Voltage Event Signature Classification for Power Quality Disturbance Identification

Buchizya Kumwenda
Lecturer, School of Engineering, Electrical Department
The Copperbelt University
Kitwe, Zambia
buchizya.kumwenda@cbu.ac.zm, kumwenda.j.buchizya@gmail.com

Abstract

The power quality indices corresponding to voltage event are provided by IEC 61000-4-30 and IEEE 1564 standards. The popularization of machine learning techniques in modern and smarter power grids inherently characterized with randomized big data motivates their inclusion. We aimed to establish key voltage disturbance event features that a machine learning model can use to attain a high classification and prediction accuracy. The feature extraction was achieved using the singular value decomposition and wavelet transform of a three-phase system represented as a space phasor model with the corresponding ellipse parameters and the shape index feature extracted. The 6 by 500 dataset with the six (6) features (rms, event duration, 3rd harmonic, 5th harmonic, 7th harmonic, shape index) was used to classify voltage sag, swell, interruption, harmonics, and normal conditions. MATLAB R2021a classification learner application was used to create a supervised machine learning model by training twenty-nine (29) classifiers and comparing their performance using the accuracy (%), confusion matrix, total cost validation, prediction speed and training time. The results indicated that without the shape index, the voltage sag and voltage interruption events were misclassified in some scenarios for all the classifiers, with the highest accuracy obtainable of 99.6% by four (4) classifiers. The inclusion of shape index feature improved the trained models to 100% classification accuracy for fourteen (14) classifiers. The training, computing and processing is required to be of high performance and accuracy to increase the online situational awareness of network operators when such systems are implemented.

Keywords
Decomposition, Shape Index, Voltage Event and Supervised Learning.

1. Introduction
The growing interest in power quality issues due to the continues developments in the power grid structure, i.e. its generation composites that now include prosumers, and high integration of renewables, transmission and distribution modes are now bidirectional and the utilization of power has loads that are linear and nonlinear with demand response capability. It is the fact that power quality monitoring demands a large storage of data for analysis to instigate appropriate actions depending on the operational mode. A commercially available classification system must, therefore, be reliable and show high accuracy. The trained model must be able to understand all possible system conditions. This is important because once implemented, the user has is not supposed to redo the training process. The training, computing and processing is required to be of high performance and accuracy to increase the online situational awareness of network operators when such systems are implemented. This paper presents the classification of the common power quality events related to voltage deviations such as voltage dip, swell, interruption, and harmonics. These are expected to be experienced in modern systems having high renewable energy integration with converters, electric vehicles and other non-linear loads.

1.1 Objectives
- To determine the optimal sampling frequency using the relationship of the root mean square value in per unit of the un-disturbed red phase at various sampling frequencies.
- To generate synthetic training dataset from a power network analysis and modelling
- To extract key features in the three phase voltage system that uniquely describes each voltage event i.e. the equivalent voltage signature using the principal component analysis (PCA)
- To establish existence of patterns among identified features

© IEOM Society International

- Train a machine learning model using different classifiers and rank performance based on the accuracy, prediction speed and training time
- Investigate the performance of the classification system subjected with different voltage events

2. Literature Review

The modernization of power grids with a high integration of renewable power stations into the power network is a trend in power industry development aimed at reducing global warming through decarbonation (Kulikov et al. 2022). The use of power electronic devices and nonlinear loads may lead to critical deviations of power quality indicators from their normal operation. Early detection and time identification of power system disturbance events are essential for situational awareness and reliable operation of the power grid (Ren et al. 2020).

The IEEE standard 1100 defines power quality as the concept of powering and grounding sensitive (e.g. electronic) equipment in a manner that is suitable to the operation of that equipment. (McGranaghan and Santoso 2007) stated that power quality monitoring can be achieved offline at the central processing locations and online within the instruments. For rapid analysis and dissemination, online approaches provide indicators for supporting key decisions during abnormal conditions that may result from faulty conditions. However, for system evaluation, problem characterization and maintenance offline approaches tend to be sufficient.

Signal processing techniques are widely employed in the analysis of power quality events. The establishment of normal power quality performance of the power system provides a reference system that enables the utility to identify abnormal characteristics. The frequency, magnitude and root mean square (rms) deviation study includes formulations of events such as undervoltage, overvoltage, unbalance (for 3-phase systems), voltage variations (sags, swells, interruptions), flicker, transients or surges, harmonics (sub, super and inter-harmonics – total harmonic distortions (THD), total demand distortion (TDD)), rapid voltage change (RVC) and effective grounding and bonding, electromagnetic interference, electrostatic discharge and power factor (Bagheri et al. 2022). The features are analyzed on magnitude duration scatter plots. Statistical methods such as the minimum and maximum, mean, standard deviation, count and accumulative probability have been used. The type of voltage event can be analyzed by evaluating changes in magnitude and phase angle on the phases as a result of faults. The ABC classification of symmetrical and unsymmetrical fault events was consequently established. The single event characterization can be applied to establish the corresponding voltage event and its impact on the grid and the associated equipment.

Machine learning techniques are applied when there is no exact mathematical relationship among predictors, but there exist a pattern. Supervised learning relies on the availability of enough data points matched or classified to act as training data points for the machine learning model. The large dataset required for the analysis will imply that data driven methods of analyzing power system conditions will be required. They include the use of signal processing and statistical techniques such as multi-resolution analysis, wavelet transform, empirical mode decomposition, the stockwell transform and signal energy transformation method. It was reported by (Akmaz 2022) that stockwell transform can be applied to obtain classification features were 30 different classification features were obtained. It was observed, however, that the use of all classification features complicated the approach and prolonged the training and testing times. Consequently, (Kwon and Sim 2013) was motivated to examine how data set structures or characteristic affect the performances of classification algorithm. (Bagheri et al. 2022) reported that the mixture of semi-major axis, semi-minor axis and the direction of the major axis forms a satisfactory base to characterize and classify voltage events. There are a number of factors to consider that influence voltage events such as fault types, fault location, the earthing or grounding system and the connection of power quality instruments. (Guillen et al. 2018) applied power spectral density index classification feature in time and frequency for analysing transient information for a power system subjected to faulty condition. (Bellan and Furga 2018) reported that the investigation of transients in three phase systems can be achieved using the Clarke transformation.

