# A Novel Feature Selection Method for the Detection and Classification of Power Quality Disturbances

Aslam Shaik [1], A. Srinivasula Reddy [2]

[1] JNTUH, Hyderabad, India.
[2] CMREC, Hyderabad, India.
[1] Corresponding author email: aslam.careers@gmail.com

***Abstract*** **- We propose an automated recognition method using entropy and correlation characteristics to recognize single and combined Power Quality Disturbances (PQD). We determine the linear and non-linear dependencies among power quality disturbance signals to extract the most relevant features and information. Since the provision of a perfect knowledge about different PQD signals is required for recognition, we try to extract the perfect and near perfect features for every disturbance. We then: i) evaluate the joint occurrence probabilities and propose a new method of 'joint mutual information maximization', ii) evaluate the cross-correlation properties and propose a second stage feature selection method. Multi-class Support Vector Machine algorithm is formulated for classification. Extensive simulations are carried out by incorporating different PQD signals like Sag, Swell, Flicker, Transient, etc. to check the performance of the proposed method. The performance is evaluated through a performance metrics for Accuracy, Detection rate and False Alarm Rate under different environments. The results show that the proposed method can effectively recognize single and combined PQ disturbances.**

***Keywords - PQD signals, Mutual Information, Cross-Correlation, Support Vector Machine, Accuracy***

## I. INTRODUCTION

In the recent years, the Power Quality (PQ) issues have gained a lot of research interest due to an increased demand of microprocessor and/or power electronics based non-linear controlled loads. Due to the power quality problems, these devices may also malfunction and deviates from the regular behavior [1]. Further, the electricity is an essential resource and also a commercial product that is evaluated not only by its reliability and also by its quality. A low quality power causes too much damage for sensitive electronic equipment like household devices and it also have a severe effect in the smart grids, high speed trains and new energy outlets etc. Hence in order to find the power quality disturbances, it is very important to analyze and recognize these PQ disturbances.

PQ disturbances cover a wide frequency range with significant variations in the magnitude and can be non-stationary or stationary signals. Based on the occurrence of disturbances, initially they are two types; one is single disturbances and second is mixed or combined disturbances. Voltage sag, voltage swell, voltage interruption, transient impulse, oscillation transients, harmonics and flicker are some examples for single PQ disturbances. The mixed disturbances are said to be occurred if more than one PQ disturbances are occurred at a time. Voltage sag with harmonics and voltage swell with harmonics are the two best examples for mixed type disturbances. Actually, the PQ disturbances are often mixed disturbances, which are occurred in the combined from different disturbances, and the various components, and these disturbances are more complex. In addition, the interaction between different single disturbances may cause the aliasing and even failure characteristics, and will result in the wrong analysis effecting on the recognition accuracy. Hence there is a necessity to design an efficient PQD detection framework which improves the recognition accuracy resulting to a perfect decision to accomplish a correct remedy in the power systems and power related appliances [2], [3].

Based on the methodology accomplished for PQD detection, the overall architecture is formulated into two phases, preprocessing and classification. In the preprocessing phase, the required features are extracted from the PQD signals and in the classification phase, the extracted features are compared with a pre-evaluated feature set and finally classified into one of the PQD class. Though there has been a vast research carried out over PQD detection, most of the approaches not focused on the features significance in the classification phase which results in the extra complexity and processing time. Furthermore, the recent approaches didn't consider the dependency between the features of different classes.

Focusing towards the extraction of relevant features by eliminating the unwanted redundant features from PQD signals, this paper proposes two new feature selection methods based on the entropy and cross-correlation properties. Considering the Mutual Information between the PQD signals, the effect of one signal occurrence over the other signal can be analyzed and by including this process for feature extraction makes the detection system robust to various power related applications. Further the set of features are processed through support vector machine algorithm for classification purpose. Experimental Simulations are carried out over different PQD signals under two different cases, namely normal and noisy, to prove the performance enrichment.

Rest of the paper is organized as follows: Section II describes the literature survey details. The details of power quality disturbances and their control parameters are illustrated in section III. The details or proposed feature selection and detection mechanism is illustrated in section IV. Experimental evaluation is described in section V and finally the conclusions and future scope are described in section VI.

## II. LITERATURE SURVEY

Detection of Power quality disturbances is so important and to do so, different authors focused on the detection of different PQDs. Some of the approaches focused only on the feature extraction techniques and they didn't given much importance for the classifier. On the other hand, some approaches focused only on the classifier and didn't consider the features importance. Based on these strategies, the overall literature survey is classified into two classes, feature extraction approaches and classification approaches. A brief outline about these two classes is illustrated in the following subsections.

### *A. Feature Extraction Approaches*

Under this class of approaches, the feature extraction is done in the spatial domain and in transform domain. In the spatial domain, the characteristics of PQD signals are evaluated with respect to the time whereas in the transform domain, the characteristics are extracted with respect the frequency.

The S-Transform (ST) is unique in that it provides frequency-dependent resolution while maintaining a direct relationship with the Fourier spectrum. The features obtained from ST are distinct, understandable, and immune to noise. Taking advantage of STs, Zhao et.al.,[4] proposed a new method of detecting and classifying power-quality disturbances. Totally, through this method eight types of single power disturbance and two types of complex power disturbance are recognized. However, the computational complexity of ST is observed to be high and also less efficiency in the time-frequency traversal. To overcome these two issues, a modified version of ST, named as Modified incomplete S-transform (MIST) is developed by Li et.al.,[5]. Wavelet transform is a multi-scale time-frequency transform with multi-resolution analysis, which is widely used to analyze PQ disturbance signals [6].Based on these advantages of wavelet transform, a new approach was developed by Saurabh Kamble [7], to detect, localize, and investigate the feasibility of classifying various types of PQDs. Voltage sag, swell, transient and harmonics are the main PQ problems mainly focused in [7]. The approach is based on wavelet transform analysis, particularly the discrete wavelet transform (DWT). The key idea is to decompose a given disturbance signal using DWT which represent a smoothed version and a detailed version of the original signal. These decomposed signals are used to extract features using many mathematical operations like peak, variance, mean deviation and skewness.

