❒   1626

Naïve bayes and linear discriminate analysis based diagnostic analytic of harmonic source identification

M.H Jopri[1], MR Ab Ghani[2], A.R Abdullah[3], T Sutikno[4], M Manap[5], J. Too[6]

[1,5]Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
[2,3,6]Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
[4]Deparment of Electrical Engineering, Universitas Ahmad Dahlan, Indonesia

Article Info	ABSTRACT
	The diagnostic analytic type of harmonic source is a vital research due to diagnose and identify type of harmonic source that exist in the power system. This paper presents a comparison of machine learning (ML) algorithm namely as the Naïve Bayes (NB) and linear discriminate analysis (LDA) in identifying and diagnosing the harmonic sources. The MLs inputs are the voltage and current feature sets that estimated from the time-frequency representation (TFR) of S-transform analysis. Four specific cases of harmonic source location are considered in this research, whereas harmonic voltage (H_V) and harmonic current (H_C) source type-load are used in the diagnosing process. The sufficiency of the proposed methodology is tested and verified on the IEEE 4-bust test feeder, and to prevent overfitting, the K-fold cross-validation technique is implemented for performance evaluation. To identify the best ML, the performance measurement consist of the accuracy, precision, geometric mean, F-measure, sensitivity, and specificity are conducted.

Corresponding Author:

Jopri, Mohd Hatta,
Faculty of Electrical and Electronic Engineering Technology,
Universiti Teknikal Malaysia Melaka (UTeM),
Jalan Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
Email: hatta@utem.edu.my

1. INTRODUCTION

Power quality signals are considered as the natural pathway to detect the harmonic sources, and since the resulting voltage and current are the additions of multiple action potential traversing the electric circuit [1, 2]. Research in the identification, detection and classification using machine learning algorithm has become one of the major interests, which allows the multiple harmonic sources identification become viable [3-5]. In the identification process, the features are first extracted, and then the machine learning algorithm is implemented in order to generate a better understanding around the potential features or solutions [6-8]. The harmonic sources had increased and may lead to the potential risk to the existing of the electrical power system [9-11]. Accordingly, the harmonic current flows also lead to extra overheating, losses and overloading to the power system equipment [12, 13]. Accordingly, it is imperative to distinguish the most reasonable methodologies to recognize and diagnose the kind of the harmonic sources in the power system [14, 15]. Different techniques have been proposed by researches due to diagnose the sort of harmonic sources dependent on various hypothetical standards, highlights, advantages, and downsides [16, 17]. Moreover, a high skill of technical experience is required to appropriately diagnose the harmonic source type [10, 18]. The research literatures on the diagnosis technique utilizing the stochastic methodology and indexing of power quality been clarified in [19, 20]. Be that as it may, this strategy uses Fourier and wavelet transform of signal processing methods, while numerous limitations of these methods are clarified in [21, 22]. Moreover, a diagnosis analysis utilizing a short-time Fourier transform (STFT) and S-transform dependent on a rule-based classifier is presented in [23, 24].

In any case, a technical knowledge is required due to set the rule-based classifier parameter [25, 26]. The power quality signal analysis is a complex because the existence of complex linear and non-linear of the signals, thus an old statistical strategy for auto-regressive integrated moving average, is presented [27]. To increase the accuracy of the ARIMA method, the vector autoregressive (VAR) model and artificial neural networks (ANN) are presented. This is because of capture the uncover the non-linear aspects and the linear interdependence, though made this technique troublesome and complex to apply. Numerous researches utilized statistical learning theory (SLT) for the pattern recognition.