There are several software that have been developed for classification and prediction such as TensorFlow, Shogun, Apache Mahout, Apache Spark MLlib, Oryx 2, Pytorch, H20.ai, RapidMiner, KNIME, Keras and WEKA (Kiranmai and Laxmi 2018) . The MathWorks have also created a classification learner app that can train models to classify data using supervised machine learning. The app enables users to explore data, select features, specify validation schemes, train models and assess the results. Some of the available classification model types include decision trees, discriminant analysis, support vector machines (SVM), logistic regression, nearest neighbors (NN), naïve bayes, kernel approximation, ensemble, and neural network classification (MathWorks 2023).

© IEOM Society International

Lastly, most commercially available power quality equipment sample at a rate of 256 samples per period as most of the power quality disturbance events frequency content is less than 5000 Hz. The sampling frequency dependent on the application and is required to be appropriately selected. In modeling events, data can be gathered from simulations or from actual historical events (Axelberg 2007). The need for open source datasets for power system modelling is emphasized by (Wiese et al. 2019; Zhenglong et al. 2022). (Tan and Ramachandaramurthy 2015) has developed models that may be use to generate some power quality events using MATLAB/Simulink software. Successful recognition of power system events in of paramount importance. (Miranda et al., 2019) demonstrated 100% recognition accuracy by convolution neural network (CNN) of four different classes of events (generator tripping, load disconnection, line tripping and inter-area oscillation) from a pool of phasor measurement unit (PMU) data. Though, applied to real systems, the classification and localization accuracies tend to be less than 100%, for instance, (Ren et al., 2020) yielded 84 and 91%, respectively.

It can be assessed that various approaches are needed to extract features from the three phase system, subjected under various events, classification of such events is possible with the help of appropriate classifiers. The paper presents steps that can be used to do the reconstruction of input signals, extraction of key features, train a machine learning model using a combination of features and identify the key ones for accurately predicting future events. The classifiers are also ranked based on performance during the training and testing process.

The remainder of the paper is divided into sections 3, 4, 5 and 6. Section 3 provides the methods used to extract features, section 4 highlights how the data was collected. The results and discussion is elaborated in section 5. The conclusion and future study is presented in section 6.

3. Methods

3.1 Sampling frequency selection and signal reconstruction

Here, the Nyquist –Shannon theorem or Whittaker-Nyquist Shannon sampling theorem is typically applied where the sampling frequency is taken to be at least twice the original signals fundamental frequency. For the purpose of getting the rms value of the original signal of 0.7071 p.u the algorithm in Figure 1 was used to obtain the sampling frequency.

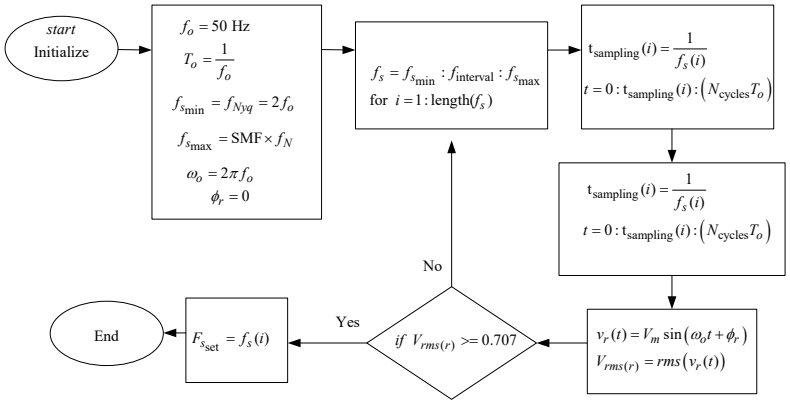


Figure 1. The sampling frequency selection algorithm

3.2 Data set generation and feature extraction

Consider the three phase voltage system with phasor representation shown in Figure 2.

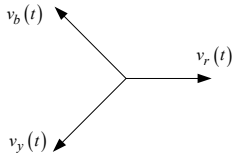


Figure 2. Three Phase Phasor Representation

© IEOM Society International

(Bellan and Furga, 2018) applies the Clarke transformation to derive the space vector transformation of the instantaneous three phase system using equation (1).

$$\begin{pmatrix} x_\alpha(t) \\ x_\beta(t) \\ x_0(t) \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \\ 1/2 & 1/2 & 1/2 \end{pmatrix}\begin{pmatrix} v_r(t) \\ v_y(t) \\ v_b(t) \end{pmatrix} \tag{1}$$

where, $x_\alpha(t)$, $x_\beta(t)$ functions form the space vector of the three phase system and $x_0(t)$ provide the zero sequence voltage. The three phase can therefore be represented by equations (2) and (3).

$$\vec{x(t)} = x_\alpha(t) + x_\beta(t) = \frac{2}{3}\begin{bmatrix} 1 & e^{j2\pi/3} & e^{j4\pi/3} \end{bmatrix}\begin{bmatrix} v_r(t) \\ v_y(t) \\ v_b(t) \end{bmatrix} \tag{2}$$

$$x_0(t) = \frac{1}{3}\left(v_r(t) + v_y(t) + v_b(t)\right) \tag{3}$$

The phase voltages may also be represented as having a rotating positive angular frequency phasor and a negative angular frequency phasor equations 4 -

$$v_r(t) = V_{\text{rms}}\sin\left(\omega t + \varphi_r\right) \tag{4}$$

$$V_{\text{rms}}e^{j(\omega t+\varphi_r)} = V_{\text{rms}}\left(\cos\left(\omega t + \varphi_r\right) + j\sin\left(\omega t + \varphi_r\right)\right) \tag{5}$$

$$V_{\text{rms}}e^{-j(\omega t+\varphi_r)} = V_{\text{rms}}\left(\cos\left(\omega t + \varphi_r\right) - j\sin\left(\omega t + \varphi_r\right)\right) \tag{6}$$

$$v_r(t) = \frac{V_{\text{rms}}}{2j}\left(e^{j(\omega t+\varphi_r)} - e^{-j(\omega t+\varphi_r)}\right) \tag{7}$$