Further combining the advantages of both the ST and DWT, some more approaches are developed [8-11].Compared with traditional wavelets; the wavelet packet has better properties. In addition, wavelet packet transform can yield more frequency sub-bands. However, due to the complexity of combined PQ disturbance signals, the actual research shows that there are problems with feature extraction, such as energy leakage and aliasing between adjacent scales with the application of various methods based on wavelet transform. Furthermore, the multiwavelet is another form of wavelet which has better properties than traditional wavelets. Multiwavelet packet transformation has more high-frequency information. In [12], multiwavelet packet entropy based feature selection is developed to extract the required set of features from transmission lines. Multiwavelet packet energy entropy, time entropy, Shannon singular entropy, and Tsallis singular entropy are defined as the feature extraction methods of transmission line fault signals. In [14, 15], due to the advantages of Tsallis singular entropy, some more extension are developed namely, Tsallis wavelet time entropy (TWTE), Tsallis wavelet energy entropy(TWEE), wavelet packet Tsallis singular entropy (WPTSE) and Tsallis wavelet singular entropy (TWSE).The application of these entropies has gained an increased efficiency in the analysis of PQ disturbances through wavelet transform. Further so many approaches are discovered by analyzing the transformed characteristic of PQD signals to detect them various applications. However, the accomplishment of transformation over the PQD signals builds an extra complexity.

On the other hand, the spatial domain approaches were also gained a lot of research interest in the PQD analysis. Considering the linear and non-linear dependencies between the PQD signals, a Mutual information based feature selection (MIFS) [16, 18] approach is proposed by Moravej et.al., [17]. In [17], to get optimal features for the classifier two stage of features election has been used. In first MIFS and in the second stage correlation feature selection (CFS) techniques are used for feature extraction from signals to build distinguished patterns for classifiers. MIFS can reduce the dimensionality of inputs, speed up the training of the network and get better performance and with CFS can get optimal features. Based on single channel independent component analysis (SCICA), a new method was developed in [19] for single and multiple power quality disturbance classification. This method decouples the power system signal into its independent components, which are classified by specialized classifiers. The classifier outputs are combined by using a logic that gives the final classification

[20]. Five classes of single disturbances and twelve of multiple disturbances are considered.

*B. Classification Approaches*

Design of an optimal classifier is also an important in the detection of PQD signals in power systems. A classifier needs to be designed in such a way that it should be robust in the PQD detection for different power related applications. Furthermore the complexity is also needs to be considered in the classifier design. Having the reduced computational complexity, the classifier can achieve effective results even in the case of larger volume datasets and different unknown signals. According to the earlier studies Support Vector machine has proved as an efficient classifier which achieves increased detection accuracy in the PQD detection.

A neural-network based dual methodology was proposed byValtierraet.al., [21] to detect and classify single and combined PQDs. In this dual methodology, the first one is an adaptive linear network which estimates the inter-harmonics and harmonics that allows the computation of total harmonic distortion index (HDI) and the root-mean square (RMS) voltage. These indices help in the detection of swells, sags, outages, inter-harmonics and harmonics. On the other hand, a Feed-Forward Neural Network (FFNN) [29] accomplished for pattern recognition helps in the detection of spikes, flicker, notching and oscillatory transients.

Further to improve the PQ disturbances classification, the SVM is combined with decision tree and a new method is developed by Ray et.al., [22]. Several features are extracted through the hyperbolic ST, and only a set of optimal features are extracted through the Genetic Algorithm (GA). Considering different application scenario, this method is simulated through modified Nordic-32 bus system Photovoltaic systems and wind energy systems. Sparse Signal Decomposition (SSD) based a new PQ disturbances detection and classification method was developed by Manikandan et.al., [23] on over-complete hybrid dictionary (OHD) matrix. Initially, the PQ signal is decomposed through SSD with an OHD matrix containing sinusoidal and elementary waveforms. Through the SSD decomposition with OHD, the impulsive and oscillatory features waveform distortions such as notching and harmonics are adequately obtained in the resultant detail signal. The remaining features in the approximation signals includes the fundamentals of dc-offset, swells, sags, flicker, and interruptions. After obtaining the required features, they are classified through the hierarchical decision tree algorithm into single and multiple PQ disturbances.

Incorporating two transforms, Hyperbolic ST and DWT, an integrated feature selection approach is proposed by Hajian, M and Foroud A [24]. Moreover, a new efficient feature selection mechanism, namely Orthogonal Forward Selection (OFS) by incorporating Gram Schmidt and forward selection is applied to attain the best subset of features. Different classifiers are accomplished to determine the best classifier. In this automatic approach, the variable parameters of different classifiers are optimized through the most popular meta-heuristic algorithm, Particle Swarm Optimization (PSO).Combining the ensemble empirical mode decomposition (EEMD) and multi-label learning, a new method was developed by Zhigang et.al., [25] for the classification of power quality complex disturbances. EEMD [26] is adopted to extract the features of complex disturbances, which is more suitable to the non-stationary signal processing. Rank wavelet support vector machine (rank-WSVM) [27, 28] is proposed to apply in the classification of complex disturbances. A new method for the classification of PQ disturbances was proposed by Subra De and Sudipta [30] through the cross-correlation combining with Fuzzy logic. Due to the accomplishment of fuzzy logic, the computational complexity of this method is observed to be less and totally seventeen types of PQ disturbances are recognized. The proposed scheme is also applied in IEEE 33-bus distribution system and validated by a real-time simulator. A method for detecting and classifying transmission line faults using cross correlation and k-nearest neighbor has been presented in [31].

