COMPARATIVE ANALYSIS OF MLP-RBF BASED NETWORKS FOR DETECTION AND CLASSIFICATION OF POWER QUALITY DISTURBANCES

Abstract

The most important factor in determining the general economic expansion of a civilization is electrical energy. The quality of power was greatly impeded by the widespread development of automated electronic equipment and the downsizing of micro-electronics utilized in power systems. In order to detect and categorize power quality issues, this research compares the capabilities of radial basis function-based classifiers and multilayer perceptron neural networks. The input feature space for the comprehensive design optimization of the RBF and MLP-NN classifiers is comprised of simple statistical parameters. Principal component analysis and sensitivity analysis are also looked at for the dimensionality reduction. With an accuracy of up to 99.81% in classifying the fundamental Power Quality disruptions, the optimized classifier is sufficiently resilient.

Key words: Power Quality, Fourier transform, Wavelet transform, Artificial intelligence, MLP, PCA

Authors

Dr. Swapnil B. Mohod
Dean
Post Graduate Studies
Prof Ram Meghe College of Engineering & Management,
Badnera-Amravati (MS), India.
swapnilmohod@gmail.com

Ketki R. Ingole
Faculty in Computer Science & Engineering
Sipna's College of Engineering
Amravati (MS), India.

I. INTRODUCTION

In today's context, Power Quality (PQ) holds paramount significance within both the power system and industrial sectors. The widespread implementation of highly sensitive automated control strategies aimed at enhancing system stability often leads to deviations from the normal supply values. These disturbances can distort the original waveform, and the increasing intricacy of the system can result in significant repercussions, including substantial economic losses in the event of component failures. Hence, it is imperative to recognize that Power Quality issues fundamentally stem from consumer-driven concerns [1]- [2].

The scope of Power Quality (PQ) issues encompasses a range of problems, including issues like overheating, motor failures, inaccuracies in metering, and the malfunctioning of protective equipment. An essential task in this field is the extraction of disturbance features from a vast array of power signals. To tackle this problem, PQ analysis researchers use a variety of methods, including the fast Fourier transform (FFT), fractional Fourier transform [3]–[4], and wavelet transform [5]–[7]. Moreover, fuzzy logic, support vector machines, and artificial neural networks (ANNs) are utilized to address PQ events of this kind.

II. PQ DISTURBANCES

Power Quality (PQ) disturbances can be categorized into two main groups: stationary and non-stationary. Stationary disturbances are characterized by time-invariant statistical properties in their waveforms. On the other hand, non-stationary disturbances have time-variant statistical characteristics in their waveforms. There are further subcategories of non-stationary disturbances, such as power frequency variations, waveform distortions, voltage imbalances, fluctuations, and short- and long-duration voltage variations.

1. **Short-Duration Voltage Variations:** Any variations in the supply voltage lasting less than a minute are categorized as short-duration voltage variations. These fluctuations may manifest as interruptions, voltage swells, or sags.

2. **Long-Duration Voltage Variations:** Extended-duration voltage variations, occasionally termed overvoltage or under voltage, along with sustained interruption, encompass any deviations in the supply root mean square (rms) voltage at the fundamental frequency lasting more than one minute.

3. **Voltage Imbalance and Voltage Fluctuation:** When the three-phase supply voltages have unequal magnitudes and possibly unequal time displacements, voltage imbalance results. Whereas voltage flicker describes extremely quick variations in the supply voltage, voltage fluctuation is a systematic random variation in the voltage.

4. **Waveform Distortion:** This is the steady-state departure of the voltage or current waveform from the ideal sine wave. DC offset and harmonics are the most common manifestations of these aberrations.

5. **Power Frequency Variation:** These are typically brought on by abrupt variations in the system's load [9].

III. DIAGNOSIS OF PQ DISTURBANCES

In order to let network operators take appropriate PQ mitigation procedures and effective remedies, it is imperative that PQ disturbances be diagnosed quickly. Power signals are often analyzed using a variety of signal processing techniques, such as the S-transform, wavelet transform, and Fourier transformation methods. Additionally, artificial intelligence (AI) techniques have made their way into the power system domain. Some researchers have also successfully applied a hybrid approach that combines the above-mentioned methods for diagnostic purposes. This section will delve further into each of these methodologies.

1. **Fourier transforms (FT):** Complex periodic functions are represented in Power Quality (PQ) analysis using the Fourier Transform (FT) as the combination of simple waves, expressed mathematically through sine and cosine functions::

- **Discrete Fourier Transform (DFT):** Discrete Fourier Transform is another name for the Fourier transform, which is primarily used on repeated signals. The DFT and frequency function of a discrete signal with finite length x[n] are provided by equations (1) and (2), respectively;

$$X(f) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] e^{\frac{-j2\Pi nf}{N}} \quad (1)$$

$$x(n) = \sum_{n=0}^{N-1} X[f] e^{\frac{j2\Pi nf}{N}} \quad (2)$$

- **Fast Fourier Transform (FFT):** The Fast Fourier Transform (FFT) plays a significant role in digital signal processing, enabling the analysis of frequency spectra. In 1948, Cooley and Tukey introduced a method for computing the N-point Discrete Fourier Transform (DFT) with a computational complexity of 2N instead of N^2. When we decompose the signal x[n] into its odd and even parts using the FFT, it can be expressed as follows:

$$FFT(x,f) = \frac{1}{2N}\sum_{n=0}^{N-1} x[2n] e^{\frac{-j\Pi(2n)f}{N}} + \frac{1}{2N}\sum_{n=0}^{N-1} x[2n+1] e^{\frac{-j\Pi(2n+1)f}{N}} \quad (3)$$

FFT reduces the computing effort from N in DFT to N log N multiplications for the same statement. This is one of the main advantages of FFT versus DFT.

- **Short Time Fourier Transform (STFT):** STFT (Short-Time Fourier Transform) is employed for analyzing signals that exhibit time-varying spectra. It involves a repetitive process of multiplying the time series by time-shifted windows and subsequently performing a Discrete Fourier Transform (DFT) on each windowed segment. The window function serves the purpose of temporally localizing the time-domain data prior to extracting frequency domain details. For continuous-time signals, the STFT can be expressed as:

$$X(\tau,\omega)=\int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-j\omega t}dt \quad (4)$$

where; x(t)= signal to be transformed, w(t)=window function, $X(\tau,\omega)$ = FT of $x(t)\omega(t-\tau)$. In case of discrete signal with 'm' as discrete time-shift, it can be expressed as

$$X(m,\omega)=\sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \quad (5)$$

Nevertheless, non-stationary signal resolution is constrained by the constant window length of a signal's STFT. Wavelet Transform (WT), a signal processing method, has been extensively used in PQ analysis to solve this resolution issue utilizing STFT.

2. **Wavelet Transform:** Wavelet Transform allows the decomposition of a signal into multiple components for in-depth analysis, offering various time-frequency resolutions. In contrast to the Short-Time Fourier Transform (STFT), the size of the smoothing window in Wavelet Transform varies depending on the analyzed frequency. There are two main types of wavelet transform that are commonly encountered: Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT). The following is a representation of the Continuous Wavelet Transform (CWT):

$$XWT(\tau,s)=\frac{1}{\sqrt{s}}\int x(t).\Psi(\frac{t-\tau}{s})dt \quad (6)$$

where; x(t)=signal to be analysed, $\psi(t)$ is the mother wavelet, s and τ represent scale and translational parameters respectively.

For the purpose of breaking down the signal into a group of mutually orthogonal wavelets, DWT is a discrete counter of CWT. Using a discrete set of wavelet scales and translations, WT complies with certain predetermined guidelines.

3. **Artificial Intelligence (AI):** Artificial Neural Networks (ANN) are inherently self-learning systems, and in addition to traditional training algorithms, hybrid algorithms and combined approaches involving ANN and methods like wavelet and S-transform have been employed. ANN possesses the ability to approximate non-linear functions and has gained extensive utilization within this field. It has found practical use in industrial applications, as well as in emerging concepts like distributed generation, particularly for Power Quality (PQ) analysis and enhancement.