Thus,

$$v_y(t) = \frac{V_{\text{rms}}}{2j}\left(e^{j(\omega t+\varphi_y)} - e^{-j(\omega t+\varphi_y)}\right) \tag{8}$$

$$v_b(t) = \frac{V_{\text{rms}}}{2j}\left(e^{j(\omega t+\varphi_b)} - e^{-j(\omega t+\varphi_b)}\right) \tag{9}$$

where; $\varphi_b = \varphi_r - 120°\left(\equiv \varphi_r - \frac{2\pi}{3}\text{rads}\right)$; $\varphi_y = \varphi_r + 120°\left(\equiv \varphi_y - \frac{2\pi}{3}\text{rads}\right)$

Therefore, the space vector can be written as equation 10.

$$\vec{x(t)} = \left|x_{\text{p}}\right|e^{j(\omega t+\varphi_{\text{p}})} + \left|x_{\text{n}}\right|e^{-j(\omega t-\varphi_{\text{n}})} \tag{10}$$

Therefore, decomposition the space phasor model into real and imaginary components generates the data matrix for constructing the ellipse shape in Figure 3. This approach is employs the principal component algorithm to determine the principal variables that estimate the ellipse (Ignatova et al., 2009).

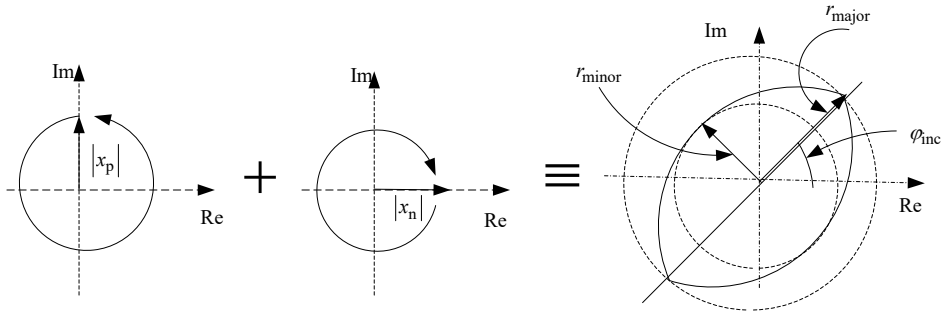


Figure 3. Representation of space vector

© IEOM Society International

The major axis, minor axis and inclination angles may be obtained using equations (11), (12), and (13), respectively.

$$r_{\text{major}} = \left|x_{\text{p}}\right| + \left|x_{\text{n}}\right| \tag{11}$$

$$r_{\text{minor}} = \left|\left|x_{\text{p}}\right| - \left|x_{\text{n}}\right|\right| \tag{12}$$

$$\varphi_{\text{inc}} = \frac{1}{2}\left(\varphi_{\text{p}} + \varphi_{\text{n}}\right) \tag{13}$$

The inclination angle of zero yields an ellipse with a horizontal major axis, while a 90 degree inclination angle produces an ellipse with a vertical major axis. Shape index (SI) is a parameter used to characterize the shape followed by the space vector in the complex plane obtained using (14).

$$S.I = \frac{r_{\text{minor}}}{r_{\text{major}}} \tag{14}$$

The value signifies the similarity index or relationship of the space vector to a circle. That is, SI of 1 signifies 100 % similarity, a value of zero indicates lack of similarity and yields a straight line space vector shape. The balanced three phase sinusoidal system results into a circle. Similarly, balanced voltage dips results into a circle. However, when the value of SI lies between zero and one, the space vector forms an ellipse shape as can be seen in Figure 4. This is so for unbalanced voltage events. When the power system is subjected to voltage harmonics, the effect corresponds to distorted circles or ellipses.

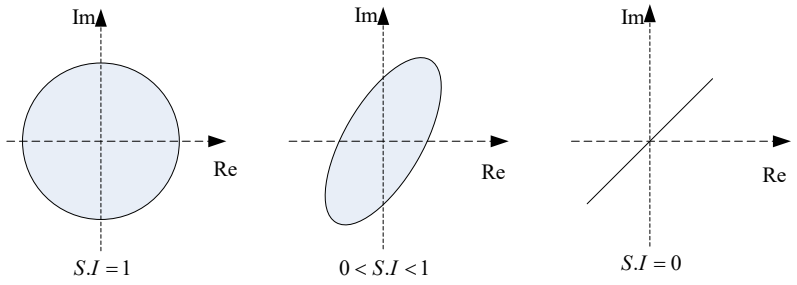


Figure 4. Shape index as a correlation coefficient of space vector shape

The ellipse parameters are obtained from singular value decomposition of the covariance matrix derived from equation (16)

$$\vec{x}\left(t\right) = \left[x_{t,R} \quad x_{t,I}\right]_{n\times 2} \tag{15}$$

$$\text{cov}\left(x_{t,R}, x_{t,I}\right) = \frac{1}{n-1}\sum_{i=1}^{N}\left(x_{t,R} - \mu_R\right) * \left(x_{t,I} - \mu_I\right) \tag{16}$$

In matrix notation, we get equation (17)

$$\text{cov}\left(x_{t,R}, x_{t,I}\right) = \frac{1}{n-1}\left[\left(x_{t,R1} - \mu_R\right) \quad \left(x_{t,R2} - \mu_R\right) \quad \ldots \quad \left(x_{t,Rn} - \mu_R\right)\right]\begin{bmatrix}\left(x_{t,I1} - \mu_R\right)\\ \left(x_{t,I1} - \mu_R\right)\\ \ldots \\ \left(x_{t,In} - \mu_R\right)\end{bmatrix} \tag{17}$$

In short, equation (18)

$$\text{cov}\left(x_{t,R}, x_{t,I}\right) = \frac{1}{n-1} M_R^T \otimes M_I \tag{18}$$

© IEOM Society International

where M_R^T is the 1 by n matrix with the $x_{t,Rn} - \mu_R$ elements transposed and M_I is a n by 1 column matrix the $x_{t,In} - \mu_R$ elements

Another useful way of extracting features from a waveform is by signal decomposition using wavelet transforms. Figure 5 illustrates one dimensional signal decomposition that was used in this research.