## III. POWER QUALITY DISTURBANCE SIGNALS

Due to occurrence of different disturbances in the power related applications, there are so many types of disturbances, occurred in an alone and also in the combined form. In this section, the most popular and identified disturbances are only illustrated and they are only considered for the classification purpose. According to the scientific standards, every disturbance signal is represented through a mathematical representation and they are explained in the table.1. A pure sine wave with frequency 50 Hz and magnitude at 1.0 p.u. as well as the another PQ disturbance signals such as sag, swell, harmonics, interruption, sag with harmonics, swell with harmonics, flicker, notch, spike, and oscillatory transients are generated through Matlab simulation software. The total recording time of the signal is 0.4 s. The reference frequency of pure sine wave is 50 Hz. The PQ disturbance signal generation models and their control parameters are shown in Table 1.

TABLE.1 POWER QUALITY DISTURBANCE SIGNALS AND ITS CONTROLLING PARAMETERS

| PQD | Class | Formula | Parameter Variations |
|---|---|---|---|
| Normal | C1 | $s(t) = \sin(\omega_n t)$ | $\omega_n = 2\pi \times 50$ rad/sec |
| Swell | C2 | $s(t) = [1 + \alpha(u(t - t_1) - u(t - t_2))] \sin(\omega_n t)$ | $0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$ |
| Sag | C3 | $s(t) = [1 - \alpha(u(t - t_1) - u(t - t_2))] \sin(\omega_n t)$ | $0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$ |
| Flicker | C4 | $s(t) = [1 + \alpha \sin(2\pi\beta t)] \sin(\omega_n t)$ | $0.1 \leq \alpha \leq 0.2, 5Hz \leq \beta \leq 20Hz$ |
| Interruption | C5 | $s(t) = [1 - \alpha(u(t - t_1) - u(t - t_2))] \sin(\omega_n t)$ | $0.9 \leq \alpha \leq 1, T \leq t_2 - t_1 \leq 9T$ |
| Oscillatory transient | C6 | $s(t) = \sin(\omega_n t) + \alpha \exp(-(t - t_1)/\tau)(u(t - t_2))\ \sin(2\pi f_n t)$ | $0.1 \leq \alpha \leq 0.8$ , $0.5T \leq t_2 - t_1 \leq 3T$ , $300Hz \leq f_n \leq 900Hz, 8ms \leq \tau \leq 4m$ |
| Harmonics | C7 | $s(t) = \alpha_1 \sin(\omega_n t) + \alpha_3 \sin(3\omega_n t) + \alpha_5 \sin(5\omega_n t) + \alpha_7 \sin(7\omega_n t)$ | $0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \sum \alpha_i^2 = 1$ |
| Swell with harmonics | C8 | $s(t) = \left[1 + \alpha\left(u(t - t_1) - u(t - t_2)\right)\right] \times [\alpha_1 \sin(\omega_n t) + \alpha_3 \sin(3\omega_n t) + \alpha_5 \sin(5\omega_n t) + \alpha_7 \sin(7\omega_n t)]$ | $0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$, $0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \sum \alpha_i^2 = 1$ |
| Sag with harmonics | C9 | $s(t) = \left[1 - \alpha\left(u(t - t_1) - u(t - t_2)\right)\right] \times [\alpha_1 \sin(\omega_n t) + \alpha_3 \sin(3\omega_n t) + \alpha_5 \sin(5\omega_n t) + \alpha_7 \sin(7\omega_n t)]$ | $0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$, $0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \sum \alpha_i^2 = 1$ |
| Notch | C10 | $s(t) = \sin(\omega_b t) - sign(\sin(\omega_b t)) \times \sum_{n=0}^{9} k\left[u\left(t - (t_1 + 0.2n)\right) - u(t - (t_2 + 0{,}02n))\right]$ | $0.1 \leq k \leq 0.4, 0 \leq t_1, t_2 \leq 0.5T$, $0.01T \leq t_2 - t_1 \leq 0.05T$ |
| Spike | C11 | $s(t) = \sin(\omega_b t) + sign(\sin(\omega_b t)) \times \sum_{n=0}^{9} k\left[u\left(t - (t_1 + 0.2n)\right) - u(t - (t_2 + 0{,}02n))\right]$ | $0.1 \leq k \leq 0.4, 0 \leq t_1, t_2 \leq 0.5T$, $0.01T \leq t_2 - t_1 \leq 0.05T$ |

## IV. PROPOSED DETECTION AND CLASSIFICATION APPROACH

The complete details of proposed approach are illustrated in this section. To achieve improved classification accuracy, this paper proposes two new feature selection techniques based on the Mutual Information and cross correlation. Here the proposed feature selection technique based Mutual information is named as Joint Mutual Information Maximization based Feature Selection (JMIMFS) and the feature selection based on cross-correlation is named as correlation based feature selection (CFS). After obtaining the required subset of features from every class of PQ signal, Multi-class Support Vector Machine (MC-SVM) [37] is accomplished for classification. Since the MIFS a non-linear feature extraction technique, it extracts the set of optimal features in the non-linear signals also and the CFS is for linearly dependent signals.

### *A. Feature Selection*

Feature selection (FS) is an important aspect in the PQ detection and classification system. In earlier there are so many methods are developed to achieve an efficient performance in the feature selection purposes in various applications. Among the earlier developed FS methodologies, MIFS approaches have gained more significant results in the detection performance. The MI was first developed by Battiti in 1994 [16], also known as a first order incremental search algorithm. Battiti proposed MI to select the more relevant features form the initial set of 'n' features. Instead of calculating the Joint MI between the selected features and the class label, Battiti's MI evaluates MI between the candidate feature and the class, relationship between the candidate feature and the already selected features. Further there are so many variants are proposed based on MI such as MIFS-U [32], mRMR [33], NMIFS [34], MIFS-ND [35] and JMI [36]. However the following drawbacks are observed with the earlier MIFS approaches.

1. Class Irrelevancy: In the earlier methods, the redundancy is measured based on the MI value between the candidate feature and features in the selected feature subset, but didn't considered the class label. If the MI between the candidate feature and selected feature in the subset is less, then the candidate features is considered as redundant feature, but this dissertation is wrong because the redundant candidate feature may share different information with the another class.