In [28], a utilization of neural network based on the classification algorithm for recognizing the disturbance signals is introduced. In any case, this technique isn't appropriate to build the frequency spectrum range, hence brings about the loss of the low energy of frequency components [29]. A computerized advancement algorithm with the mechanics of characteristic genetics and regular choice alternative is known as genetic algorithm (GA) additionally been presented in diagnostic analysis. The GA fundamental to compute the highest endurance probability of the fittest individual. The main benefit of this method is, it use variable coding set as a substitute of the actual factors. In any case, the downside of GA such not, hard to choose the fitness function, and the accuracy is dependent on the problem knowledge [30]. The fuzzy logic (FL) and probabilistic neural network are utilized in [31] to recognize kind of harmonic sources. In any case, the complete exactness of this strategy is decline whenever new type of load is existed [19]. Besides, the FL has a downside, for example, the structure of the rule-based is difficult, and FL isn't adaptable to new disturbance events. The logistic regression is one the well known characterization algorithms in the AI field. The logistic regression is normally utilized for classification with linear regression function [21]. In any case the logistic regression unfit to take care of the non-linear data as this technique required the availability of all the significant independent variables [32]. To date, many existing machine learning algorithms such as support vector machine (SVM), k-nearest neighbor (KNN), linear discriminate analysis (LDA), artificial neural network (ANN), and random forest (RF) have successfully applied to other engineering applications, [6, 33].

This paper proposes high accuracy, and costs effective technique to identify and diagnose the type of harmonic sources in the distribution system using machine learning methods and single-point measurement approach at the point of common coupling (PCC) [26, 34, 35]. The selected machine learning methods are Naïve Bayes (NB) and linear discriminate analysis (LDA). The NB is a simple classifier based on Bayes' theorem that assumes all the features are independence [36], whereas the linear discriminate analysis (LDA) is faster and computationally less expensive. At last, the best machine learning method for identifying and diagnosing harmonic sources is selected based on the performance measurement criteria such as the accuracy, precision, geometric mean, and F-measure [3].

2. RESEARCH METHODOLOGY
2.1. Machine learning methods
In this research, two popular machine learning algorithms, namely Naïve Bayes (NB) and linear discriminate analysis (LDA) for the identification and diagnostic of harmonic sources are proposed. The NB and LDA are chosen due to their promising performance in previous work. To prevent overfitting, the K-fold cross-validation technique is implemented for performance evaluation [36, 37]. Initially, the data is divided equally into K parts, and each part is used for testing in succession. Meanwhile, the remaining parts are used to train the classifier. The process is repeated for K times and the average results are recorded.

Naïve Bayes (NB) is another common algorithm in this world. The NB is a simple classifier based on Bayes's theorem that assumes all the features are independence [36]. In this work, the NB with normal distribution function is applied. On the other hand, Linear Discriminate Analysis (LDA) is one of the robust machine learning algorithms among the popular classifiers. The LDA projects the feature in the most discriminative ways for identification process [36]. In this work, LDA with basic linear function is implemented.

3. PROPOSED METHOD
Based on recent literature, the proposed technique can be fulfilled using the measurement technique at the PCC and presented in Figures 1 and 2 [38]. Moreover, the harmonic signals are analyzed using S-transform and two types of harmonic sources comprise of the harmonic current source (H_C) and harmonic voltage source (H_V) type-load [1]. Four specific cases considered, which are [2, 35]: (1) no harmonic source in the power system (N-N), (2) harmonic source located at the downstream (N-H) of the PCC, (3) harmonic source located at the downstream and upstream of the PCC (H-H), and (4) harmonic source located at the upstream of the PCC (H-N). The main aim of this research is to identify and diagnose type of harmonic sources in the power system.
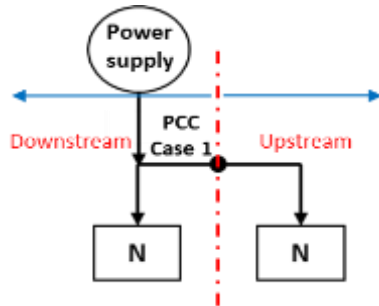
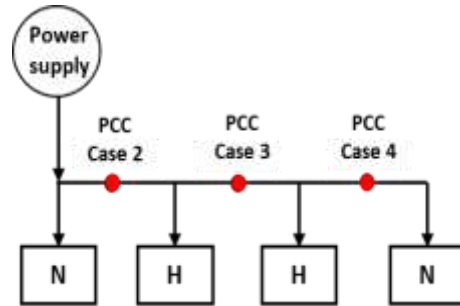Figure 1. Upstream-Downstream for Case 1

Figure 2. IEEE 4-bus test feeders for Case 2, 3 and 4

where N is a non-harmonic source, which is resistor load, H is harmonic producing load and H can be H_C or H_V, respectively.