IV. EXPERIMENTAL SETUP AND DATA GENERATION

The classifier based on neural networks is built and refined to identify and categorize PQ disturbances. The following is the initial experimental setup for data collection:

In experimentation following PQ disturbances are considered,

- Voltage Sag
- Voltage Swell
- Arcing load influence
- Short circuit condition

Figure 1: Experimental Setup

In the laboratory setup, an induction motor is employed, powered by the mains at 230V, single-phase, 50Hz, with a rating of one HP. The motor, a squirrel cage induction type, is utilized to analyze voltage sags and swells in the system through repetitive ON/OFF operations. A welding machine operating at the same voltage and frequency is utilized to introduce actual arcing load effects into the system. Short-circuited welding electrodes are used to simulate short circuit scenarios for experimental purposes.

A Tektronix Digital Storage Oscilloscope (DSO) of model TPS 2014 B is used to record current signals. This oscilloscope has a 100 MHz bandwidth and a 1 GHz sampling rate that may be adjusted. Tektronix voltage probes, rated at 1000V and with an approximate bandwidth of 200 MHz, are used to record current signals. A total of 100 sets of signals representing a range of mains supply circumstances are captured at a sampling frequency of 10 kHz throughout the trials.

To replicate a weak system environment within the laboratory, a 2 ½ core cable, measuring 200 meters in length, is employed. This setup allows for the observation of the effects of voltage sags, swells, and arcing loads.

V. FEATURE EXTRACTION

The necessary features of the collected data are extracted by wavelet transform analysis in a MATLAB environment.

Five categories of PQ events—such as Induction motor sag, arc load sag, sag from welding machine short circuit, swell from Induction motor switching off, and healthy condition—as well as their key characteristics are taken into consideration to illustrate the wavelet technique's feature extraction capacity. To execute successive signal decomposition utilizing the MRA technique for the reconstruction and decomposition of the observed signal, WT uses two functions: the scaling function ϕ and the wavelet function ψ. The distinct features of each instance are displayed in Figures 2 through 6. Wavelet function ψ functions as a A high pass filter generates a high-frequency component referred to as the detailed function (d), while the scaling function¯convolves the signal with a low pass filter , resulting in a low-frequency component known as the approximate function (a) in the decomposed signal.

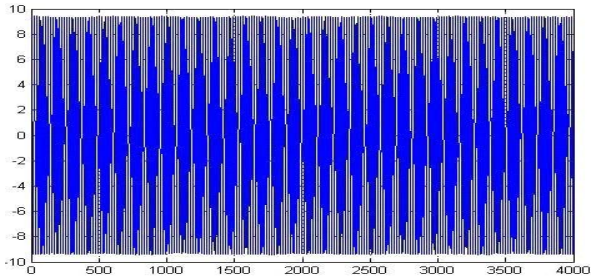


Figure 2: Normal condition

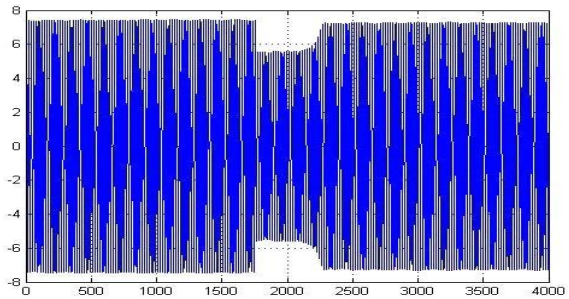


Figure 3: Sag due to Induction Motor Start

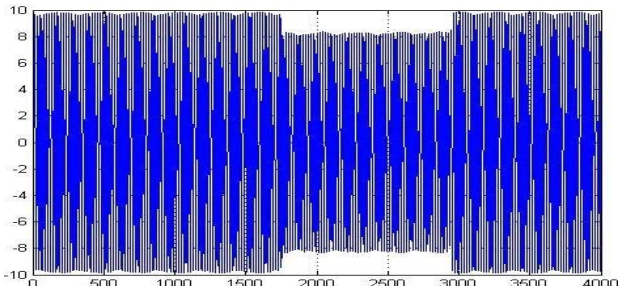


Figure 4: Sag due to welding machine short circuited

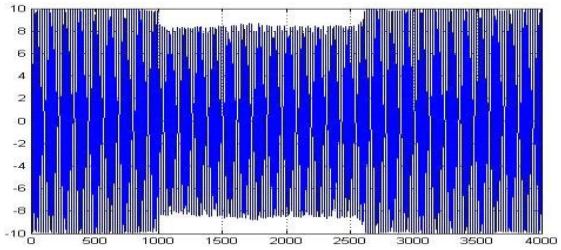


Figure 5: Sag due to arc load of welding machine

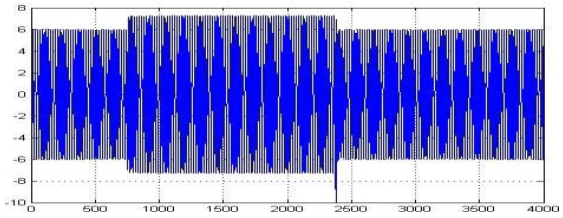


Figure 6: Swell due to Induction Motor switching off

Different power quality disturbances are classified based on statistical characteristics of waveforms. Specifically, "sample" statistics will be derived from the available data, including the kurtosis coefficient, maximum and minimum values of the skewness coefficient, and the root mean square (RMS) of zero-mean signals (standard deviation). Additionally, Pearson's coefficient of skewness is considered, as defined by:

$$g_2 = \frac{3\left(\bar{x} - \tilde{x}\right)}{S_x} \tag{7}$$

$\bar{x}$, $\tilde{x}$ and S_x is sample mean, sample median, and standard deviation, in that order.

The sample coefficient of variation v_x is defined by;

$$v_x = \frac{S_x}{\bar{x}} \tag{8}$$

Data set for the r^{th} sample moment about the sample mean is:

$$m_r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^r}{n} \tag{9}$$

Spread about the center is shown by m2, skewness about the center is indicated by m3, and the amount of data massed at the center is indicated by m4. Using the second, third, and fourth moments, the sample coefficients of skewness and kurtosis are defined directly.

, respectively. $g_3 = \frac{m_3}{\left(\sqrt{m_2}\right)^3}$ (10)

$$g_4 = \frac{m_4}{\left(\sqrt{m_2}\right)^4} \tag{11}$$

The following is the covariance between the dimensions j and k sample:

$$c_{jk} = \frac{\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)\left(x_{ik} - \overline{x_k}\right)}{(n-1)} \tag{12}$$

For dimensions *j* and *k*, r_{jk} the ordinary correlation coefficient is given as;

$$r_{jk} = \frac{c_{jk}}{S_j - S_k} \tag{13}$$

VI. DISTURBANCE CLASSIFIER USING ANN

1. **Multi Layer Perceptron:** The proposed classifier for disturbances is the Multilayer Perceptron (MLP) Neural Network. The input layer comprises 14 input parameters, while the output layer consists of six Processing Elements (PE's) corresponding to six critical conditions. The data processing and analysis are conducted using Neuro Solution 5.7, XLSTAT-2010, and MATLAB 7.1.

   The general learning algorithm employed for training and classification is described as follows:

   Initialization of Weights:

   - First, set the weights to tiny, random numbers.
   - Step 2: Perform steps 3–10 if the stopping condition is false.
   - Step 3: Complete steps 4–9 for every training pair.

2. **Feed Forward:**

   - Step 4: After receiving the input signal xi, each input unit sends it to every unit in the hidden layer.
   - Step 5: The weighted input signals of each hidden unit (j=1,…,p) are added together.

$$z_{-inj} = v_{oj} + \sum_{i=1}^{n} x_i v_{ij} \tag{14}$$

Applying the activation function $Z_j = f(z_{inj})$ here the activation function is $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$ and sends this signal to all units in output units.