2. Over Estimation of feature significance: in the case of high correlation of candidate feature with one or some pre-selected features, the candidate feature is assumed to share more information about the features selected in the subset, but at the same time the candidate feature can be an independent feature from the majority features in the

selected feature subset. In such condition, the value of goal function is high, despite the redundancy of the candidate feature to some features within the subset. This problem was occurred in the methods like MIFS-U, mRMR, MNIFS, MIFS-ND which follows a cumulative summation and forward search algorithm search to approximate the solution.

*1. JMIMFS*

In this paper, a new FS method is proposed based on the MI based FS, JMIM. JMIM is a combined from the Joint MI (JMI) and Maximum of the Minimum (MIM). JMIM is aimed to address the above problems, class irrelevancy and the Over estimation of feature significance, which occurs when the cumulative summation is employed.

The FS process should be in such a way that for a given Full feature set F of size N, it needs to select a subset of features S, S⊆F, with dimensions K, K≤N, by which the classification accuracy should be equal or better compared to the classification accuracy obtained through the full set of feature F. Simply it can also be defined as a FS that extracts the features which have maximum MI with the class label, i.e., I(S; C).

Based on these aspects, the feature relevance is defined as, for an already selected feature subset, S, the feature $f_i$ is said to be more relevant than the feature $f_j$ if the MI between $f_i$ and S with respect to the class C $(I(f_i, S; C))$ is greater than the MI between the feature $f_j$ and S with respect to the class C $\left(I(f_j, S; C)\right)$, simply, $(I(f_i, S; C)) > \left(I(f_j, S; C)\right)$.

Further the feature relevance can also be defined through the Joint MI. Let F be the full set of features, S be the subset of features which was already selected for the Feature set F. Let a feature $f_i$, $f_i \in F - S$, and $f_s \in S$, the m-Joint MI is defined as the MI between $f_i$ and the features present in the already selected feature subset S. The minimum value of m-Joint MI is referred as minimum joint MI, i.e., $\min_{s=1,2,\dots,K} I(f_i, f_s; C)$. A larger value of joint MI of $f_i$ and the features in the subset S denotes a high relevance with the class label C. Further a larger value of joint MI also denote that the m-joint MI of other features, $f_j$, $f_i$ and $f_j \in F - S$ denotes the minimum joint MI between the features $f_j$ and $f_i$ . Simply it denotes that, compared to the feature $f_i$, the feature $f_j$, shares less information towards the class label C. According to the above definitions, the feature which shares maximum information is said to be more relevant.

Further a new definition is given for redundancy from the given set of features F, and a selected feature subset S, a feature $f_i$ is said to redundant to the selected feature subset S if it doesnot share new information with the class C. If the feature $f_i$ is highly correlated with a feature $f_s$, $f_s \in S$, then the probability of mass functions of $f_i$, $f_s$ and $(f_i, f_s)$ are equal, i.e., $P(f_i) \cong P(f_s) \cong P(f_i, f_s)$.

Based on the above discussions, to overcome the problem of over estimation feature significance, this work accomplished a new method called JMIM to select the optimal feature set by which the classification accuracy increase with less number of features extracted from dataset. It is a combined form of Joint MI and maximum of minimum, through which the most relevant features are chosen. The new criterion for the FS according to the JMIM is formulated as:

$$f_{JMIM} = \arg \max_{f_i \in F-S} \left( \min_{f_s \in S} (I(f_i, f_s; C)) \right) \quad (1)$$

Where

$$I(f_i, f_s; C) = I(f_s; C) + I\left(f_i; {}^{C}/_{f_s}\right) \quad (2)$$

$$I(f_i, f_s; C) = H(C) - H\left({}^{C}/_{f_i}, f_s\right) \quad (3)$$

$$I(f_i, f_s; C) = \left[-\sum_{c\in C} p(c)\log(p(c))\right] - \left[\sum_{c\in C} \sum_{f_i\in F-S} \sum_{f_s\in S} log\left(\frac{p(f_i, f_s, {}^{c}/_{f_s})}{p({}^{f_i}/_{f_s})p({}^{c}/_{f_s})}\right)\right] \quad (4)$$

This method follows the iterative forward greedy search algorithm to find there levant feature subset of size k within the feature space. The JMIMFS reduces the almost all redundant features from the PQ signals and preserves only the features which are high informative and contributes to the class. Approximately, the first stage feature selection technique reduces 40% of features from the full set of features.

*2. CFS*

After extracting a set of features in the first stage, they are further subjected to the second stage feature selection, CFS. Here the correlation is measured between the features to find the similarity between them. Specifically, the CFS evolves the linearly dependent features, the obtained subset from the first stage are linearly related to each other. Finding those linear dependencies makes the detection system computationally inexpensive and also achieves more classification accuracy. To analyze the obtained subset further deeply, a new measure called Pearson product-moment Correlation Linear Coefficient (PCLC) is evaluated which derives the linear correlation between the features. This measure is fast and accurate is measuring the correlation between random linearly dependent variables and also it is insensitive to non-linear correlations.

Given two variables X and Y, the PCLC (X; Y) or r (X; Y) is measured as:

$$r(X;Y) = \frac{A}{B.\ C} (5)$$

Where:

$$A = n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i - \sum_{j=1}^{n} y_j \quad (6)$$

$$B = \sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{j=1}^{n} x_j\right)^2} \quad (7)$$

$$C = \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{j=1}^{n} y_j\right)^2} \tag{8}$$

The value of $r(X;Y)$ falls into a definitely closed interval [-1, 1]. A value of -1 or 1 indicates a high correlation between two variables. A value close to zero indicates the weak relationship between them. Further to select a feature with maximum correlation the PCLC based feature selection is formulated as:

$$C_r = \underset{f_i \in F}{\operatorname{argmax}} \left( r(f_i; C) - \frac{1}{|S|} \sum_{f_s \in S} \frac{corr(f_i; f_s)}{corr(f_i; C)} \right) \tag{9}$$

Form the above expression, the feature which represents the maximum correlation is chosen as a relevant feature for the class C and the feature which does not has a maximum correlation is removed or eliminated. This PCLC based feature selection is carried out over subset of features to further select the optimal features in a specific disturbances class. In the training phase, for every attack class, the feature with maximum correlation is extracted and trained through MC-SVM [37]..