Figure 3 shows an overview of the proposed method. Initially, the voltage and current signals are measured at the PCC. Next, the S-transform analysis is applied to transform the voltage and current signals into time-frequency representation (TFR). Then, the signal parameters are then estimated from the TFR, and the parameters form into two feature sets: (1) current feature set, (2) voltage feature set. The feature sets are normalized and then fed into the ML for the diagnosis of harmonic sources. The LDA and KNN are applied to diagnose the NN, NH, HH, and HN cases for both H_C and H_V.
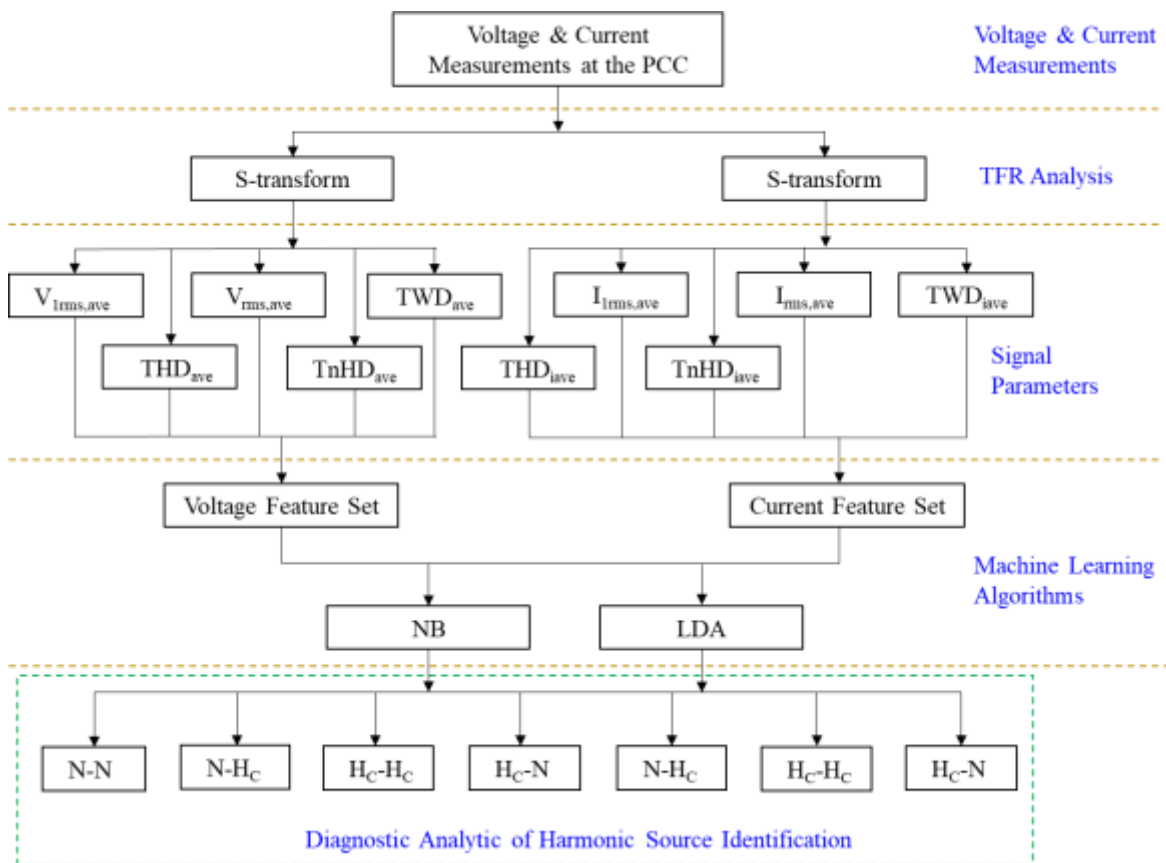