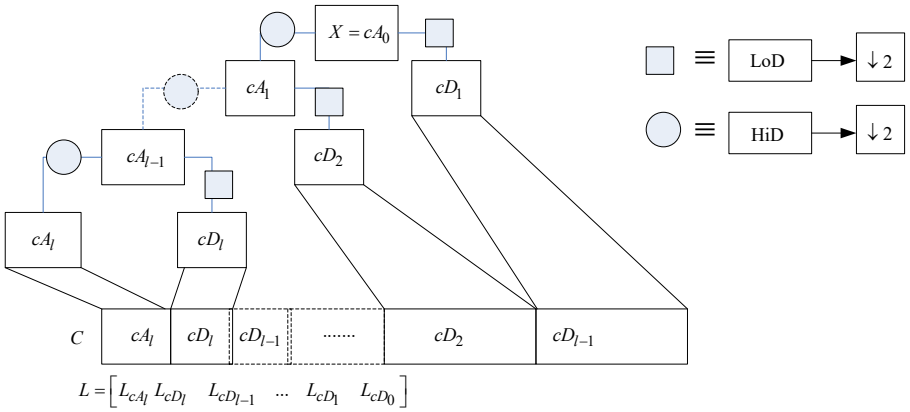


Figure 5. One dimensional discrete wavelet transform

Given a signal X, of length N, the number of levels that are required to be performed during the discrete wavelet transform (DWT) is given by equation 19. At each level, the approximate coefficients and detail coefficients are determined by convolving the input signal with a low pass filter (LoD) and the high pass filter (HiD), thereafter a dyadic decimation (down sampling) is performed, respectively. The available analytic wavelets include Morse wavelet, Morlet (Gabor) wavelet and Bump.

$$l \le \log_2(N) \tag{19}$$

The correlation coefficients between signals under the normal condition event and those during disturbances or abnormal power quality events is a measure of linear dependence given by equation 20, the Pearson correlation coefficient.

$$\rho(X_1, X_2) = \frac{1}{N-1}\sum_i^N \left(\frac{X_{1i} - C}{\sigma_{X_1}}\right)\left(\frac{X_{2i} - \mu_{X_2}}{\sigma_{X_2}}\right) = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \tag{20}$$

where μ is the mean, σ the standard deviation. That is, the correlation coefficients can also be obtained by dividing the covariance of X_1 and X_2 by the product of the mean and the standard deviation.

More features can be extracted by obtaining the coefficient matrix when signals $X_1, X_2, \ldots, X_N$ are compared is given by equation 21. Accordingly, the correlation between detailing coefficients or approximation coefficients obtained after wave decomposition at different levels for each phase can be obtained for each event.

$$R = \begin{bmatrix} \rho(X_1, X_1) & \rho(X_1, X_2) & \ldots & \rho(X_1, X_N) \\ \rho(X_2, X_1) & \rho(X_2, X_2) & \ldots & \rho(X_2, X_N) \\ \ldots & \ldots & \ldots & \ldots \\ \rho(X_N, X_1) & \rho(X_N, X_2) & \ldots & \rho(X_N, X_N) \end{bmatrix} \tag{21}$$

3.3 The training process

A supervised machine learning approach was used to develop a model for predicting voltage events on the power system. The process of data synthesis or capture, pre-processing, training, evaluation and validation can be depicted in form of a flow chart in Figure 6.

© IEOM Society International

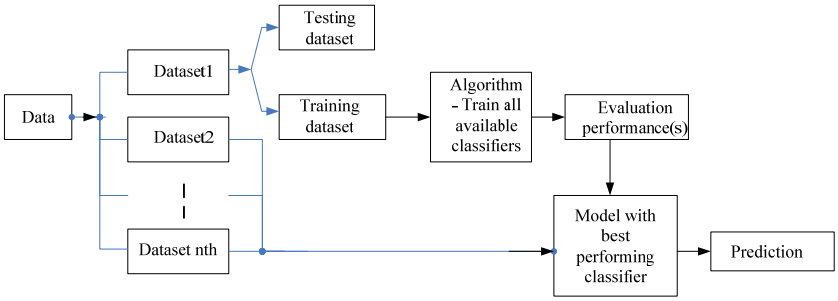


Figure 6. The machine learning – training, testing and validation procedure

The classifier settings are as follows (MathWorks 2023);

- Decision Trees include the Fine tree, medium tree, and coarse tree. For the fine tree, the maximum number of splits is 100, split criteria uses the Gini's diversity index with no surrogate decision trees. For the medium tree twenty (20) is the maximum number of splits whereas four (4) for the coarse tree.
- Discriminant include linear and quadratic with a full covariance structure.
- Naïve Bayes include Gaussian Naïve and Kernel Naïve Bayes with unbounded support.
- Support Vector Machines (SVM) kernel functions includes linear, quadratic, cubic, fine Gaussian (scale of 0.56), medium Gaussian (scale of 2.2) and coarse Gaussian (scale of 8.9).
- Nearest neighbor (NN) classifiers include fine (number of neighbors set to 1, Euclidean distance metric with equal distance), medium (number of neighbors set to 10, Euclidean distance metric with equal distance), coarse (number of neighbors set to 100, Euclidean distance metric with equal distance), cosine (number of neighbors set to 10, cosine distance metric, equal distance weight), cubic (number of neighbors set to 10, minikowski distance metric, equal distance weight) and weighted (number of neighbors set to 10, Euclidean distance metric with distance weight squared inverse) KNN.
- Ensembles with methods that include Boosted Trees (AdaBoost, learner type by decision trees with maximum number of splits 20), subspace discriminant (number of learners 30, subspace dimension of 3), subspace KNN (nearest neighbor, number of learners 30, subspace dimension of 3), RUSBoosted trees (decision trees with maximum number of splits of 20, number of learners 30, learning rate of 0.1)
- Neural Networks methods include narrow (number of fully connected layers of 1, first layer size of 10, activation by ReLU, iteration limit of 1000, regularization strength (lambda) of 0), medium (first layer size of 25), wide (first layer size of 100), bilayered (number of fully connected layers of 2, first layer size of 10, second layer size of 10) and trilayered neural networks (number of fully connected layers of 3, first layer size of 10, second layer size of 10, third layer size of 10).