- Step 6: Each output unit (y_k , *k=1,…,m*) sums its weighted input signals ,

$$y_{-ink} = w_{ok} + \sum_{j=1}^{p} z_i w_{jk} \quad (15)$$

And apply its activation function to calculate the output signals $Y_k = f\left(y_{-ink}\right)$ here the activation function is

$$\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x}) \quad (16)$$

3. Back Propagation Error:

- Step 7: The goal pattern for each output unit (, k=1,...,m) is determined by computing the error information term.

$$\delta_k = \left(t_k - y_k\right) f\left(y_{-ink}\right)$$

- Step 8: Each hidden unit (z_j ,j=1,…,p) sums its delta inputs from units in the layer above

$$\delta_{-inj} = \sum_{k=1}^{m} \delta_i w_{jk} \quad (17)$$

The error information term is calculated as

$$\delta_j = \delta_{-inj} f\left(z_{-inj}\right) \quad (18)$$

4. Updation of Weight and Biases:

- Step 9: Each output unit ($y_{k,}$ *k=1,…,m*) updates its bias and weights *(j=0,…,p)*

$$w_{jk}\left(t+1\right) = w_{jk}(t) + \alpha\delta_k z_j + \mu\left[w_{jk}(t) - w_{jk}(t-1)\right] \quad (19)$$

Where α is learning rate and μ is momentum factor

And each hidden unit *(z_j , j=1,…,p)* updates its bias and weights *(i=0,…,n)*

$$v_{jk}\left(t+1\right) = v(t) + \alpha\delta_j x_i + \mu\left[v_{ij}(t) - v_{ij}(t-1)\right] \quad (20)$$

- Step 10: Test the stopping condition

The neural network is trained with randomized data, and this process is repeated five times using various random weight initializations to eliminate bias. Authentic learning and generalization are encouraged by this method, especially for various hidden layers. Figure 7 illustrates the study's findings that the step learning rule and the tanh-axon transfer function produce better results. The number of Processing Elements (PEs) in the concealed layer is varied during the experimentation. Figure 8 shows that 11 PEs in the hidden layer yields the optimum network performance.

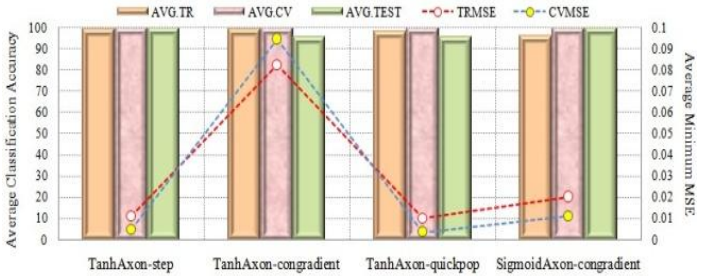
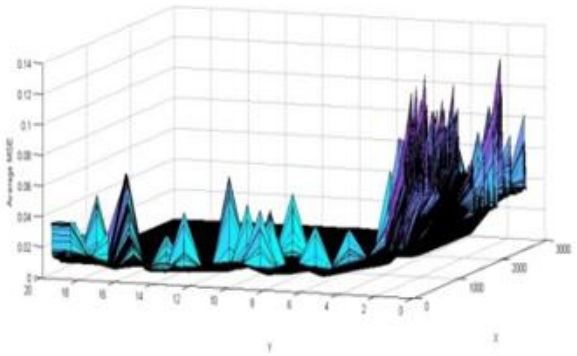


Figure 7: shows how average MSE and classification accuracy vary depending on the learning algorithm and transfer function.



Y: No. of Epoch X: No. of PE's

Figure 8: shows how the number of processing elements in the hidden layer affects the average MSE during training.

By comparing the average minimum MSE, the hidden layer and output layer parameters, or step size, are chosen. The performance is displayed in Figures 9 and 10, where the ideal step size values for the hidden layer and output layer are, respectively, 0.8 and 0.9.

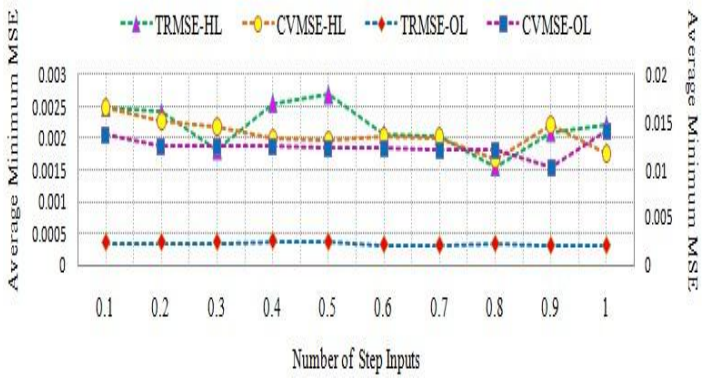


Figure 9: Average minimum MSE variation on TR and CV datasets with step sizes for the hidden and output layers

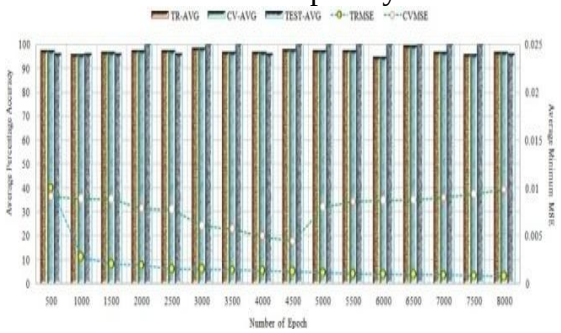


Figure 10: Average MSE Variation for Training and CV with Average Classification Accuracy for the Total Number of Epoches

VII. SELECTION OF ERROR CRITERION

Supervised learning necessitates the use of a metric for assessing the network's performance. Error criteria members are compared to an expected response, and any discrepancies are reported to the appropriate learning procedure. Gradient descent learning is used to identify the proper metric to assess sensitivity. The cost function, represented by J, should ideally approach zero as the network converges to the desired response, although it usually stays positive. Various cost functions have been presented in the literature, where "p" is defined as p = 1, 2, 3, 4, and so on, and the criterion is represented as L-1, L-2, L-3, L-4, and so forth. The cost function is employed to define components within the Error Criteria family.

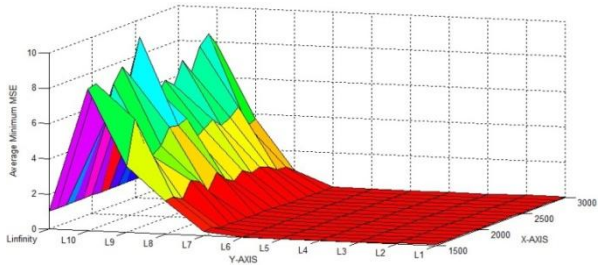
Error Criteria family components are defined by a cost function in the form:

$$J(t) = \frac{1}{2}\sum_{i}(d_i(t) - y_i(t))^p \tag{21}$$

and error functions:

$$e_i(t) = -(d_i(t) - y_i(t)) \tag{22}$$

Since the expected response and the network's output are d (t) and y (t), respectively, a number of error criteria have been tried in order to determine which one is right. In the end, the L-2 criterion, which is displayed in Fig. 11, produces the best results.



Y-AXIS: Error Criterion X-AXIS: No. of Epochs

Figure 11: Variation in Average MSE with Error Criterion

From the above experimentation, selected parameters for MLP NN are given below.

Transfer: TanhAxon Learning Rule: Step
Number of Inputs: 14
Number of Hidden Layers: 01
Number of PEs in Hidden Layer: 11

Step Learning Rule: Hidden Layer: Transfer function: tanh
Step size: 0.8 Transfer function: tanh Output Layer: Learning Rule: step
Size of step: 0.9
There are 4500 epochs in total.
237 Training Exemplars = 70%, Cross Validation Exemplars = 15% are the number of connection weights.

15% of testing exemplars
Time required per epoch per exemplar: 0.1822 micro –secs.

The proposed neural network (NN) undergoes training on various datasets and is subsequently subjected to careful validation to ensure that its performance remains independent of the specific data partitioning scheme used. The NN's performance should consistently demonstrate optimality across all datasets, as assessed through mean square error (MSE) and classification accuracy. The designed multilayer perceptron (MLP) undergoes five rounds of training with different random weight initialization and is tested on separate testing, cross-validation (CV), and training datasets.

For both training and testing, two different ways to data tagging are used: "data tagging by percent" and "data tagging by various groups." The Leave-N-Out technique is also applied. A useful method for evaluating the model's generalization skills is leave-n-out training, which is especially helpful with tiny datasets. With this approach, the full dataset can be used for training and testing. The network is trained iteratively by the algorithm,

which removes a different subset of the data each time for testing. The outputs from each tested subset are then combined into a thorough testing report. The model then undergoes one more training round with all of the data, and the weights from this last training run can be applied to additional testing.

The entire dataset is split into four groups in order to assess the accuracy of the classification and learning capacity. First and second groups (i.e., 50% of the data) are the training data; third and fourth groups (i.e., 25% of the data each) are set aside for cross-validation and testing (e.g., 1234: 1, 2-TR, 3-CV, 4-Test). For every group, a total of twenty-four combinations are prepared, and the network is trained and evaluated. The variance in the percentage of data tagged for each error criteria and the average minimum MSE are evaluated, with the L2 error criteria showing the lowest MSE and the best classification accuracy. Figures 12, 13, and 14 give the full set of results.