Figure 3. An overview of the proposed method

3.1. Voltage and current feature sets

In this research, the voltage and current feature sets consist five signal parameters that estimated from the voltage and current signals of the PCC, respectively [5] :

- The average instantaneous RMS of voltage and current ($V_{rms,ave}$ and $I_{rms,ave}$),
- The average instantaneous RMS fundamental of voltage and current ($V_{1rms,ave}$ and $I_{1rms,ave}$),
- The average instantaneous total harmonic distortion of voltage and current (THD_{ave} and THD_{iave}),
- The average instantaneous total nonharmonic distortion of voltage and current ($TnHD_{ave}$ and $TnHD_{iave}$) and
- The average instantaneous total waveform distortion of voltage and current (TWD_{ave} and TWD_{iave})

3.2. Performance measurement of machine learning methods

In this research, several features are estimated from the power quality signals. Hence, two feature sets, namely voltage feature set and current feature set are collected. In order to prevent numerical problem, these estimated features are normalized between 0 and 1. Afterward, the features are fed into the NB and LDA for the identification process, which aims at identifying the harmonic sources. It is worth noting that NB and LDA are computed for 10 runs and the average results are recorded. For performance measurement, six popular evaluation metrics, namely accuracy, precision, geometric mean, sensitivity, specificity, and F-measure are used, and they can be defined as follows [3, 4]:

3.2.1 Accuracy

The accuracy is a metric used to measure how accurate the proposed harmonic source diagnosis system can be. It is defined as [37],

$$Accuracy = \frac{No.\,of\,corrected\,diagnose\,samples}{Total\,number\,of\,samples} \tag{1}$$

3.2.2 Precision

Precision is the proportion of the expected positive samples that were corrected, and it can be expressed as,

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3.2.3 Geometric mean

Geometric mean (G-mean) is an enhanced version of sensitivity and specificity, and it can be calculated as [36],

$$Geometric_{mean} = \sqrt{Sensitivity \times Specificity} \tag{3}$$

3.2.4 Specificity

Specificity is another commonly used metric in diagnosis, and it can be calculated as [37],

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

3.2.5 Sensitivity

Sensitivity is a commonly used metric in diagnosis, and it can be expressed as [36],

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

3.2.6 F-measure

F-measure is an important metric that supports accuracy, and it is used to characterize the performance of a classifier. It can be defined as [37],

$$F_{measure} = \frac{2TP}{2TP + FN + FP} \tag{6}$$

Where TP is the true positive, TN is the true negative, FN is the false-negative, and FP is the false positive, which can be obtained from the confusion matrix.

4. RESULTS AND ANALYSIS

Table 1 shows the results of accuracy, precision, geometric mean, F-measure, sensitivity and specificity for the identification of the harmonic sources using NB and LDA for voltage feature set. On the other side, a clear visualization of performance measurements can be found in Figure 4. From Table 1, one can see that the performance of NB and LDA were very low, which cannot identify the harmonic sources in an effective manner. Even though NB performed better than LDA in voltage feature set, but NB scored the accuracy of 40%, which was lower than half and this is not suitable for real time application. The reason might be the weak discriminability of voltage features, thus leading to worst performance in current work. Based on the results obtained, voltage features were not the good selection for harmonic source identification.
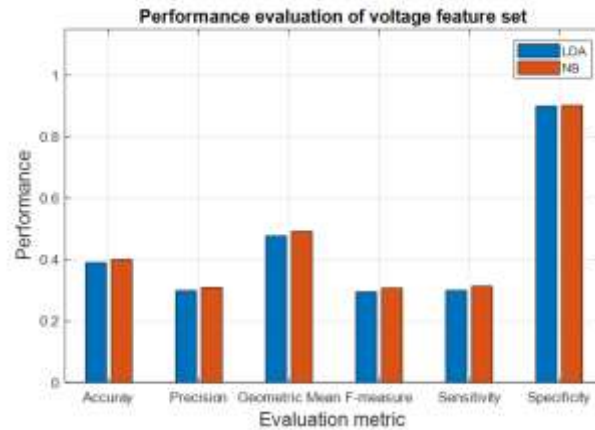

