information Integration* **22** (2021), 100204.
- [18] R. Machlev, A. Chachkes, J. Belikov, Y. Beck and Y. Levron, Open source dataset generator for power quality disturbances with deep-learning reference classifiers, *Electric Power Systems Research* **195** (2021), 107152.
- [19] K. Sekar, K. Kanagarathinam, S. Subramanian, E. Venugopal and C. Udayakumar, An improved power quality disturbance detection using deep learning approach, *Mathematical Problems in Engineering* **2022** (2022).
- [20] A. Saxena, A.M. Alshamrani, A.F. Alrasheedi, K.A. Alnowibet and A.W. Mohamed, A Hybrid Approach Based on Principal Component Analysis for Power Quality Event Classification Using Support Vector Machines, *Mathematics* **10**(15) (2022), 2780.
- [21] S.T. Suganthi, A. Vinayagam, V. Veerasamy, A. Deepa, M. Abouhawwash and M. Thirumeni, Detection and classification of multiple power quality disturbances in Microgrid network using probabilistic based intelligent classifier, *Sustainable Energy Technologies and Assessments* **47** (2021), 101470.
- [22] B. Eristi and H. Eristi, Classification of Power Quality Disturbances in Solar PV Integrated Power System Based on a Hybrid Deep Learning Approach, *International Transactions on Electrical Energy Systems* **2022** (2022).
- [23] H. Sindi, M. Nour, M. Rawa, Ş. Öztürk and K. Polat, A novel hybrid deep learning approach including combination of 1D power signals and 2D signal images for power quality disturbance classification, *Expert Systems with Applications* **174** (2021), 114785.
- [24] C.I. Chen, S.S. Berutu, Y.C. Chen, H.C. Yang and C.H. Chen, Regulated Two-Dimensional Deep Convolutional Neural Network-Based Power Quality Classifier for Microgrid, *Energies* **15**(7) (2022), 2532.
- [25] A. Yılmaz, A. Küçükler and G. Bayrak, Automated classification of power quality disturbances in a SOFC&PV-based distributed generator using a hybrid machine learning method with high noise immunity, *International Journal of Hydrogen Energy* **47**(45) (2022), 19797–19809.
- [26] J. Li, H. Liu, D. Wang and T. Bi, Classification of power quality disturbance based on s-transform and convolution neural network, *Frontiers in Energy Research* **9** (2021), 708131.
- [27] Y. Xi, Z. Chen, X. Tang, Z. Li and X. Zeng, Type identification and time location of multiple power quality disturbances based on KF-ML-aided DBN, *IET Generation, Transmission & Distribution* **16**(8) (2022), 1552–1566.
- [28] S.Z. Motlagh and A.A. Foroud, Power quality disturbances recognition using adaptive chirp mode pursuit and grasshopper optimized support vector machines, *Measurement* **168** (2021), 108461.
- [29] A. Vinayagam, M.L. Othman, V. Veerasamy, S. Saravan Balaji, K. Ramaiyan, P. Radhakrishnan, ... and N.I. Abdul Wahab, A random subspace ensemble classification model for discrimination of power quality events in solar PV microgrid power network, *PLoS one* **17**(1) (2022), e0262570.
- [30] A. Vinayagam, V. Veerasamy, P. Radhakrishnan, M. Selperumal and K. Ramaiyan, An ensemble approach of classification model for detection and classification of power quality disturbances in PV integrated microgrid network, *Applied Soft Computing* **106** (2021), 107294.
- [31] S. Hemalatha, A. Johnny Renoald, G. Banu and K. Indirajith, Design and investigation of PV string/central architecture for bayesian fusion technique using grey wolf optimization and flower pollination optimized algorithm, *Energy Conversion and Management* **286** (2023), 117078. <https://doi.org/10.1016/j.enconman.2023.117078>.
- [32] K. Saravanan and A. Johnny Renoald, Optimizing Energy Utilization in the Weaving Industry: Advanced Electrokinetic Solutions with Modified Piezo Matrix and Super Lift Luo Converter, *Electric Power Components and Systems*, (2023). DOI: 10.1080/15325008.2023.2262458
- [33] J.R. Albert, K. Ramasamy, V.J. Michael Jerard, et al., A Symmetric Solar Photovoltaic Inverter to Improve Power Quality Using Digital Pulsewidth Modulation Approach, *Wireless Pers Commun* **130** (2023), 2059–2097. <https://doi.org/10.1007/s11277-023-10372-w>.
- [34] M. Periasamy, T. Kaliannan, S. Selvaraj, V. Manickam, S. Androse Joseph and J.R. Albert, Various PSO methods investigation in renewable and nonrenewable sources, *International Journal of Power Electronics and Drive Systems* **13**(4) (2022), pp. 2498–2505, DOI: 10.11591/ijpeds.v13.i4.pp.2498-2505.
- [35] J.R. Albert, et al., An Advanced Electrical Vehicle Charging Station Using Adaptive Hybrid Particle Swarm Optimization Intended for Renewable Energy System for Simultaneous Distributions, *Journal of Intelligent and fuzzy system* **43**(4) (2022), pp. 4395–4407. DOI: 10.3233/JIFS-220089.
- [36] Babypriya, B. Renoald, A. Johnny, M. Shyamalagowri and R. Kannan, An Experimental Simulation Testing of Single-diode PV Integrated MPPT Grid-tied Optimized Control Using Grey Wolf Algorithm, *Journal of Intelligent and fuzzy system* **43**(5) (2022), 5877–5896, DOI: 10.3233/JIFS-213259.
- [37] J.R. Albert, Design and Investigation of Solar PV Fed Single-Source Voltage-Lift Multilevel Inverter Using Intelligent Controllers, *J Control Autom Electr Syst* **33** (2022), 1537–1562. <https://doi.org/10.1007/s40313-021-00892-w>.
- [38] J.R. Albert, et al., Investigation on load harmonic reduction through solar-power utilization in intermittent SSFI using particle swarm, genetic, and modified firefly optimization algorithms, *Journal of Intelligent and fuzzy system* **42**(4) (2022), pp. 4117–4133. DOI:10.3233/JIFS-212559.
- [39] M.R. Sundarakumar, et al., A Heuristic Approach to Improve the Data Processing in Big Data Using Enhanced Salp Swarm Algorithm (ESSA) and MK-means Algorithm. **45** (2023), 2625–2640. DOI: 10.3233/JIFS-231389.