4. Data Collection and Classification

Equally distributed rms values for each phase was extracted at the start of the triggering point of the disturbance. Apart from the rms value, the harmonic component feature is considered as part of the voltage waveform component during events. That is, odd harmonics such as the third harmonic, fifth (5^{th}) and seventh (7^{th}) harmonics are prominent and are the ones captured. The shape index feature was formulated by performing singular value decomposition of the space phasor model of the recorded three phase voltage waveforms. Lastly, since each voltage event has a specific duration of occurring on the power system, duration feature is also added. Consequently, a feature vector of six (6) components or predictors for each three phase recording is established and each combination having the correct

© IEOM Society International

classification. The feature vector (FV) can be simplified using equation 28. The number of training datasets for each class was 100 (i.e. $N = 100$).

$$FV = \begin{bmatrix} \text{rms} & \text{duration} & \text{V3h} & \text{V5h} & \text{V7h} & \text{SI} \\ \text{rms}_1 & \text{duration}_1 & \text{V3h}_1 & \text{V5h}_1 & \text{V7h}_1 & \text{SI}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{rms}_N & \text{duration}_N & \text{V3h}_N & \text{V5h}_N & \text{V7h}_N & \text{SI}_N \\ \text{rms}_{N+1} & \text{duration}_{N+1} & \text{V3h}_{N+1} & \text{V5h}_{N+1} & \text{V7h}_{N+1} & \text{SI}_{N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{rms}_{2N} & \text{duration}_{2N} & \text{V3h}_{2N} & \text{V5h}_{2N} & \text{V7h}_{2N} & \text{SI}_{2N} \\ \text{rms}_{2N+1} & \text{duration}_{2N+1} & \text{V3h}_{2N+1} & \text{V5h}_{2N+1} & \text{V7h}_{2N+1} & \text{SI}_{2N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{rms}_{3N} & \text{duration}_{3N} & \text{V3h}_{3N} & \text{V5h}_{3N} & \text{V7h}_{3N} & \text{SI}_{3N} \\ \text{rms}_{3N+1} & \text{duration}_{3N+1} & \text{V3h}_{3N+1} & \text{V5h}_{3N+1} & \text{V7h}_{3N+1} & \text{SI}_{3N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{rms}_{4N} & \text{duration}_{4N} & \text{V3h}_{4N} & \text{V5h}_{4N} & \text{V7h}_{4N} & \text{SI}_{4N} \\ \text{rms}_{4N+1} & \text{duration}_{4N+1} & \text{V3h}_{4N+1} & \text{V5h}_{4N+1} & \text{V7h}_{4N+1} & \text{SI}_{4N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{rms}_{5N} & \text{duration}_{5N} & \text{V3h}_{5N} & \text{V5h}_{5N} & \text{V7h}_{5N} & \text{SI}_{5N} \end{bmatrix}_{5N\times 6} \rightarrow \begin{bmatrix} \text{class} \\ \text{V}_{\text{sag}} \\ \vdots \\ \text{V}_{\text{sag}} \\ \text{V}_{\text{swell}} \\ \vdots \\ \text{V}_{\text{swell}} \\ \text{V}_{\text{interruption}} \\ \vdots \\ \text{V}_{\text{interruption}} \\ \text{V}_{\text{harmonics}} \\ \vdots \\ \text{V}_{\text{harmonics}} \\ \text{V}_{\text{normal}} \\ \vdots \\ \vdots \\ \text{V}_{\text{normal}} \end{bmatrix}_{5N\times 1} \quad (28)$$

5. Results and Discussion

Each disturbance recording consist of three phases of voltage waveforms with a nominal frequency of 50 Hz. The optimal sampling frequency was determined by establishing the relationship of the root mean square value in per unit of the un-disturbed red phase at various sampling frequencies. Figure 7 indicates that sampling frequencies above 20 kHz are able to accurately reproduce the original signal. It has been reported that some power quality equipment have sampling frequencies in the range of 20 kHz to 40 kHz. Some operate at even higher sampling frequencies of 400 kHz, in which case the rms approximates 0.7071 p.u.

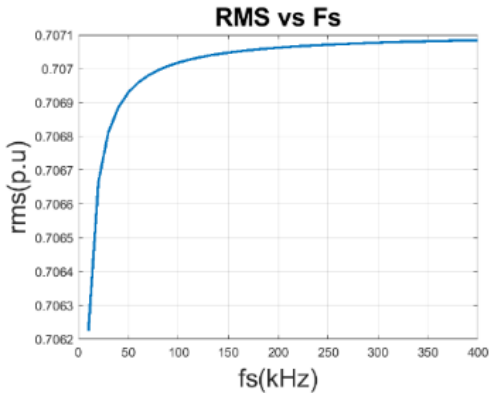


Figure 7. The RMS value for red phase voltage (sinusoidal) in per unit at various Sampling frequency

5.1 Numerical Results

The features extracted using correlation of waveforms with the maximum level of decomposition of 15 determined by equation (19). Taking a level of 3 using the Deubechies level 3 (i.e. Db3) wavelet signal decomposition results in the following parameters to describe the original signal balanced three phase waveform in Table 1.

Table 1. Correlation coefficients of obtained one dimensional decomposition using Deubechies with 3 levels for a balanced upfaulted three phase system

SI	ry	rb	yb	Cd1ry	Cd1rb	Cd1yb	Cd2ry	Cd2rb	Cd2yb	Cd3ry	Cd3rb	Cd3yb
1	-0.5	-0.5	-0.5	-1	-1	+1	-1	-1	+1	-1	-1	+1

© IEOM Society International

where Cd1ry is the correlation coefficient between the detailing coefficients of red phase level 1 decomposition and yellow phase decomposition. Similarly, Cd2ry is at level 2 and Cdr3 and level 3.