## V. SIMULATION RESULTS

### A. Experimental Data

To simulate the proposed PQD framework, MATLAB software was used. Initially the PQ signals are generated according to the mathematical representations shown in table.1. By varying the control parameters in every mathematical formula, for every class of PQD, number of signals is generated. For example, the control parameters for swell signal is α. For a given range constraints of α, it was varied α=0.1,0.2,0.3,…,0.8. Further the range constraint of time (T) is 0.4 to 3.6. With an increment of T = 0.1, the total number of possible values can be generated is 33 and for every T values, the α is varied. Based on these observations, the total number of possible swell signals can be generated are 33×8=264. Similarly, for every PQD class, number of signals is generated and the signals generated for every class are represented in table.2. All these signals are generated by varying their respective control parameters. These signals are divided into two groups, training set and testing set. The total number of signals divided as training and testing signals are represented in table III.

TABLE II. SIGNALS GENERATED WITH VARYING CONTROL PARAMETERS

| 1. PQD | 2. Varying parameters | 3. Total signals |
|---|---|---|
| Normal | 4. $\omega_n = 2\pi \times 50$ rad/sec | 5. 10 |
| Swell | 6. T=0.4, 9×T=3.6, $\alpha = 0.1,0.2,0.3,\dots,0.8$ | 7. 33×8=264 |
| Sag | 8. T=0.4, 9×T=3.6, $\alpha = 0.1,0.2,0.3,\dots,0.9$ | 9. 33×9=297 |
| Flicker | 10. $\alpha = 0.1,0.2$, $\beta = 5,6,7,\dots,20$ | 11. 16×2=32 |
| Interruption | 12. T=0.4, 9×T=3.6, $\alpha = 0.9$ and 1.0 | 13. 33×2=66 |
| Oscillatory transient | 14. T=0.5×T to 3×T=0.2 to 1.2, $f_n = 300,400,\dots.,900$, 15. $\alpha = 0.1,0.2,0.3,\dots,0.8$ | 16. 6×8×6=288 |
| Harmonics | 17. $\alpha_3, \alpha_5, \alpha_7$=0.05, 0.1, 0.15 | 18. 3×3×3=27 |
| Swell with harmonics | 19. $\alpha_3, \alpha_5, \alpha_7$=0.05, 0.1, 0.15, $\alpha = 0.1,0.2,0.3,\dots,0.8$ | 20. 27×8=216 |
| Sag with Harmonics | 21. $\alpha_3, \alpha_5, \alpha_7$=0.05, 0.1, 0.15, $\alpha = 0.1,0.2,0.3,\dots,0.9$ | 22. 27×9=243 |
| Notch | 23. k=0.1,0.2,0.3,0.4, $t_1$=0,0.1,0.2,…,0.5, $t_2$=0,0.1,0.2,…,0.5 | 24. 4×6×6=144 |
| Spike | 25. k=0.1,0.2,0.3,0.4, $t_1$=0,0.1,0.2,…,0.5, $t_2$=0,0.1,0.2,…,0.5 | 26. 4×6×6=144 |
| **27. Total** | | **28. 1732** |

TABLE III. TRAINING AND TESTING SET CONSIDERED FOR SIMULATION

| 29. PQD | 30. Class Label | 31. Total signals | 32. Training Set | 33. Test Set |
|---|---|---|---|---|
| Normal | C1 | 34. 10 | 35. 7 | 36. 3 |
| Swell | C2 | 37. 264 | 38. 192 | 39. 72 |
| Sag | C3 | 40. 297 | 41. 213 | 42. 84 |
| Flicker | C4 | 43. 32 | 44. 21 | 45. 11 |
| Interruption | C5 | 46. 66 | 47. 45 | 48. 21 |
| Oscillatory transient | C6 | 49. 288 | 50. 209 | 51. 79 |
| Harmonics | C7 | 52. 27 | 53. 18 | 54. 9 |
| Swell with harmonics | C8 | 55. 216 | 56. 159 | 57. 57 |
| Sag with Harmonics | C9 | 58. 243 | 59. 175 | 60. 66 |
| Notch | C10 | 61. 144 | 62. 108 | 63. 36 |
| Spike | C11 | 64. 144 | 65. 108 | 66. 36 |
| **Total** | | **67. 1732** | **68. 1255** | **69. 477** |

### *B. Performance Evaluation*

In this paper, the Accuracy, Precision, DR, FAR and F-Score are considered to evaluate the performance of proposed approach. The basis for these metric evaluations is confusion matrix and it is represented in table IV.

TABLE.IV. SAMPLE CONFUSION MATRIX

| | | Predicted | |
|---|---|---|---|
| | | Normal | Attack |
| Actual | Normal | TP | FN |
| | Attack | FP | TN |

Based on the obtained TP, TN, FP and FN values from the confusion matrix, performance metrics are evaluated and the respective mathematical representation is given as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Detection\ rate\ or\ Recall = \frac{TP}{TP+FN} \quad (12)$$

$$False\ Alarm\ Rate = \frac{FP}{TN+FP} \quad (13)$$

$$F-Score = \frac{2*Recall*Precision}{Recall+Precision} \quad (14)$$

Where:
TP= True Positives, TN = True Negatives, FP= False Positives and FN = False Negatives.

### *C. Results*

Initially the complete train set is subjected to the feature extraction through the proposed two feature selection approaches, JMIMFS and CFS. Finally the obtained optimal feature set of every PQD signal was trained through SVM algorithm. Here totally, 1298 signals are processed for training and 434 signals are processed for testing. Further the test set is subjected to testing process and the classified results are formulated in the following table IV. To check the efficiency of proposed approach, a further simulation carried out in the presence of noise. In this simulation, all the disturbance signals are subjected to noise addition and then processed for testing. The obtained classification results in this case are represented in the following table V. Further to check performance of proposed approach, the test signals are subjected to noise contamination and those noisy PQ disturbances are processed for testing. The obtained classification results under the noisy experiments are displayed in table VI.