Figure 12: Average minimal MSE variation with test on Training and Testing dataset andpercentage of data marked for training

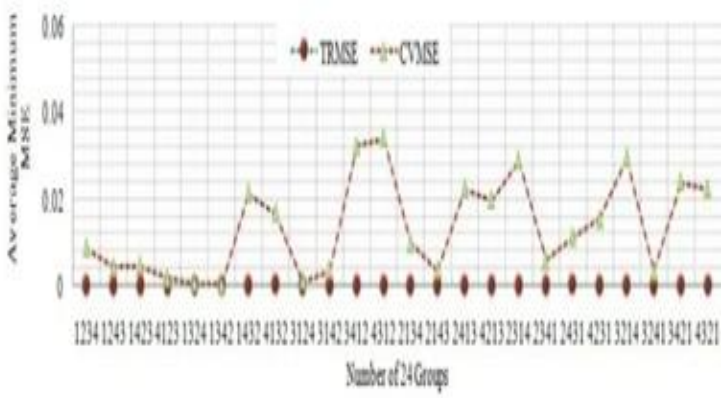


Figure 13: Changes in the average minimum MSE for each of the 24 dataset groups with training and CV

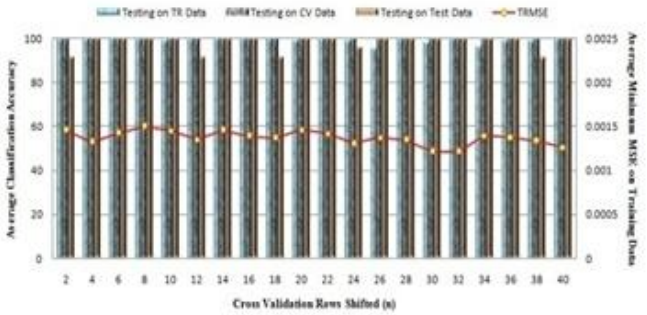


Figure 14: Average MSE Variation with Test on Training, CV, and Testing Datasets with Shifted CV Rows (n)

The final improved MLP-based classifier yields a 99.18% classification accuracy with an average training MSE of 0.00083773 and an average cross-validation MSE of 0.025529.

VIII. RADIAL BASIS FUNCTION

Radial basis function (RBF) networks are hybrid networks with a nonlinear structure, typically composed of a single hidden layer containing processing elements (PEs). Unlike conventional sigmoidal functions, Gaussian transfer functions are employed in this layer. Unsupervised learning rules are applied to determine the centers and widths of these Gaussians, while supervised learning is implemented in the output layer. It is important to note that RBF networks generally demonstrate faster learning rates compared to multilayer perceptrons (MLPs). The application of this network is recommended, especially in situations where the number of exemplars is extremely limited or widely dispersed, leading to inherent ambiguity in clustering.

In the case of standard radial basis function networks (RBFs), the supervised component of the network is primarily tasked with generating a linear combination of the output derived from the unsupervised layer. Consequently, the default configuration involves having no hidden layersThe supervised segment can be converted from a simple linear perceptron to a multilayer perceptron (MLP) by adding more hidden layers. The six conditions of PQ disturbances that the network is intended to classify are: Arc load, Welding machine short circuit, Swell owing to Induction Motor, Sag due to Induction Motor, and Line to Ground.

1. **The general learning algorithm is as follows:** The response of the Gaussian activation function is nonnegative for all values of x. We define the function as

$$f(x) = \exp(-x^2) \tag{23}$$

and its derivative

$$f'(x) = -2x\exp(-x^2) = -2f(x) \tag{24}$$

The radial basis function and the back propagation network in the Gaussian function are not the same.

Step 1: Initialize the weights (set to small values)
Step 2: do step 3-10, if stopping condition is false
Step 3: do step 4-9 for each input
Step4: all units in the hidden layer receives input signals from each input unit (x_i , $i=1,\ldots, n$)
Step 5: Calculate the radial basis function
Step 6: Choose the centers of the radial basis functions. The centers are chosen using the collection of input vectors. A appropriate sample of the vector space input has been ensured by selecting a sufficient number of centers.
Step 7: The output of i_m unit $v_i(x_i)$ in the hidden layer

$$v_i(x_i) = e\left\{\sum_{j=1}^{r}\frac{\left[x_{ji}-(x'_{ji})^2\right]}{\sigma_i^2}\right\} \qquad (25)$$

Where x_{ji} - centre of the RBF unit for input variable
x'_{ji} - j^{th} variable of input pattern
σ_i -Width of the i^{th} RBF unit

Step 8: Initialize the weights in the output layer of the network to some small random value
Step 9: output calculation of the neural network

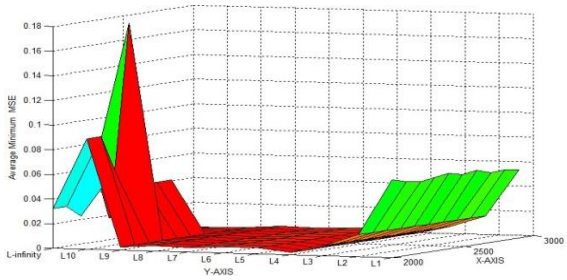
$$y_{net} = \sum_{i=1}^{H} w_{im} v_i(x_i) + w_0 \qquad (26)$$

Where
H-number of hidden layer nodes (RBF Function)
y_{net} -Output value of m^{th} node in output layer for the n^{th} incoming pattern
w_{im} - Weight between i^{th} RBF unit and m^{th} output node
w_0 - Biasing term at n^{th} output node

Step 10: Calculate the error and check the stopping condition

To ensure genuine learning and generalization across various parameters, biasing is removed, and the neural network is exposed to randomized data that undergoes retraining five times with distinct random weight initializations.

2. **Selection of Error Criterion:** Analogous trials are conducted to determine the appropriate criterion for error. Fig. 15 displays variations in average minimum MSE based on various error criteria.
   Finally L-4 criterion gives the optimal results.



Y: Error Criterion X: No. of Epochs

Figure 15: Variation in Average MSE with Error Criterion

The supervised and unsupervised parameters, including the cluster centers, metric, and competitive rule, have been optimized through testing. Results of the experiments are displayed in Figs. 16 through 19.

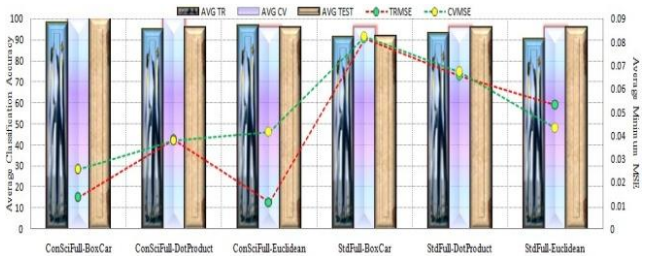
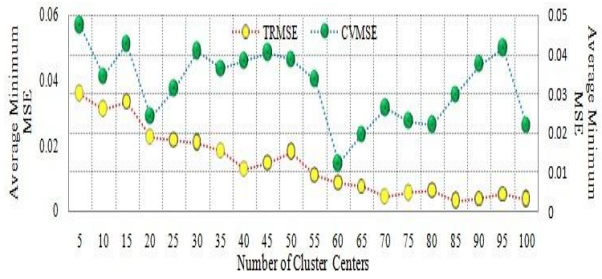


Figure 16: Variation in Average Classification Accuracy with Competative rule and metric



Variation in Average Minimum MSE for Various Numbers of Cluster Centers (Fig. 17)

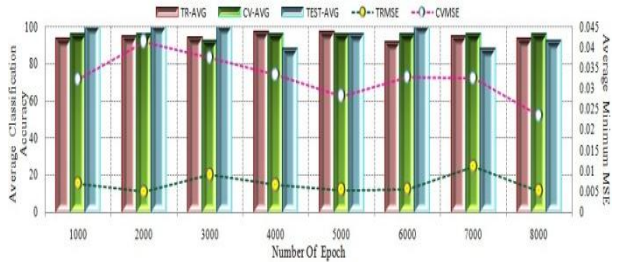


Figure 18: Average Variation Arrangement Precision and Mean lowest MSE using Epoch

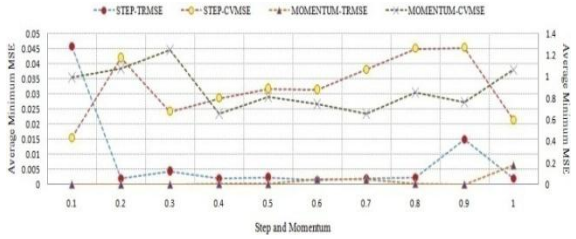


Figure 19: Variation in Average Minimum MSE for Step and Momentum with Training and CV

Lastly, the following parameters are used in the creation of the RBF classifier:

There are 14 inputs.

