Available online at www.sciencedirect.com
ScienceDirect
Procedia Technology 21 (2015) 163 – 170
Procedia Technology

SMART GRID Technologies, August 6-8, 2015

# Variance based offline Power Disturbance Signal Classification using Support Vector Machine and Random Kitchen Sink

Aneesh C[a*], Hisham PM[a], Sachin Kumar S[a], Maya P[b], K P Soman[a]

[a]*Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore – 641 112, India*
[b]*Dept of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Coimbatore – 641 112, India*

**Abstract**

In this paper, five classes of different power quality disturbances such as swell, sag, harmonics, sag with harmonics and swell with harmonics are synthesized using MATLAB/SIMULINK software which is further decomposed into 8 intrinsic mode functions using variational mode decomposition (VMD). VMD is an adaptive signal processing method that decomposes the signal in to several intrinsic mode functions (IMF) or components. The variance calculated from each of the mode is taken as feature representation. It is found that sines and cosines of variance vector of eight different IMF candidates of a signal acts as feature vector that can accurately extract salient and unique nature of the power disturbances. The classification is performed using Support Vector Machines (SVM) and Random Kitchen Sink (RKS) algorithm. The classification results in a highest accuracy of 94.44% for RKS method when compared to SVM.

Peer-review under responsibility of Amrita School of Engineering, Amrita Vishwa Vidyapeetham University
*Keywords:*Variational mode decomposition; Random Kitchen Sink; Power system, Support vector machine; variance

## 1. Introduction

Non-linear loads contain power electronic switching devices which distort the power system waveforms that were purely sinusoidal at the stage of generation. It draws current only for short duration of each mains cycle during which voltage drops at the point of common coupling with power system. Examples of such loads are controlled and

*Corresponding author. Tel.: +918754699468
*E-mail address:* a4aneeshc@gmail.com


Peer-review under responsibility of Amrita School of Engineering, Amrita Vishwa Vidyapeetham University
doi:10.1016/j.protcy.2015.10.084

uncontrolled converters, variable speed drives, consumer electronic equipment, computers and peripherals and uninterruptable power supplies. Various techniques have been employed to clearly identify and classify the power quality problems. In [1] G.Rata et al. uses both Fourier transform and wavelet transform for harmonic pollution analysis. Among them wavelet transform approach is widely used as it is able to provide the time information of harmonics which is missing in the Fourier transform and most of the PQ disturbances have non-stationary transient nature. In [2], Abdel Galil etal .utilizes wavelet transform for feature extraction. Here disturbance segments are extracted from voltage signals. Tracking and detecting algorithms such as Kalman filter (for sudden changes in disturbances) or Teager energy operator (for slow flickers like disturbances) are utilized. 11 level of decomposition is done for the test signal and energy of every obtained signal is found. Classification is done with the help of decision tree which uses C4.5 algorithm based upon Inductive inference algorithm and reported an overall accuracy of 90.4%. In [3], Haibo He et al. performs a Db4 wavelet feature extraction. 11 dimensional wavelet feature vector is created by a 10 level wavelet decomposition. Seven classes of different power quality disturbances with 200 signals per class was considered. The classification schema was done using a form of neural network called SOLAR (self-organizing learning array). This method provided an accuracy of 94.93%. In [4], Murat Uyar et al. explains a wavelet norm entropy based feature extraction method for PQ disturbance classification. The P-value for norm entropy calculation selected was 1.75. and obtained an accuracy of 95.71%. In [5], uses a different feature extraction method in which a standard deviation curve plotted for wavelet transformed coefficients at different resolution levels was applied for classification. In [6], DWT integrating with PNN (Probabilistic Neural Network) is used for PQ disturbance recognition.In this paper, a simple variance based classification is discussed. Six categories of power disturbance signal are synthesized for performing experiments. Each such signal is decomposed using VMD, into eight intrinsic modes or sub-signals. Through experiments, eight modes are found out which are sufficient to represent the signal. On observing each mode of a signal, the variance information about each mode can be found out which captures the disturbance information so that it can be used as feature for distinguishing itself with different disturbances.