Figure 4. Performance measurement of NB and LDA using voltage feature set

Table 1. The performances of NB and LDA using voltage feature set

Evaluation metrics	Voltage feature set	
	NB	LDA
Accuracy	0.4000	0.3900
Precision	0.3093	0.2987
Geometric mean	0.4919	0.4767
F-measure	0.3072	0.2948
Sensitivity	0.3129	0.2993
Specificity	0.9016	0.8999

Table 2 displays the results of accuracy, precision, geometric mean, F-measure, sensitivity and specificity for the identification of the harmonic sources using NB and LDA for current feature set. Moreover, a configuration of performance measurements for current feature set can be found in Figure 5. As can be seen, the performances of NB and LDA were good in this work. The result shows that current features were able to discriminate the classes between harmonic sources, which contribute to the optimal performances in current work. From Table 2, LDA achieved better results compared to NB. The reason might be the high linearity of current features that gives the LDA to provide an excellent performance when identifying the harmonic sources.

Figure 6 and Figure 7 demonstrate the confusion matrix of NB and LDA for the identification of harmonic sources using current feature set. Note that the confusion matrix of NB and LDA for voltage feature sets are not shown due to their weak performances in this work. In these Figures, it shows that NB and LDA were able to identify the harmonic sources very well, which provided high class-wise accuracy of above 90%. With NB, the Hc-Hc and N-Hv were perfectly classified (100%). However, the N-Hc and Hv-N were lower than 95% class-wise accuracy, especially Hv-N (87.16%). On the other hand, LDA scored 100% class-wise accuracy on Hc-N, N-Hv, and 99.5% on Hc-Hc. Only two classes, N-Hc and Hv-Hv were lower than 95%, which was superior than NB in harmonic sources identification. The result affirms the LDA can usually provide higher performance than NB in harmonic sources identification.

Table 2. The performances of NB and LDA using current feature set

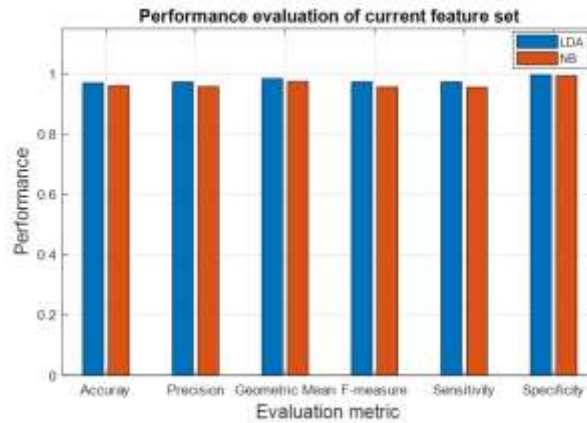| Evaluation metrics	Current feature set	
	NB	LDA
Accuracy	0.9600	0.9700
Precision	0.9576	0.9733
Geometric mean	0.9744	0.9842
F-measure	0.9561	0.9730
Sensitivity	0.9557	0.9729
Specificity	0.9934	0.9959

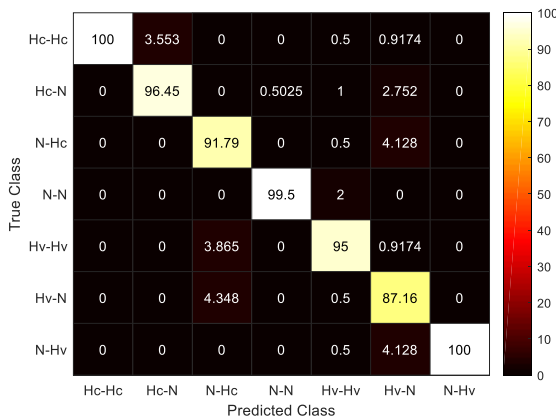Figure 5. Performance measurement of NB and LDA using current feature set