Competitive Regulation: All-Inclusive Measurement: Boxcar
There are no hidden layers. The output layer is
Learning Rule: 0.1 Step Size, Momentum Momentum: 0.7 Tanh the transfer function 5000 epochs in total
Unmonitored Education: Uppermost Epoch: 1500
Max Epoch for Supervised Learning: 3500
Learning Rate: beginning at 0.01 and decreasing to 0.001
There are sixty cluster centers.
There are 1266 connection weights.
70% of the examples are used for training, while 15% are used for cross-validation.
15% of the exemplars for testing Time required for training per exemplar is 0.604 micro-secs RBF classifiers are trained and tested on similar data sets, or groups of data. Figures 20 through 22 display the results.

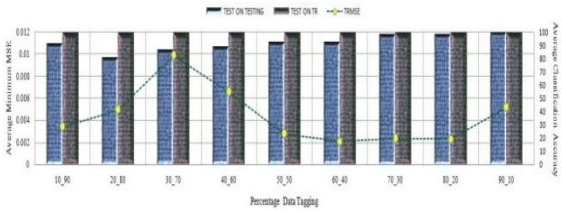


Figure 20: Average Variation Arrangement Percentage of data tagged for training and accuracy of testing on the Testing and Training dataset

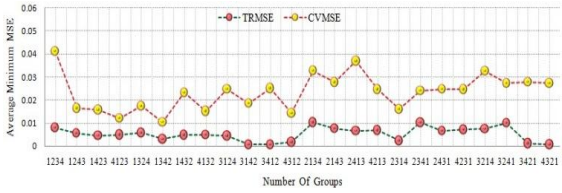


Figure 21: shows the average Minimum MSE variation with training and CV for each of the 24 dataset groups.

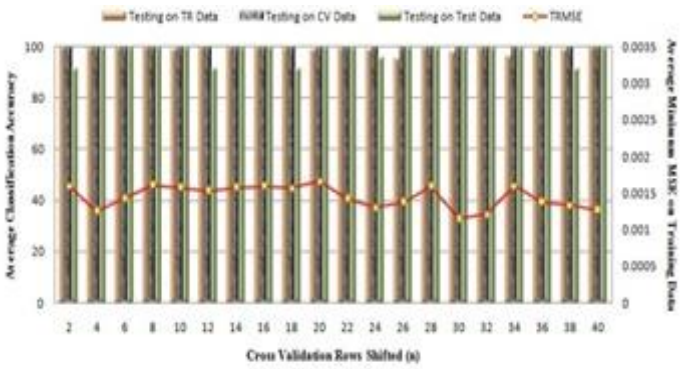


Figure 22: Average MSE Variation with Test on Training, CV, and Testing Datasets with Shifted CV Rows (n)

After optimization, the RBF-based classifier yielded a 98.05% classification accuracy, with an average cross-validation mean square error of 0.0232185 and an average training mean of 0.00519002.

IX. DIMENSIONALITY REDUCTION OF CLASSIFIERS

1. **Sensitivity Analysis:** Dealing with an excessive quantity of input features is a significant difficulty after feature extraction, which can possibly impair accuracy and result in considerable processing demands. In order to solve this problem, the most sensitive values are carefully analyzed and chosen to be input parameters. The selection of inputs is driven by the analysis of mean squared error (MSE) and the percentage of classification accuracy. These selected parameters are employed in the multilayer perceptron (MLP) classifier. Fig. 23 shows graphically the differences in MSE and classification accuracy percentage with various sets of the most sensitive inputs. The findings clearly show that only the most sensitive criteria produce the best results. After nine inputs, the classifier is tested and fine-tuned, and the complete results are shown in detail in Figures 24 to 27.

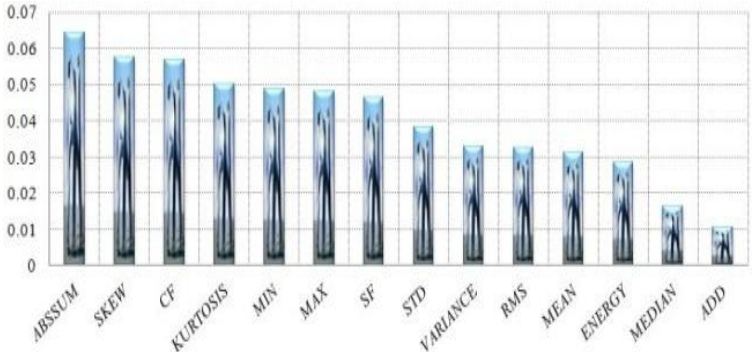
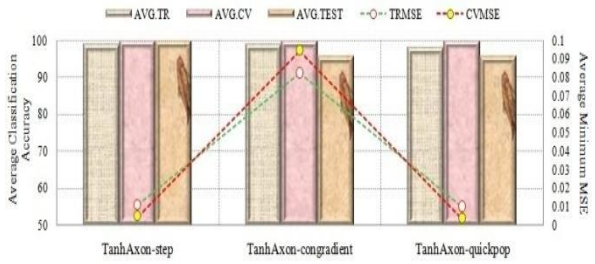


Figure 23: Sensitivity Analysis about Mean



Variation of classification accuracy and average MSE for learning rule and transfer function .

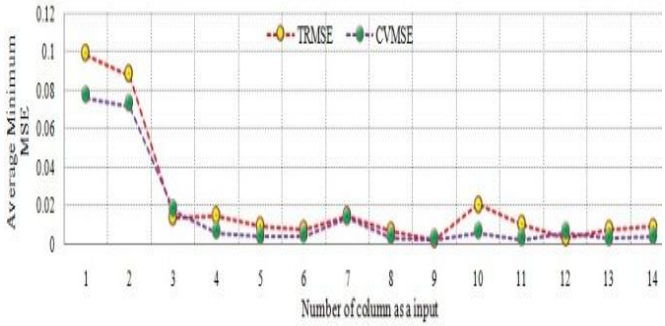


Figure 24:
Item 25. Average minimal MSE variation with number of inputs on training and CV datasets Choosing an Error Criterion

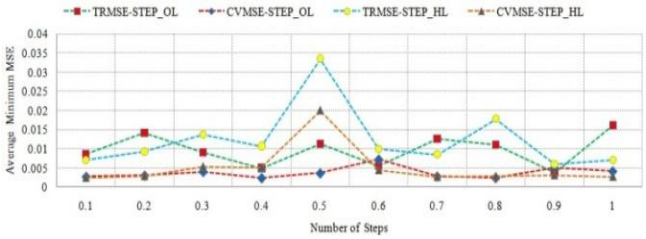
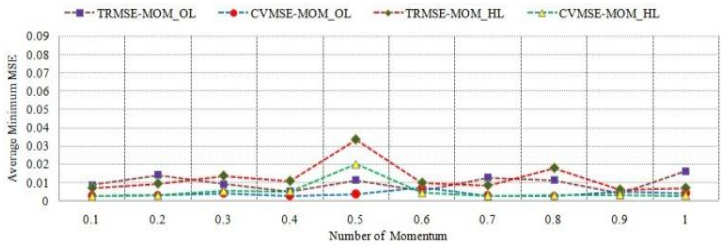


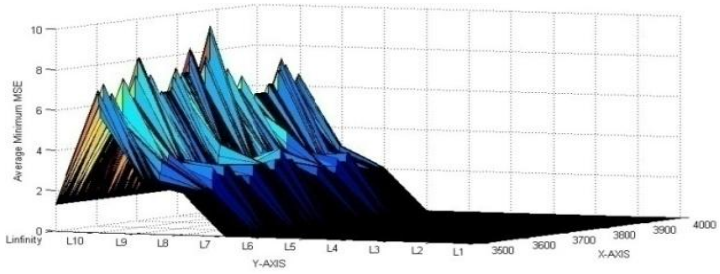
Figure 26: Average minimal MSE variation on TR and CV datasets with step sizes for the hidden and output layers



Variation of average minimum MSE on TR, CV dataset with output layer and hidden layer momentum Figure: 27

The L-2 criterion was selected, as shown in Fig. 28, after a number of error

criteria were tested to find the optimal error criterion for the expected response.