TABLE.V. CLASSIFICATION RESULT OF TESTING SIGNALS (CONFUSION MATRIX)

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1 | **3** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| C2 | 0 | **70** | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 72 |
| C3 | 0 | 1 | **81** | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 84 |
| C4 | 1 | 0 | 0 | **9** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 11 |
| C5 | 0 | 1 | 2 | 0 | **18** | 0 | 0 | 0 | 0 | 0 | 0 | 21 |
| C6 | 1 | 0 | 0 | 1 | 0 | **75** | 0 | 0 | 1 | 1 | 0 | 79 |
| C7 | 0 | 1 | 0 | 0 | 0 | 0 | **8** | 0 | 0 | 0 | 0 | 9 |
| C8 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | **55** | 1 | 0 | 0 | 57 |
| C9 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | **63** | 0 | 1 | 66 |
| C10 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | **34** | 0 | 36 |
| C11 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | **32** | 36 |

TABLE VI. CLASSIFICATION RESULT OF TESTING SIGNALS IN THE PRESENCE OF NOISE (CONFUSION MATRIX)

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1 | **3** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| C2 | 0 | **67** | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 72 |
| C3 | 0 | 1 | **79** | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 84 |
| C4 | 1 | 0 | 0 | **8** | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 11 |
| C5 | 0 | 1 | 2 | 0 | **16** | 0 | 0 | 1 | 1 | 0 | 0 | 21 |
| C6 | 0 | 1 | 0 | 1 | 0 | **73** | 0 | 2 | 1 | 0 | 1 | 79 |
| C7 | 0 | 1 | 0 | 0 | 1 | 0 | **7** | 0 | 0 | 0 | 0 | 9 |
| C8 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | **53** | 1 | 1 | 0 | 57 |
| C9 | 0 | 2 | 0 | 1 | 0 | 0 | 1 | 0 | **61** | 0 | 1 | 66 |
| C10 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | **32** | 1 | 36 |
| C11 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | **31** | 36 |

Based on the obtained classification results for all types of PQ disturbances, the performance is measured through the performance metrics according to the formulae specified in the above subsection. Further to verify the performance enrichment of proposed framework, the obtained results are compared with the results obtained through conventional approaches, Cross-Correlation combined with fuzzy logic (CCS + Fuzzy) [30], Cross-Correlation combined with K-nearest Neighbor (CCS + K-NN) [31], Flexible entropy based feature selection combined with MC-SVM (FEFS+MC-SVM) [37] and Mutual Information based feature selection combined with SVM (MIFS + SVM) [17]. The measured Accuracy, Detection rate, precision, False Alarm Rate and F-score are represented in the following figures.

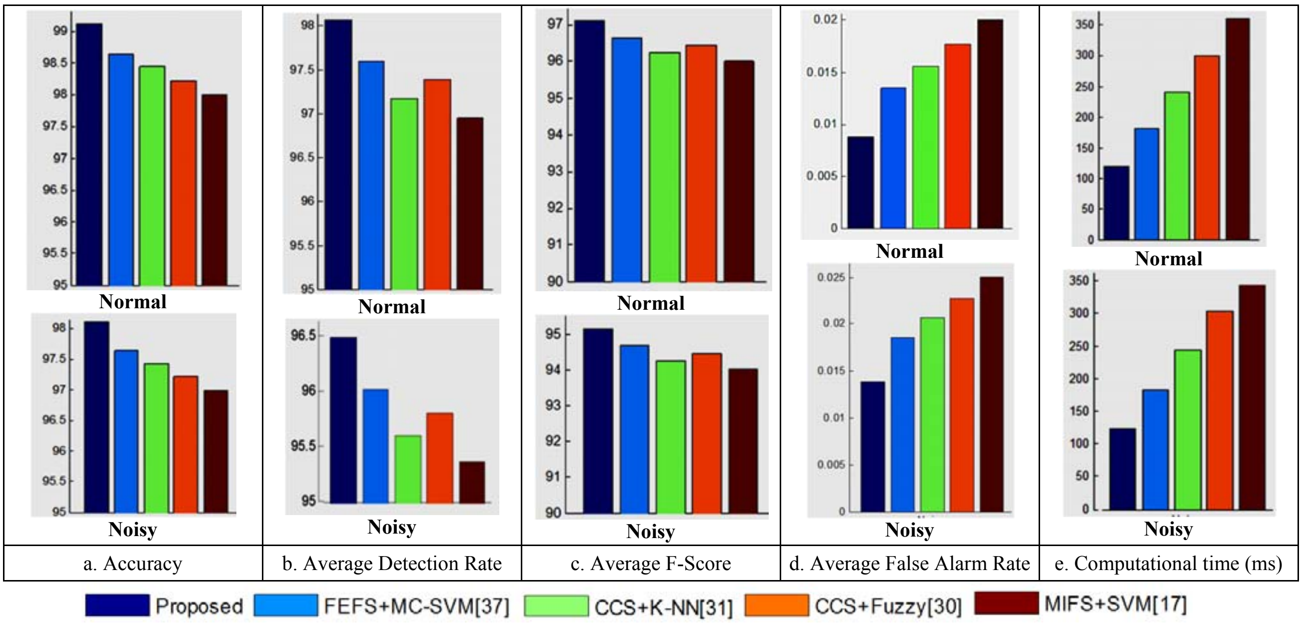


Figure.1 Results of comparison.