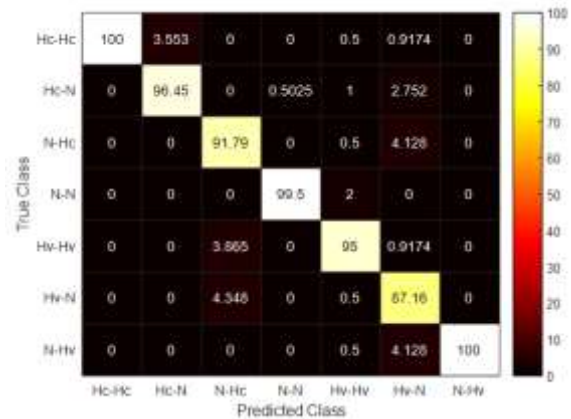Figure 6. Confusion matrix of NB using the current feature set

Figure 7. Confusion matrix of LDA using the current feature set

5. CONCLUSION

In this research two powerful classifiers, NB and LDA are proposed to identify and diagnose the harmonic sources in power quality system. In addition, the performances of two different feature sets, namely voltage feature set and current feature set are examined. Our experimental results show that voltage features were not good for harmonic sources identification, which cannot provide satisfactory performance. On the one hand, current features were excellent in identification process, which contributes to the highest discriminative power in this work. The results also prove that the performance of LDA was significantly better than NB in harmonic sources identification. In future, the LDA can be applied to other engineering applications.