## 2. Variational Mode Decompositon

The goal of Variational Mode decomposition (VMD) [7] is to decompose a signal into different modes (or sub-signals) called intrinsic mode functions (IMF's) having finite band widths using calculus of variation method. VMD tries to find out the modes concurrently using ADMM (Alternating Directional Method of Multipliers) optimization technique, hence it is called a non-recursive algorithm. In VMD method when all the decomposed modes are added together, it reconstructs the original signal back with certain sparsity properties. VMD gives a reasonable result when compared to Empirical Mode Decomposition (EMD) which was introduced by Huang et.al [8] and Empirical Wavelet Transform (EWT) [9]. EMD doesn't have a strong mathematical base and this algorithm fails in noise affected situations [10]. The original formulation of VMD casted as optimization problem is continuous in time domain. The formulation is given as [7],

$$\min_{u_k,\omega_k}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\}$$

$$\text{s.t.}\quad \sum_k u_k = f \tag{1}$$

The augmented lagrangian $L$ method converts the constrained formulation in equation 1, into an unconstrained one as

$$L(u_k,w_k,\lambda)=\alpha\sum_k\left\|\partial_t\left[\left((\delta(t)+(j/\pi t))*u_k(t)\right)e^{-j\omega_k t}\right]\right\|_2^2+\left\|f-\sum_k u_k\right\|_2^2+\left\langle\lambda,f-\sum_k u_k\right\rangle \tag{2}$$

In VMD, the solution is obtained in the frequency domain using ADMM method applied on to a sequence of iterations [7].

## 3. Random Kitchen Sink algorithm

Random kitchen sink algorithm is a new machine learning algorithm for classification of non-linearly separated data set. An explicit feature mapping $\phi(x)$ is obtained corresponding to Radial basis function (RBF) Kernel. This in combination with regularized least square algorithm for regression allows us to obtain a simple classifier that can be used for real time applications. The idea behind Random kitchen sink algorithm is that RBF kernel is a real Gaussian function and hence it is symmetric. The Fourier transform of a real Gaussian function is again a real. Gaussian function and can be interpreted as a multivariate Gaussian probability density function. This reinterpretation allows to get an explicit method for deriving feature function. The mathematical derivation is as follows. Define the dimension :

Let $K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle = e^{\frac{-1}{\sigma}\|x_1 - x_2\|_2^2}$ (3)

$$e^{\frac{-1}{\sigma}\|x_1-x_2\|_2^2} = e^{\frac{-1}{\sigma}(x_1-x_2)^T(x_1-x_2)} = e^{\frac{-1}{2}(x_1-x_2)^T\Sigma^{-1}(x_1-x_2)} \quad \text{where } \Sigma = \begin{bmatrix} 2\sigma & 0 & \dots & 0 \\ 0 & 2\sigma & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 2\sigma \end{bmatrix} \tag{4}$$

The kernel can be expressed as a Gaussian probability density function. To be a proper density function it must be scaled. Since covariance matrix is diagonal , the Gaussian density function can be written as product of n Gaussian function and hence the associated variables are independent when kernel is interpreted as probability density function. Let,

$x_1 - x_2 = z$ , then the kernel function can be written as $f(z) = e^{\frac{-1}{2}z^T\Sigma^{-1}z}$ . Let $F(\Omega)$ represent Fourier Transform of $f(z)$ . That is

$$F(\Omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(z)e^{-jz^T\Omega}dz \tag{5}$$

Since $f(z)$ is Gaussian $F(\Omega)$ is again Gaussian. To be precise it is a multivariate Gaussian function with variance ( to be verified regarding sigma ) . We interpret $F(\Omega)$ as a Gaussian (multivariate) density function. Now