The simulations were done using a Lenovo Laptop with Intel® Celeron ® CPU B830 @1.80GHz, x64 –based, 4 GB RAM.
The first case used 500 observations with five (5) predictors or features and a 5 fold-cross validation scheme using 29 classifiers, the results are represented in Table 2. It can be noted that here, no classifier was able to record 100% accuracy performance. However, some classifiers scored high performances such as the decision tree classifiers (fine, medium and coarse) and subspace with 99.8% accuracy revealed to be the best options with the first dataset. Other classifiers also recorded good performances such as the neural networks and support vector machines. However, the quadratic discriminant and Gaussian naïve bayes failed to be trained. It was observed that the training time and accuracy performance had no correlation.

Table 2. Classification learner accuracy, prediction speed and training time

No.	Classifier description or model type	Accuracy (validation)	Total cost validation	Prediction speed	Training time
1	Fine Tree	99.8%	2	~2400 obs/sec	36.401 sec
2	Medium Tree	99.8%	2	~8600 obs/sec	2.443 sec
3	Coarse Tree	99.8%	2	~9900 obs/sec	1.7726 sec
4	Subspace KNN Ensemble	99.8%	2	~460 obs/sec	8.7365 sec
5	Trilayered Neural Network	99.6%	N/A	~10000 obs/sec	2.5875 sec
6	Quadratic SVM	99.6%	4	~1600 obs/sec	4.7448 sec
7	Bagged Trees Ensemble	99.6%	2	~1100 obs/sec	10.509 sec
8	Medium Neural Network	99.6%	N/A	~9100 obs/sec	2.1527 sec
9	Cubic SVM	99.4%	4	~3400 obs/sec	4.2102 sec
10	Fine Gaussian SVM	99.4%	3	~2400 obs/sec	3.5254 sec
11	Fine KNN	99.4%	6	~3000 obs/sec	11.014 sec
12	Narrow Neural Network	99.4%	N/A	~7800 obs/sec	7.18301 sec
13	Wide Neural Network	99.2%	N/A	~8700 obs/sec	2.2168 sec
14	Linear SVM	99.0%	6	~2600 obs/sec	13.598 sec
15	Bilayered Neural Network	99.0%	N/A	~11000 obs/sec	2.3971 sec
16	Medium Gaussian SVM	98.4%	9	~2700 obs/sec	4.6118 sec
17	Weighted KNN	98.0%	10	~5000 obs/sec	1.6419 sec
18	Cosine KNN	97.8%	11	~4600 obs/sec	2.404 sec
19	Medium KNN	96.4%	18	~4100 obs/sec	2.0968 sec
20	Cubic KNN	96.4%	19	~4000 obs/sec	1.9734 sec
21	Coarse Gaussian SVM	90.2%	49	~3400 obs/sec	2.6216 sec
22	Linear Discriminant	88.0%	59	~5300 obs/sec	6.0945 sec
23	Subspace Discriminant Ensemble	88.0%	61	~740 obs/sec	8.7573 sec
24	Coarse KNN	83.0%	85	~3900 obs/sec	2.166 sec
25	Kernel Naïve Bayes	79.0%	105	~6600 obs/sec	10.107 sec
26	Boosted Trees Ensemble	20.0%	400	~6200 obs/sec	8.851 sec
27	RUSBoosted Tress	20.0%	400	~9500 obs/sec	2.2621 sec
28	Quadratic Discriminant	Failed	N/A	N/A	N/A
29	Gaussian Naïve Bayes	Failed	N/A	N/A	N/A

5.2 Graphical Results

For a balanced three phase system the results of a space phasor model (SPM) are close to a unit circle. During voltage disturbance events, the resulting shape, i.e. ellipse has parameters unique to each event. Therefore, the ellipse parameters can be used as indicators and a way of characterizing such events as shown in Figure 8.

© IEOM Society International

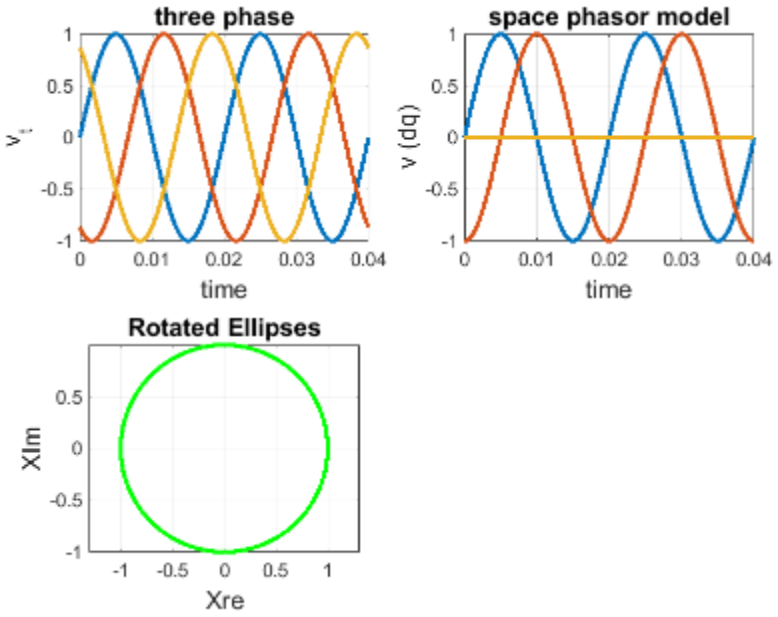


Figure 8. A way of characterizing events

To evaluate the classifier performance, the confusion matrix, Receiver Operating Characteristic (ROC) curve, parallel coordinates plot, and the scatter plot are assessed. For example the classification learner results for Fine Tree classification technique are illustrated in Figure 9. It can be shown that the voltage sag and voltage interruption events were misclassified for some instances. This can be seen on the scatter plot of the rms and duration of these events tends to be similar in some cases.