REFERENCES
[1]   N. H. T. Huda, and A. R. Abdullah, "Power quality signals detection using S-transform," *Proceedings of the 2013 IEEE 7th International Power Engineering and Optimization Conference,* 2013.
[2]   A. R. Abdullah, N. M. Kassim, M. Manap, N. A. Ngatiman, and M. R. Yusoff, "Localization of Multiple Harmonic Sources for Inverter Loads Utilizing Periodogram," *J. Teknol.,* vol. 8, no. 2, pp. 87-91, 2016.
[3]   J. Too, A. R. Abdullah, and N. Mohd Saad, "Binary Competitive Swarm Optimizer Approaches for Feature Selection," *Computation,* vol. 7, no. 2, p. 31, 2019.
[4]   N. A. Abidullah, G. Z. Peng, and A. R. Abdullah, "A new two points method for identify dominant harmonic disturbance using frequency and phase spectrogram," *Int. Rev. Electr. Eng.,* vol. 9, no. 2, pp. 453–459, 2014.
[5]   A. R. Abdullah, T. Sutikno, M. Manap, M. R. Ab. Ghani, and M. R. Yusoff, "A critical review of time-frequency distribution analysis for detection and classification of harmonic signal in power distribution system," *Int. J. Electr. Comput. Eng.,* vol. 8, no. 6, pp. 4603-4618, 2018.
[6]   Too, Abdullah, and Mohd Saad, "A New Quadratic Binary Harris Hawk Optimization for Feature Selection," *Electronics,* vol. 8, no. 10, p. 1130, 2019.
[7]   J. Too, A. R. Abdullah, and N. M. Saad, "A new co-evolution binary particle swarm optimization with multiple inertia weight strategy for feature selection," *Informatics,* vol. 6, no. 2, 2019.
[8]   A. R. Abdullah, M. Manap, T. Sutikno, and M. R. A. Ghani, "An Improved of Multiple Harmonic Sources Identification in Distribution System with Inverter Loads by using Spectrogram," *Bull. Electr. Eng. Informatics,* vol. 7, no. 2, pp. 244-256, 2018.
[9]   A. R. Abdullah, et al., "Performance Evaluation of Real Power Quality Disturbances Analysis Using S-Transform," *Appl. Mech. Mater.,* vol. 752-753, pp. 1343-1348, 2015
[10]   a. R. Abdullah, N. a. Abidullah, N. H. Shamsudin, N. H. H. Ahmad, and M. H. Jopri, "Power Quality Signals Classification System Using Time-Frequency Distribution," *Appl. Mech. Mater.,* vol. 494-495, pp. 1889-1894, 2014.
[11]   A. R. Abdullah, M. Manap, T. Sutikno, and M. R. A. Ghani, "An identification of multiple harmonic sources in a distribution system by using spectrogram," *Bull. Electr. Eng. Informatics,* vol. 7, no. 2, pp. 244-256, 2018.
[12]   A. R. Abdullah, N. S. Ahmad, N. Bahari, M. Manap, A. Jidin, and M. H. Jopri, "Short-circuit switches fault analysis of voltage source inverter using spectrogram," *Int. Conf., Electr. Mach. Syst. (ICEMS),* pp. 1808-1813, 2013.
[13]   N. H. H. A. A. R. Abdullah, N. A. Abidullah, and N. H. Shamsudin, "Performance Verification of Power Quality Signals Classification System," *Appl. Mech. Mater.,* vol. 753, pp. 1158-1163, 2015.
[14]   N. A. Abidullah, A. R. Abdullah, N. H. Shamsudin, and N. H. T. H. Ahmad, "Real-time power quality signals monitoring system," *Proceeding-2013 IEEE Student Conf. Res. Dev. SCOReD,* pp. 433-438, 2013.
[15]   N. H. H. and Abidullah, N. A, Abdullah, A. R., Zuri_Sha'ameri, A., Shamsudin, and N.H., Ahmad, "Real-Time Power Quality Disturbances Detection and Classification System," *World Appl. Sci. J.,* vol. 32, no. 8, pp. 1637-1651, 2014.
[16]   A. R. Abdullah, M. Manap, M. F. Habban, and T. Sutikno, "An Accurate Classification Method of Harmonic Signals in Power Distribution System by Utilising S-Transform," *TELKOMNIKA (Telecommunication Comput. Electron. Control.,* vol. 15, no. 1, pp. 12-20, 2017.
[17]   W. Tee, M. R. Yusoff, A. R. Abdullah, N. S. N. Anwar, and H. Musa, "Spectrogram based window selection for the detection of voltage variation," *Int. J. Integr. Eng.,* vol. 11, no. 3, pp. 240-247, 2019.
[18]   A. R. Abdullah, T. Sutikno, M. Manap, M. R. A. Ghani, and A. S. Hussin, "A diagnostic analytics of harmonic source signature recognition by using periodogram," *Int. J. Electr. Comput. Eng.,* vol. 8, no. 6, pp. 5399-5408, 2018.
[19]   A. S. Hussin, A. R. Abdullah, T. Sutikno, N. M. Saad, and W. Tee, "Harmonic load diagnostic techniques and methodologies: A review," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 9, no. 3, pp. 690-695, 2018.
[20]   A. R. Abdullah, N. . Abidullah, and M. H. Jopri, "Analysis of Power Quality Disturbances Using Spectrogram and S-transform," *Int. Rev. Electr. Eng.,* vol. 3, pp. 611-619, 2014.
[21]   M. Manap, A. R. Abdullah, R. Karim, and M. R. Y. A. H. Azahar, "A verification of periodogram technique for harmonic source diagnostic analytic by using logistic regression," *TELKOMNIKA (Telecommunication Comput. Electron. Control.,* vol. 17, no. 1, pp. 497-507, 2019.
[22]   M. Manap, A. R. Abdullah, N. Z. Saharuddin, N. A. Abidullah, and N. S. Ahmad, "Performance Comparison of VSI Switches Faults Analysis Using STFT and S Transform," *Appl. Mech. Mater.,* vol. 785, pp. 210-214, 2015.
[23]   A. Abdullah, T. Sutikno, and M. Manap, "A Utilisation of Improved Gabor Transform for Harmonic Signals Detection and Classification Analysis," *Int. J.,* vol. 7, no. 1, pp. 21-28, 2017.
[24]   M. F. B. Habban, M. Manap, A. R. Abdullah, and T. Sutikno, "An evaluation of linear time frequency distribution analysis for VSI switch faults identification," *Int. J. Power Electron. Drive Syst.,* vol. 8, no. 1, pp. 1-9, 2017.
[25]   A. R. Abdullah, M. Manap, M. R. Yusoff, T. Sutikno, and M. F. Habban, "An improved detection and classification technique of harmonic signals in power distribution by utilizing spectrogram," *Int. J. Electr. Comput. Eng.,* vol. 7, no. 1, pp. 12-20, 2017.