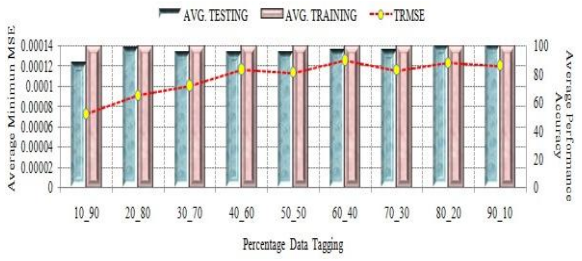
Y: Error Criterion X: No. of Epochs

Figure 28: Variation in Average MSE with Error Criterion

observing the same guidelines The NN model is ultimately optimized by a series of tests, and the resulting network is constructed using the following parameters:

There are nine inputs.
There are one hidden layer.
There are nine PEs in the hidden layer.
Secret Layer:Function of transfer: tanhLearning Guideline: velocityStep magnitude: 0.9; momentum: 0.8; result Level:
Function of transfer: tanhRule of Learning: stepMomentum: 0.7; step size: 0.4
Epochal Number: 4500
Failed Standard: L3
150 connection weights in total
Time needed for each exemplar in an epoch: 0.4077 µ-sec
Total effectiveness of training: 99.81897%
TRMSE on average: 0.000137

Ultimately, Figs. 29, 30, and 31 display the outcomes of training and testing the new NN (MLP-DR-S) classifier under the specified conditions.



Item 29. Variation of the % of data tagged for training and the average minimal MSE with test on the Testing and Training dataset

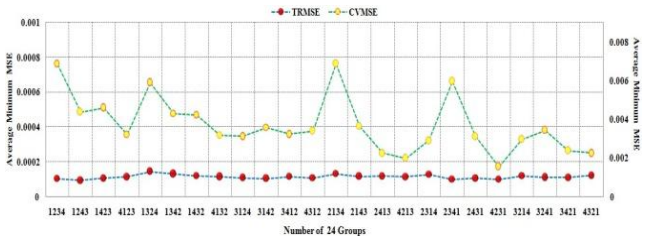


Figure 30: shows the average minimal MSE variation with training and CV for each of the 24 dataset groups.

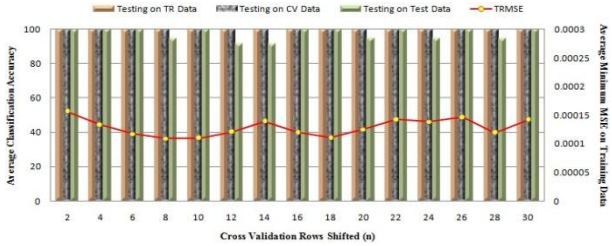
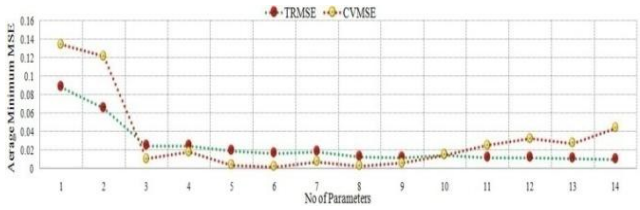
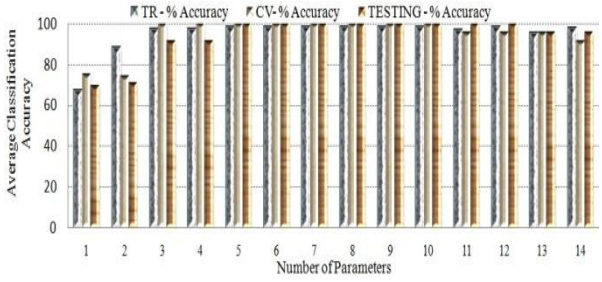


Figure 31: Average MSE Variation with Test on Training, CV, and Testing Datasets with Shifted CV Rows (n)

To determine the appropriate analysis and pinpoint the most responsive factors for utilization as input parameters, a parallel testing methodology is employed for the RBF-based classifier. Illustrated in Fig. 32, the quantity of input parameters, organized in descending order according to sensitivity, is juxtaposed with the average minimal mean squared error (MSE) on both training and cross-validation datasets. Additionally, Fig. 33 provides insight into the average categorization



No. 32. Average minimal MSE variation with number of inputs on training and CV datasets



Item 33. Variation in the average classification accuracy with test on the number of inputs, CV, and TR data

The results of the sensitivity analysis can be used to reduce the RBF network's dimensions. The remaining inputs are down to eight. Similar trials are used to build the optimal RBF classifier, which is modified as follows:

Competitive Guideline: Morality Complete Measure: Euclidean
45 cluster centers are present.
There are eight inputs.
There are 681 connection weights.
46.20% less connection weights overall
There are eight inputs.
Competitive Guideline: Morality Complete Measure: Euclidean
There are no hidden layers. The output layer is
Function of transfer: tanhLearning Guideline: velocityDecay to: 0.001 Step size: 0.8 Momentum: 0.6 Learning Rate Start at: 0.01
There are 45(80) cluster centers.
Error Standard: L7
Period: 5000
Time needed in microseconds for each exemplar in an epoch.
TR: 97.87%, CV: 97.17%, TEST: 96.80%, overall efficiency
All-around Average Error % T: 0.0012893
As a Whole Averages 0.01130 for CVMSE
RBF-DR-S (Dimensionally Reduced utilizing Sensitivity Analysis) takes 0.2683 microseconds to complete.

Lastly, a variety of circumstances are tested and trained on the new RBF (RBF-DR-S) classifier. The findings are displayed in Figures 34, 35, and 36.

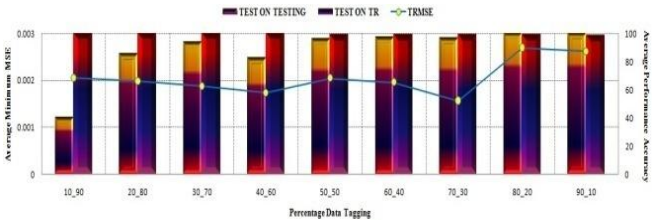


Figure 34: Average minimum MSE variation with test on Training and Testing dataset and percentage of data marked for training

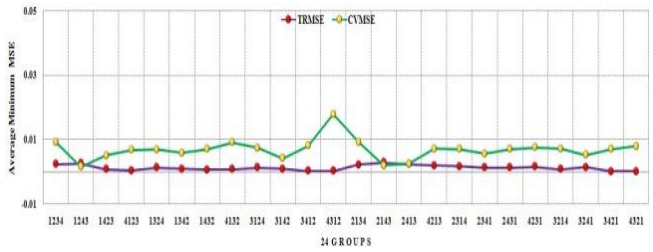


Figure 35: Variation of average minimum MSE with Training and CV with all 24 group of dataset

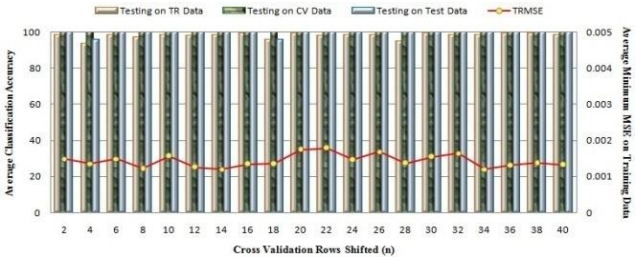


Figure 36: Variation of average MSE with Test on Training, CV and Testing dataset with CV rows shifted (n)

2. **Principal Component Analysis:** Principal Component Analysis (PCA) is another technique used for dimensionality reduction. It successfully lowers the dimensionality of the input space, resulting in a more efficient network. Pearson's rule is followed in the PCA procedure. The mathematical concept of eigenvalues—which function as a gauge of the accuracy of the projection from a 13-dimensional space to a lower-dimensional one—is the subject of attention in Fig. 37. Fig. 38 sheds light on the correlation circle by showing that, because of its closeness to the circle, "skew" is the only parameter that retains information throughout several axes.