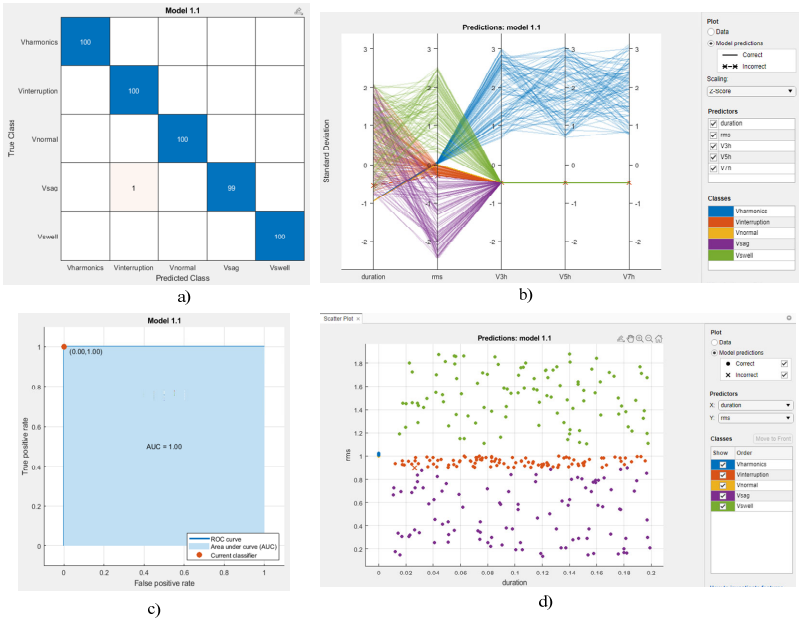


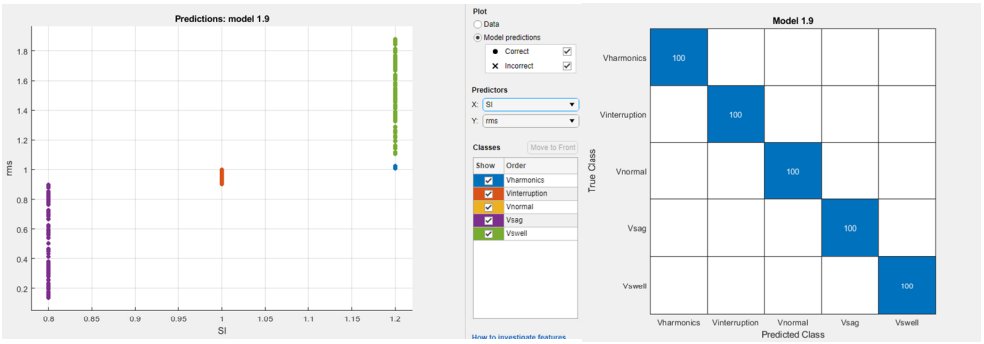
Figure 9. Fine Tree classifier results – a) confusion matrix, b) parallel coordinates, c) ROC curve, d) scatter plot

This was more so in other classifiers that recorded poor accuracy performances of 20% such as the boosted trees.

© IEOM Society International

5.3 Proposed Improvements

To have a model with 100% accuracy, the potential improvement was achieved by incorporating ellipse characteristic features and wavelet transform features in the training dataset. By including the shape index feature the voltage sag event and voltage interruption event which were initially in some instances misclassified became more distinct as can be seen in Figure 10.



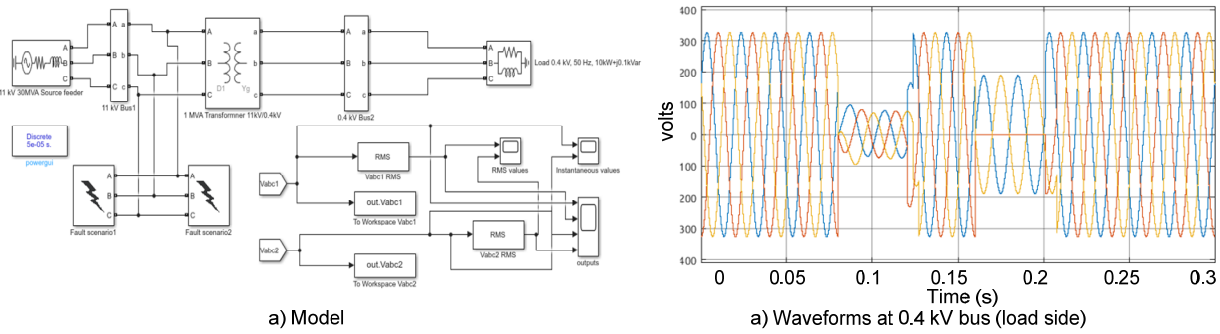
a) b)

Figure 10. SVM Quadratic classifier with Shape index (SI) feature: a) scatter plot rms vs SI, b) confusion matrix

The techniques that improved and had an accuracy (validation) of 100% were SVM (i.e. quadratic, cubic, fine Gaussian and medium Gaussian), KNN (i.e. fine, medium, cosine, cubic, weighted, and subspace ensemble) and Neural Networks (i.e. narrow, medium, wide, and bilayered). However, other techniques were unable to provide 100% classification accuracy, those included linear SVM (99.8%), decision trees (99.6% for fine, medium and coarse), subspace discriminant ensemble (99.6%), trilayered neural network (99.4%), coarse Gaussian SVM (99.2%), coarse KNN (89.8%), subspace discriminant ensemble (88%), Kernel Naïve Bayes (79%), Boosted trees ensemble and RUSBoosted trees ensemble (20%). There were three classifiers that failed to provide any predicted model, that is, the linear discriminant, quadratic discriminant and the Gaussian Naïve Bayes.

5.4 Validation

Only 80% of the dataset was used during the training process. The validation of the trained model was done using the remaining 20 %. Furthermore, a MATLAB Simulink model in Figure 11 was created to generate voltage swell and sag events testing datasets, it is a modified version to one proposed by (Tan and Ramachandaramurthy 2015).



a) Model a) Waveforms at 0.4 kV bus (load side)

Figure 11. MATLAB model for generating testing dataset for sag and swell events

The results obtained validated the results in section 5.1, 5.2 and 5.3.