[26] M. R. Yusoff, A. R. Abdullah, T. Sutikno, M. Manap, and A. S. Hussin, "An Analysis of Harmonic and Interharmonic Contribution of Electric Arc Furnace by Using Periodogram," *Int. J. Electr. Comput. Eng.,* vol. 7, no. 6, p. 3753, 2017.

[27] A. Stetco et al., "Machine learning methods for wind turbine condition monitoring: A review," *Renew. Energy,* vol. 133, pp. 620-635, 2019.

[28] R. Kumar, B. Singh, and D. T. Shahani, "Recognition of Single-stage and Multiple Power Quality Events Using Hilbert-Huang Transform and Probabilistic Neural Network," *Electr. Power Components Syst.,* vol. 43, no. 6, pp. 607-619, Apr. 2015.

[29] D. Srinivasan, W. S. Ng, and A. C. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Trans. Power Deliv.,* vol. 21, no. 1, pp. 398-405, 2006.

[30] P. Thakur and A. K. Singh, "Signal Processing and AI Based Diagnosis of Power Quality Disturbances : A Review," *Int. Conf. Energy Econ. Environ.,* pp. 1-6, 2015.

[31] M.-Y. Chen and B.-T. Chen, "Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform," *Appl. Soft Comput.,* vol. 14, pp. 156-166, 2014.

[32] J. Too, A. R. Abdullah, N. Mohd Saad, N. Mohd Ali, and H. Musa, "A detail study of wavelet families for EMG pattern recognition," *Int. J. Electr. Comput. Eng.,* vol. 8, no. 6, pp. 4221-4229, 2018.

[33] J. Too, A. R. Abdullah, and N. M. Saad, "Hybrid Binary Particle Swarm Optimization Differential Evolution-Based Feature Selection for EMG Signals Classification," *Informatics,* vol. 8, no. 3, p. 79, 2019.

[34] A. R. Abdullah, M. Manap, M. R. Yusoff, and T. Sutikno, "A fast localization of multiple harmonic sources for rectifier loads by utilizing periodogram," *Telkomnika (Telecommunication Comput. Electron. Control.,* vol. 15, no. 1, 2017.

[35] A. Abdullah, M. Jopri, and M. Manap, "An Improved Spectrogram to Identify Multiple Harmonic Sources in Distribution System with Inverter Loads," *Proc.,* vol. II, pp. 0-5, 2017.

[36] J. Too, A. R. Abdullah, and N. M. Saad, "Classification of Hand movements based on discrete wavelet transform and enhanced feature extraction," *Int. J. Adv. Comput. Sci. Appl.,* vol. 10, no. 6, pp. 83-89, 2019.

[37] J. Too, A. R. Abdullah, N. M. Saad, and N. M. Ali, "Feature selection based on binary tree growth algorithm for the classification of myoelectric signals," *Machines,* vol. 6, no. 4, 2018.

[38] A. R. Abdullah, G. Z. Peng, and S. A. Ghani, "A new vector draft method for harmonic source detection at point of common coupling," *Proc. 2014 IEEE 8th Int. Power Eng. Optim. Conf. PEOCO 2014,* pp. 110-114, 2014.