Figure.1 describes the comparative results analysis between the proposed and conventional approaches with respect to the Receiver Operating Characteristics (ROC) performance metrics. Under this section, the accuracy obtained after the classification of all PQD signals under two test cases, normal and noisy, are represented in figure 1-a. As it can be seen, the accuracy of proposed approach is high compared to the conventional approach for both test cases. Since the proposed two stage feature selection mechanism tries to extract only the informative features for every PQD signals and they are only trained to the system. The obtained optimal subset of features helps in the provision of perfect discrimination between the PQD signals. In the conventional approach, FEFS + MC-SVM, the feature extraction technique extracts the subset of features from full set based on the Mutual Information But the redundant features reduced through this method are not selected based on the joint relevancy of different classes. This method won't considered the joint MI between the selected feature subset and the class label by which some necessary features are also got eliminated and some unnecessary feature are preserved. Next, the conventional approaches, CCS + K-NN and CCS + Fuzzy are able to extract only the linearly dependent features and they don't focused much on the non-linear dependent features, by which excessive number of features are trained and tested. Lastly, the conventional method, MIFS + SVM, accomplished a two stage feature selection mechanism. But this method just considered MI as a reference criterion for the feature subset selection. To achieve an optimal feature set, the obtained MI needs to be further normalized (According to the FEFS) and needs to consider the Joint MI also.

Furthermore, the individual detection rate is also observed as high for proposed approach when it was compared with conventional approaches, as depicted in figure 1-b. The detection rate or recall is measured as the total number of correctly classified signals to the total test set of that particular class. For instance, in the table.5, the total number of signals classified as normal under normal class is shown as 3 and the total signals given for testing is also 3. Then the detection rate for the class 1 is evaluated as 100%. In this manner, the DR for remaining classes is also measured and the obtained Average DR is depicted in figure 1-b. Next the Average F-score is represented in figure 1-c and it is also observed as high compared to the conventional approaches.

The obtained average false alarm rate is depicted in figure 1-d. As can be seen, the Average FAR of proposed approach is less under both test cases when it was compared with the conventional approaches. According to the formulae of FAR (Eq.13), it is measured as the ratio of FP and the sum of TP and FP. Here the false positives define the number of signals which are classified falsely for a given positive input. From the confusion matrices, (table V and table VI), the obtained FPs for every class is observed as very less, by which the overall FAR also be very less for both test cases. Moreover, the FAR of proposed is very less when compared the FARs of conventional approaches. The main reason for being this much less FAR of proposed approach is the accomplishment of Joint MI between the selected feature set and the class at every instance of PQD signals by which the number of irrelevant features are reduced greatly.

Computational time is one more factor on which current researchers are mostly focusing. The CT is directly proportional to the size of dataset. As the dataset size increases, the CT also increases ad it reduces when the dataset size reduces. In the conventional approaches, most of the research focused to achieve an increased accuracy but not focused much on the time complexity. In the proposed approach and also the conventional approaches (only considered for comparison), the focus was made over the CT also by proposing different feature selection techniques. But they have their own drawbacks in the extraction of most optimal feature set by which the all PQD signals can be differentiated in a more discriminate way. The proposed approach got succeeded in this evaluation and extracts only the most relevant features those are related to all possible classes; hence the CT of proposed approach is less. The obtained results of CT are depicted in figure 1-e.

## VI. CONCLUSION AND FUTURE SCOPE

We presented a new system on the detection and classification of Power Quality Disturbances. With the aim of reducing the computational time complexity followed by an increased classification performance, two new feature selection approaches were developed based on the linear and non-linear dependencies between the samples constituted to formulate the PQD signals. Since the features are more important in the provision of discrimination between different signals, the proposed feature selection techniques extracts only the most significant features through which maximum information can be explored. Further the extracted optimal feature set is accomplished through machine learning algorithm to classify the test PQD signals. Here the Mutual Information and correlation properties are used for feature selection and SVM algorithm is used for classification. Various types of Simulation experiments carried out over the proposed system illustrated the effective performance in the detection and classification under normal and noisy environments. The obtained classification accuracy, detection rate and the computational time were observed as better for the proposed approach when compared with conventional approaches. On an average the Accuracy of proposed approach is enhanced by 1.5% and the computational time reduced by 100ms from the conventional approaches. Hence the proposed framework is robust for different PQD signals used to occur in different power related applications.

In future, this work can be extended by incorporating more than one classifier algorithms to achieve an increased accuracy. Since a single classifier algorithm cannot provide much differentiation for various unknown disturbance signals, two classifier algorithms would yield better performance. One classifier can be used to classify the normal signals from PQD signals and second stage classifier can be employed for more detailed classification.

## REFERENCES


[1] Kapoor, R., &Saini, M. K. (2012). Classification of power quality events – a review. International Journal of Electrical Power & Energy Systems, 43, 11–19.

[2] Kapoor, R., &Saini, M. K. (2007). A new signal processing technique for power system disturbances detection and classification. Int. J. Electrical Engineering, 88, 9–14.

[3] H. Siahkali, "Power quality indexes for continue and discrete disturbances in a distribution area," in Proc. IEEE Power and Energy Conf., Johor Bahru, Malaysia, pp. 678–683, 2008.

[4] Zhao, F.Z.; Yang, R.A. Power-quality disturbance recognition using S-transform. IEEE Trans.Power Deliv.2007, 22, 944–950.

[5] Li, L.; Yi, J.L.; Zhu, J.L. Parameter estimation of power quality disturbances using modified incomplete S-transform. Trans. China Electro Tech. Soc. 2011, 26, 187–193.

[6] Poisson, O.; Rioual, P.; Meunier, M. Detection and measurement of power quality disturbances using wavelet transform. IEEE Trans. Power Del. 2000, 15, 1039–1044.

[7] SaurabhKamble, IshitaDupare, "Detection of power quality disturbances using wavelet transform and artificial neural network", Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD), Kottayam India, 2014.

[8] P. Ray, N. Kishor, and S. Mohanty, "Islanding and power quality disturbancedetection in grid connected hybrid power system using waveletand S-transform," IEEE Trans. Smart Grid, vol. 3, no. 3, pp. 1082–1094,Sep. 2012.

[9] P. Ray, S. Mohanty, and N. Kishor, "Disturbance detection in grid connecteddistributed generation system using wavelet and S-transform," Elect.Power Syst. Res., vol. 81, no. 3, pp. 805–819, Mar. 2011.

[10] S. Santoso, W. M. Grady, E. J. Powers, J. Lamoree, and S. C. Bhatt,"Characterization of distribution power quality events with Fourier andwavelet transforms," IEEE Trans. Power Del., vol. 15, no. 1, pp. 247–254,Jan. 2000.