6. Conclusion

The research aims at establishing key voltage disturbance event features that a machine learning model can use to provide 100% classification accuracy (validation). The sampling frequency selection was achieved by minimizing the error in the root mean square value of the reconstructed waveform from its corresponding original waveform. The key

© IEOM Society International

features extracted from the three-phase voltage waveform included the rms value, presence of 3rd, 5th and 7th harmonics, duration of disturbance, and shape index obtained from elliptical shape representation of space phasor model. The classification learner application was used to create a supervised machine learning model by training twenty-nine (29) classifiers and comparing their performance using the accuracy (%), confusion matrix, total cost validation, prediction speed and training time. The results indicated that without the shape index, the voltage sag and voltage interruption events were misclassified in some scenarios for all the classifiers, with the highest accuracy obtainable of 99.6% by four (4) classifiers. Some classifiers such as the Quadratic discriminant and Gaussian Naïve Bayes failed to be trained. The inclusion of shape index feature improved some of the trained models and it was observed that fourteen (14) classifiers were able to achieve 100% classification accuracy. However, the Quadratic discriminant, Gaussian Naïve Bayes and now including the linear discriminant classifiers failed to be trained. The training, computing, and processing is required to be of high performance and accuracy to increase the online situational awareness of network operators when such systems are implemented.

The research provides additional insight into the process of power quality disturbance identification and characterization with emphasis on voltage events. It identifies that addition of shape indices for each event enables a machine learning model to have higher prediction accuracy. However, the model is limited to six features, which limits its application.

Future research may include increasing the features, training dataset and validating the efficacy using data from real power stations, noise detection and filtering, as well as fault type characterization. Furthermore, detailed comparison of the classification technique is required to narrow down to an optimal technique. The feature extraction and representation might be modified to ensure that all classifiers learn something than failing completely.

References

Akmaz, D., Classification of single and combined power quality disturbances using Stockwell Transform, relief feature selection method and multilayer perceptron algorithm, *Naturengs, MTU Journal of Engineering and Natural Sciences*, vol. 3, no. 1, pp. 13-23, 2022.

Axelberg, P., On tracing flicker sources and classification of voltage disturbances, PhD thesis, *Chalmers University of Technology*, 2007: Accessed on 15 Feb, 2023. Available on: https://citeerx.ist.psu.edu/document?repid=repl&type=pdf&doi=15cd856d3d8e5ca3346be8591a4ef0d34da6c05

Bagheri, A., Oliveira, R.A.d., Bollen, M.H.J., and Gu, I.Y.H., A framework based on machine learning for analytics of voltage quality disturbances, *Energies*, vol. 15, no. 4, 1283, 2022.

Bellan, D. and Furga, G.S., Space vector state equations analysis of three phase transients, *Journal of Electrical Systems JES*, vol. 14, no. 1, pp. 188-198, 2018.

Guillen, D., Paternina, M.R.A., Beja, J.O., Tripathy, R.K, Mendez, A.Z., Olvera, R.T., and Tellez, E.S., Fault detection and classification in transmission lines based on a PSD index, *IET Generation, Transmission & Distribution*, vol. 12, no. 18, pp. 4070-4078, 2018.

Ignatova, V., Granjon, P. and Bacha, S., Space vector method for voltage dips and swells analysis, *IEEE Transactions on power delivery*, vol. 24, no. 4, pp. 2054-2061, 2009.

Kiranmai, A.S., and Laxmi, J.A., Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy, Protection and Control of Modern Power System, 3, 29, 2018.

Kulikov, A.L., Shepovalova, O.V., Ilyushin, P.V., Filippov, S.P. and Chirkov, S.V., Control of electric power quality indicators in distribution networks comprising a high share of solar photovoltaic and wind power stations, *Energy Reports*, vol. 8, pp. 1501-1514, 2022.

Kwon, O. and Sim, J.M., Effects of data set features on the performance of classification algorithms, Expert Systems with Applications, vol. 40, no. 5, pp. 1847-1857, 2013.

Miranda, V., Cardoso, P.A., Bessa, R.J., and Decker, I., Through the looking glass: Seeing events in power systems dynamics, *International Journal of Electrical Power & Energy Systems*, vol. 106, pp. 411-419, 2019.

McGranaghan, M.F., Santoso, S., Challenges and Trends in Analyses of Electric Power Quality Measurement Data. *EURASIP J. Adv. Signal Process.* 2007, 057985, 2007.

Ren, H., Hou, Z.J., Vyakaranam, B., Wang, H., and Etingor, P., Power system event classification and localization using a convolutional neural network, *Frontiers in Energy Research*, vol. 8, pp. 1-11, 2020.

Tan, R.H.G. and Ramachandaramurthy, V.K., A comprehensive modeling and simulation of power quality disturbances using MATLAB/Simulink, Book chapter, *Intech Open*, 2015.

© IEOM Society International

Wiese, W., Schlecht, I., Bunke, W.D., Gerbaulet, C., Hirth, L., Jahn, M., Kunz, F., Lorenz, C., Muhlenpfordt, J., Reimann, J., and Schill, W.P., Open power system data- frictionless data for electricity system modelling, *Applied Energy*, vol. 236, pp. 401-409, 2019.

Zhenglong, S., Machlev, R., Chao, J., Qianchao, W., Perl, M., Belikov, J. and Levron, Y., PF-FEDG: An open-source data generator for frequency disturbance event detection with deep-learning reference classifiers, *Energy Reports*, vol. 9, pp. 397-413, 2022.

Biography

Buchizya Kumwenda is a Lecturer at the Copperbelt University, School of Engineering in the Electrical Department, Kitwe, Zambia. He has over 7 years of practical experience in teaching and research as well as academic administration. Buchizya has Bachelor of Engineering in Electrical and Electronics Engineering and Master of Philosophy in Electrical Engineering. He has published a number of papers on Solar Energy Integration into Power Grids considering ramp rate constraints, load shedding effects on residential consumers, solar home system geospatial placement and sizing considerations and electric vehicle requirements in developing countries. His papers have been presented papers in USA, Ghana, South Africa, and Rwanda. His research interests include renewable energy, optimization, reliability, power quality, and scheduling and electricity market. He is member of EIZ.

© IEOM Society International