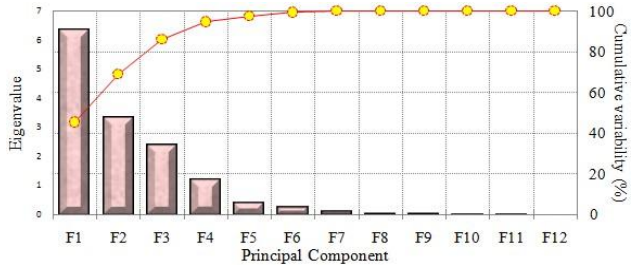


Figure 37: Eigen values, % variability, and principal component

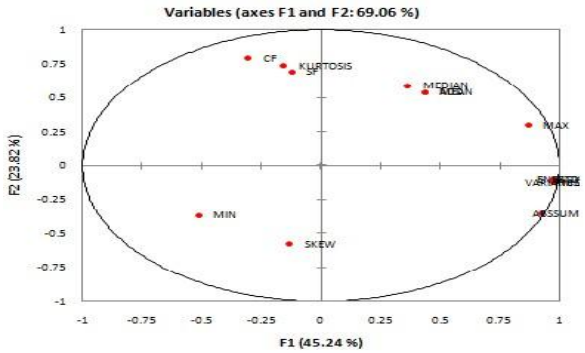


Figure 38: Correlation Circle

By varying the number of primary components used as input, we evaluated the average minimal MSE and average classification accuracy to establish the ideal number of inputs for the MLP. The analysis's findings are displayed in Figs. 39, 40, and 41, which show how the average MSE varies depending on the error criterion and the quantity of input processing elements used.

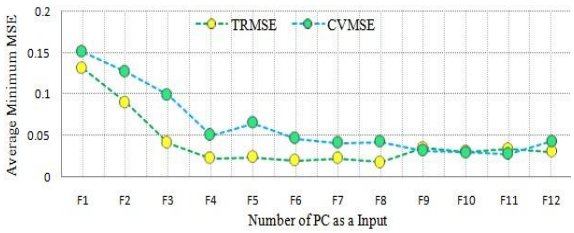
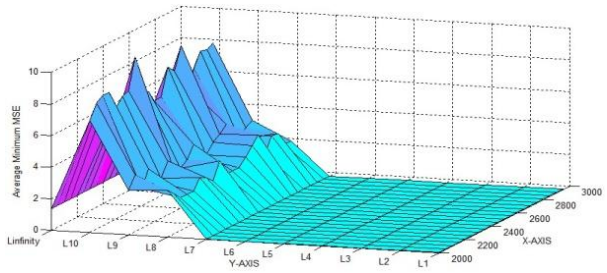


Figure 39: Variation of average minimum MSE on training and CV dataset with number of PCs as inputs



Y: Error Criterion X: No. of Epochs

Figure 40: Variation in Average MSE with Error Criterion

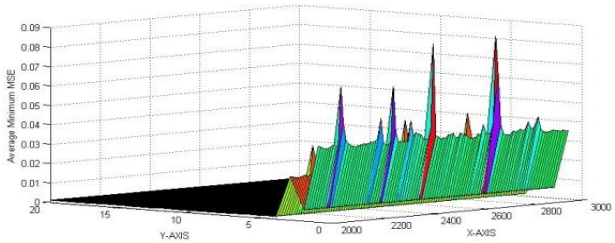


Figure 41: Variation of average classification accuracy with test on Testing, CV and TR Data with number of PEs as Inputs

The dimensionally reduced MLP (MLP DR PCA) NN model based on hidden layer and output layer parameter modifications is constructed using the following parameters.

There are seven inputs.
There are one hidden layer.
There are 14 PEs in the Hidden Layer.
Secret Layer:
Function of transfer: tanh Learning Guideline: Velocity
Measurement scale: 0.8 Acceleration: 0.8 Output Level:
Function of transfer: tanh Learning Guideline: Velocity
Error: 0.3 Momentum: 0.4 Step size Standards: L5.
Period: 4000
Time needed for each exemplar per epoch: 0.2505μ seconds
There are 202 connecting weights.

Weight loss as a percentage: 17.32%

Lastly, as illustrated in Figs. 42, 43, and 44, the new MLP (MLP-DR-PCA) classifier is trained and tested under a variety of situations.

Changes in the average minimum MSE with testing and training dataset tests and percentage of data tagged for training are shown in Figure 42.

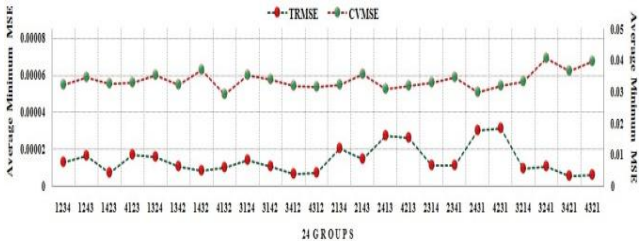


Figure 43: shows the average minimal MSE variation with training and CV for each of the 24 dataset groups.

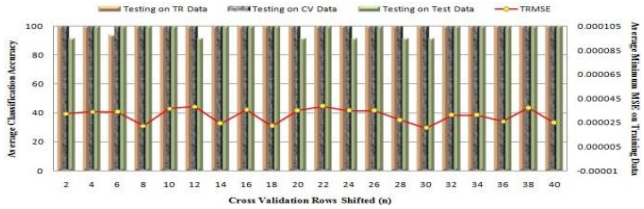


Figure 44: Average MSE Variation with Test on Training, CV, and Testing Datasets with Shifted CV Rows (n)

By using the same procedure as previously, tests are conducted to develop the RBF classifier with the fewest dimensions. The ideal RBF classifier with seven inputs is the outcome, as shown in Fig. 45.

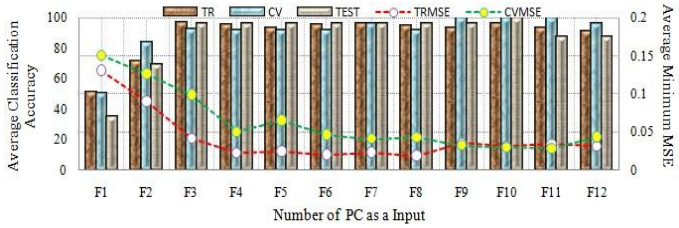
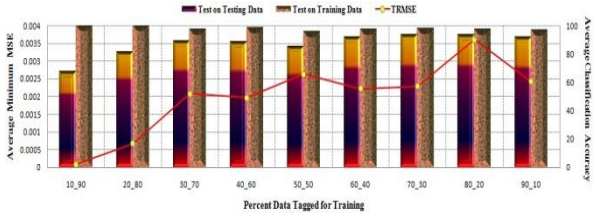


Figure 45: Variation of average classification accuracy with test on Testing, CV and TR Data with number of PCs as inputs

There are seven inputs.
Rivalries Guidelines: Normal Complete Meter: Boxcar
There are no hidden layers. The output layer is
Function of transfer: tanhRule of Learning: step
45 cluster centers are present.
There are seven inputs.
Layer of Output: Step Size: 0.5 Learning Rate: Initial at 0.01; Decays to 0.001
There are 636 connection weights.
49.76% less connection weights were used.
TRMSE on average: 0.0012893 CVMSE on average: 0.01130
RBF-DR-PCA (Dimensionally Reduced using PCA) takes 0.4600 microseconds to complete.

Lastly, as illustrated in Figs. 46, 47, and 48, the new RBF (RBF-DR-PCA) classifier is trained and assessed under a variety of situations.



Item 46. Variation of the % of data tagged for training and the average minimal MSE with test on the Testing and Training dataset

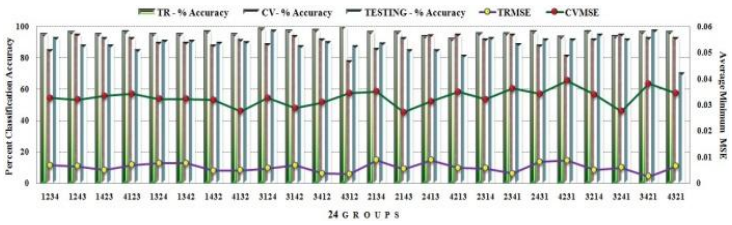


Figure 47: shows the average minimal MSE variation with training and CV with percent accuracy for each of the 24 dataset groups.