[11] Muhammad liaz, MdShafiullah, M A abido, "Classification of power quality disturbances using Wavelet Transform and Optimized ANN", 18th International Conference on Intelligent System Application to Power Systems (ISAP), Porto, Portugal, 2015.

[12] Liu, Z.G.; Han, Z.W.; Zhang, Y. Multiwavelet packet entropy and its application in transmissionline fault recognition and classification. IEEE Trans. Neural Netw. Learning Syst. 2014, 25,2043–2052.

[13] Dewal, K.Y.; Lal, M.; Shyam, A.R. Wavelet energy and wavelet entropy based epileptic brainsignals classification. Biomed. Eng. Lett.2012, 2, 147–157.

[14] Chen, J.; Li, G. Tsallis wavelet entropy and its application in power signal analysis. Entropy 2014,16, 3009–3025.

[15] Liu, Z.G.; Hu Q.L.; Cui, Y.; Zhang, Q. A new detection approach of transient disturbancescombining wavelet packet and Tsallis entropy.Neurocomputing2014, 142, 393–407.

[16] Battiti R. Using mutual information for selecting features in supervised neuralnet learning. IEEE Trans Neural Networks 1994;5(4):537–50.

[17] Z. Moravej, S. A. Banihashemi, and M. H. Velayati, "Power quality events classification and recognition using a novel support vector

algorithm", Energy Conversations and Management, Vol.50, No.3, pp. 3071–3077, 2009.
[18] Liu Tian-Yu, Li Guo-Zheng. Fault diagnosis by using selective ensemblelearning based on mutual information. In: The second internationalsymposium on optimization and systems biology (OSB'08) Lijiang, China,October 31–November 3 2008. p. 191–7.
[19] Danton D. Ferreira, Jose M. deseixas, Augusto S Cergueira, "A method based on independent component analysis for single and multiple power quality disturbance classification", Electric Power Systems Research, Volume 119, February 2015, Pages 425-431
[20] Lima, M.A.A.; Coury, D.V.; Cerqueira, A.S.; Nascimento, V.H. A method based on IndependentComponent Analysis for adaptive decomposition of multiple power quality disturbances. J.ControlAutom. Electr. Syst. 2014, 25, 80–92.
[21] Valtierra-Rodriguez, M.; de Jesus Romero-Troncoso, R.; Osornio-Rios, R.A.; Garcia-Perez, A.Detection and classification of single and combined power quality disturbances using neuralnetworks. IEEE Trans. Ind. Electron. 2014, 61, 2473–2482.
[22] Ray, P.K.; Mohanty, S.R.; Kishor, N.; Catalao, J.P.S. Optimal feature and decision tree-basedclassification of power quality disturbances in distributed generation systems. IEEE Trans.Sustain. Energy 2014, 5, 200–208.
[23] Manikandan, M.S.; Samantaray, S.R.; Kamwa, I. Detection and classification of power qualitydisturbances using sparse signal decomposition on hybrid dictionaries. IEEE Trans. Instrum.Meas. 2015, 64, 27–38.
[24] Hajian, M.; Foroud, A.A. A new hybrid pattern recognition scheme for automatic discriminationof power quality disturbances.Measurement2014, 51, 265–280.
[25] Zhigang Liu, Member, IEEE, Yan Cui, and Wenhui Li, "A Classification Method for Complex PowerQuality Disturbances Using EEMDand Rank Wavelet SVM", IEEE Transactions On Smart Grid, Vol. 6, No. 4, July 2015.
[26] Z. H. Wu and N. E. Huang, "A study of the characteristics of whitenoise using the empirical mode decomposition method," Proc. Roy. Soc.London A, vol. 460, no. 2046, pp. 1597–1611, 2003.
[27] Z. Li, Z. Weida, and J. Licheng, "Wavelet support vector machine," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no. 1, pp. 34–39,Feb. 2004.
[28] D. J. Strauss and G. Steidl, "Hybrid wavelet-support vector classification of waveforms," J. Comput. Appl. Math., vol. 148, no. 2, pp. 375–400,2002.
[29] Raj Kumar, Bhim Singh, D. T. Shahani, Ambrish Chandra, and Kamal Al-Haddad, "Recognition of Power Quality Disturbances Using S-Transform Based ANN Classifier and Rule Based Decision Tree", IEEE Transactions on Industry Applications, 2014.
[30] Subhra De1, Sudipta Debnath, "Real-time cross-correlation-based technique for detection and classification of power quality disturbances", IET Generation, Transmission & Distribution, Vol. 12 Iss. 3, 2017, pp. 688-695.
[31] Dasgupta, A., Debnath, S., Das, A.: 'Transmission line fault detection and classification using cross-correlation and K-nearest neighbor', Int. J. Knowl. Based Intell. Eng. Syst., 2015, 19, (10), pp. 183–189.
[32] Kwok, N., & Choi, C.(2002).Input features election for classification problems. IEEE Transactions on Neural Networks, 13, 143–159.
[33] Peng,H.,Long,F.,&Ding,C.(2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 1226–1238.
[34] Estévez, P. A., Tesmer, M., Perez, A., & Zurada, J. M. (2009). Normalized mutual information feature selection. IEEE Transactions on Neural Networks, 20,189–201.
[35] Hoque,N.,Bhattacharyya,D.K.,&Kalita,J.K.(2014).MIFS-ND:amutualinformation-basedfeatureselectionmethod.ExpertSystemswithApplications,41(14), 6371–6385.
[36] Yang, H., & Moody, J . (1999).Feature selection based on joint mutual information. In Proceedings of international ICSC symposium on advances in intelligent data analysis (pp.22–25).
[37] Aslam Shaik, Ankireddy Srinivasulu Reddy, "Flexible Entropy Based Feature Selection and Multi Class SVM for Detection and Classification of Power Quality Disturbances", International Journal of Intelligent Engineering and Systems, Vol.11, No.5, 2018.

.