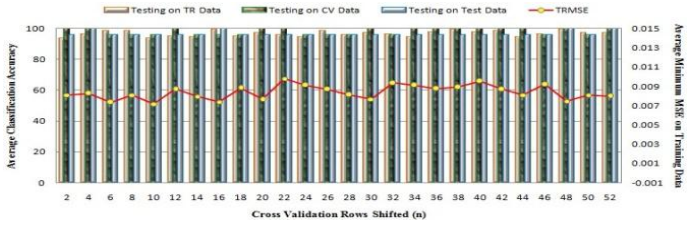


Figure 48: Average MSE Variation with Test on Training, CV, and Testing Datasets with Shifted CV Rows (n)

X. RESULTS AND DISCUSSION

The outcomes of the RBF-based classifier and the optimized MLP classifier for the classification of PQ disturbances are evaluated in this research. With an average MSE of 0.0008, the optimized MLP classifier obtained 99.18% classification accuracy, whereas the RBF-based classifier achieved 98.05% classification accuracy with an average MSE of 0.0051. These findings clearly show that the MLP-based classifier is a better fit for handling this issue.

Two dimensionality reduction techniques—Principal Component Analysis and Sensitivity Analysis—are used to improve performance and lower complexity. Classifiers are redesigned and optimized for better performance using these strategies. It is clear that the more efficient method for reducing dimensionality is Sensitivity Analysis. Sensitivity analysis reduces connection weights by 37.7% while increasing the MLP-based classifier's percent classification accuracy from 99.18% to 99.81%. Nevertheless, this method is not appropriate for the RBF-based classifier since it only marginally affects classification accuracy while reducing dimensions by 49.76%.

The dimension reduction classifier based on MLP, utilizing sensitivity analysis, serves as an effective diagnostic tool for identifying common power quality disturbances. The average MSE for the samples remains consistently reasonable, around 0.0015185. Moreover, the average classification accuracy for both training and cross-validation instances is a respectable 98.05% and 96.92%, respectively, demonstrating sound classification performance. Comparative results are presented in Table I for reference.

REFERENCES

[1] IEEE Std. 1366. IEEE standard guide for power distribution reliability (2003).
[2] IEEE 1159.3. IEEE recommended practice for the transfer of power quality data
[3] Hamid EY, Kawasaki ZI. Wavelet-based data compression of power system disturbances using the minimum description length criterion. IEEE Trans Power Deliv 2002;17(2):460–6
[4] B. Hannaford and S. Lehman, “Short Time Fourier Analysis of the Electromyogram: Fast moments and constant contraction,”IEEE Transaction on Biomedical Engg. Vol.33, pp.1173-1181,1986
[5] W.A. Wilkinson, M.D. Cox “Discrete wavelet analysis of power system transients,” IEEE Transaction on Power System 11 (1996) 2038–2044
[6] Dash PK, Panigrahi BK, Sahoo DK, Panda G. Power quality disturbance data compression, detection, and classification using integrated spline wavelet and S- transform. IEEE Trans Power Deliv 2003;18(2):595–600.
[7] Poisson O, Rioual P, Meunier M. New signal processing tools applied to power quality analysis. IEEE Trans Power Deliv. 1999;14(2):561–6.
[8] Fanibhushan Sharma, A. K. Sharma, Ajay Sharma, Nirmala Sharma “Recent Power Quality Techniques A Comparative Study”, Canadian Journal on Electrical & Electronics Engineering Vol. 1, No. 6, October 2010.
[9] IEEE Recommended Practice for Monitoring Electric Power Quality, IEEE Inc.,New York, USA, 1995.
[10] Bracewell, R. N. (2000). The Fourier Transform and its Applications. McGraw- Hill Book Co, Singapore.
[11] Qian S., Chen D., “Understanding the nature of signals whose power spectra change with time-joint analysis”, IEEE Signal Processing Magazine, 1999.
[12] Qian S., Chen D., “Discrete Gabor Transform”, IEEE transaction signal processing, Vol. 41(7), pp. 2429-2439.
[13] S. Santoso, E. J. Powers, W. Mack Grady, and P. Hofmann, “Power quality assessment via wavelet transform analysis,” IEEE Transaction on Power Delivery, vol. 11, no. 2, pp. 924–930, April 1996.

[14] D. C. Robertson, O. I Camps, J. S. Mayer, andW. B. Gish, "Wavelet and electromagnetic power system transients," IEEE Transaction on Power Delivery, vol. 11, no. 2, pp. 1050–1058, April 1996.
[15] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 674–693.
[16] Robert D. Nowak and Richard G. Baraniuk, "Wavelet-Based Transformations for Nonlinear Signal Processing", IEEE transactions on Signal Processing, 47(7) (1999) 1852–1865.
[17] L. Agrisani, P. Daponte, M.D. Apuzzo, and A. Testa, "A measurement method based on the wavelet transform for power quality analysis", IEEE Trans. Power Delivery, vol. 13, 990-998, Aug 1998.
[18] G.T. Heydt, A.W. Galli, Transient power quality problems analyzed using wavelets, IEEE Trans. Power Deliv. 12 (1997) 908–915.
[19] Huseyin Eris_ti, Ozal Yıldırım, Belkıs Eris_ti, Yakup Demir, "Automatic recognition system of underlying causes of power quality disturbances based on S-Transform and Extreme Learning Machine," Electrical Power and Energy Systems 61, 2014 553–562
[20] Wael R., Anis Ibrahim and Medhat M. Morcos, "Artificial Intelligence and Advanced Mathematical Tools for Power Quality Applications: A Survey", IEEE transactions on power delivery, VOL. 17(2), pp. 668-673, APRIL 2002.
[21] M. Kezunovic and I. Rikalo, "Automating the analysis of faults and power quality," IEEE Comput. Appl. Power, vol. 12, pp. 46–50, 1999.
[22] M. Faisal, A. Mohamed, H. Shareef, A. Hussain, "Power Quality diagnosis using time frequency analysis and rule based techniques ", Expert System with Applications, vol. 38, pp. 12592-12598, 2011.
[23] B. Biswal, M. Biswal, S. Mishra, R. Jalaja, "Automatic Classification of Power Quality Events Using Balanced Neural Tree," IEEE Transaction On Industrial Electronics, Vol. 61, No. 1, January 2014, 521B
[24] Elango M.K., Kumar A.N., Duraiswamy K, "Identification of powerQuality disturbances using Artificial Neural Networks", IEEEConference on Power and Energy Systems, pp. 1 – 6 , 2011.
[25] Milan Biswal, P. K. Dash "Measurement and Classification of Simultaneous Power Signal Patterns With an S-Transform Variant and Fuzzy Decision Tree," IEEE Transaction On Industrial Informatics, Vol. 9, No. 4, November 2013, 1819
[26] Morsi W.G., El-Hawary M.E., "A new fuzzy-wavelet based representative quality power factor for stationary and nonstationary power quality disturbances", IEEE Conference, pp.1 – 7, 26-30 July 2009
[27] Sulaiman, M.S., Mohd Yasin, F.Kamada, M., Reaz , "Expert System for Power Quality Disturbance Classifier", Power Delivery, IEEE Transactions Vol 22, Issue 3,pp 1979 – 1988. July 2007.

Table 1: Effects on Average Classification Accuracy and Various Parameters When Classifier Tested on Unseen Data

Sr. No	Classifier	Average Training Mse	Average Cross Validation Mse	Percent Classification Accuracy When Tested On			No Of Connecting Weights	Percent Reduction Of Connecting Weights	Time Per Exemplars
				Traing Data	Cross-Valid-Ation Data	Test Data			
01	MLP	0.00083773	0.025529	99.18	97.62	96.22	237	- - -	0.2030 µ-secs
02	MLP-PCA	0.00003944	0.0346457	99.00	96.76	95.11	202	14.76	0.2505 µ-secs
03	MLP-S	0.000113194	0.01358683	99.81	98.63	96.74	150	36.70	0.4077 µ-secs
04	RBF	0.0051902	0.0233185	98.05	96.92	95.47	1266	- - -	0.604 µ-secs
05	RBF-PCA	0.00606888	0.03276	96.27	94.43	92.88	636	49.76	0.4600 µ-secs
06	RBF-S	0.0012893	0.01130	97.87	97.17	96.80	681	46.20	0.2683 µ-secs