$F^{-1}(\Omega) = \langle \phi(x_1), \phi(x_2) \rangle = \int_{-\infty}^{\infty} F(\Omega)e^{jz^T\Omega}dz$ . This can be interpreted as expected value of the quantity $e^{jz^T\Omega}$ . That is,

$$E(e^{jz^T\Omega}) = \int_{-\infty}^{\infty} F(\Omega_i)e^{jz^T\Omega}dz \tag{6}$$

However our aim is to obtain a generic expression for $\phi(x)$ .

$$E(e^{jz^T\Omega}) = \frac{1}{k}\sum_{i=1}^{k} e^{jz^T\Omega_i} = \frac{1}{k}\sum_{i=1}^{k} e^{j(x-y)^T\Omega_i} = \frac{1}{k}\sum_{i=1}^{k} e^{jx^T\Omega_i}\overline{e^{jy^T\Omega_i}} = \langle \phi(x_1), \phi(x_2) \rangle = K(x_1, x_2) \tag{7}$$

$$K(x_1, x_2) = \frac{1}{k}\begin{bmatrix} e^{j(x_1-x_2)^T\Omega_1} \\ e^{j(x_1-x_2)^T\Omega_2} \\ . \\ e^{j(x_1-x_2)^T\Omega_k} \end{bmatrix} = \frac{1}{k}\left\langle \begin{bmatrix} e^{jx_1^T\Omega_1} \\ e^{jx_1^T\Omega_2} \\ . \\ e^{jx_1^T\Omega_k} \end{bmatrix}, \begin{bmatrix} e^{jx_2^T\Omega_1} \\ e^{jx_2^T\Omega_2} \\ . \\ e^{jx_2^T\Omega_k} \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} \sqrt{1/k}\,e^{jx_1^T\Omega_1} \\ \sqrt{1/k}\,e^{jx_1^T\Omega_2} \\ . \\ \sqrt{1/k}\,e^{jx_1^T\Omega_k} \end{bmatrix}, \begin{bmatrix} \sqrt{1/k}\,e^{jx_2^T\Omega_1} \\ \sqrt{1/k}\,e^{jx_2^T\Omega_2} \\ . \\ \sqrt{1/k}\,e^{jx_2^T\Omega_k} \end{bmatrix} \right\rangle \tag{8}$$

To avoid complex number computation, it can be taken as

$$\phi(x)=\sqrt{1/k}\begin{bmatrix} Cos(x^T\Omega_1) \\ \vdots \\ Cos(x^T\Omega_k) \\ Sin(x^T\Omega_1) \\ \vdots \\ Sin(x^T\Omega_k) \end{bmatrix}$$

*3.1 Regularized least square algorithm*

The regularized least square formulation for multiclass classification problem is formulated as a regression problem. The regression parameters is obtained using the following optimization problem. In this formulation $X$ denotes the data matrix, $Y$ denotes the label matrix. Using lagrangian method and properties of trace, weight matrix $W$ is obtained.

$$W^* = \arg\min_{W^{m\times n}} \|Y - XW\|_F^2 + \lambda\|W\|_F^2 \tag{9}$$

$$f(W) = \|Y - XW\|_F^2 + \lambda\|W\|_F^2 = Tr\left(\left(Y - XW\right)^T\left(Y - XW\right) + \lambda W^T W\right) \tag{10}$$

$$\frac{\partial f(W)}{\partial W} = 2\left(X^T X\right)W - X^T Y - X^T Y + 2\lambda W = 0 \Rightarrow W^* = \left(X^T X + \lambda I\right)^{-1} X^T Y \tag{11}$$

## 4. Support Vector Machine

Support Vector Machine (SVM) belongs to the class of supervised learning algorithms in which the machine is given a set of examples, data vectors $\{x_1 \ldots x_n\}$ where $x_i \in R^m$, associated with class labels, $\{y_1 \ldots y_n\}$ where $y_i \in \{-1,1\}$,to learn [11].In 3-D scenario, a plane provides a linear separation. And in n-D space a hyperplane provides the separation.

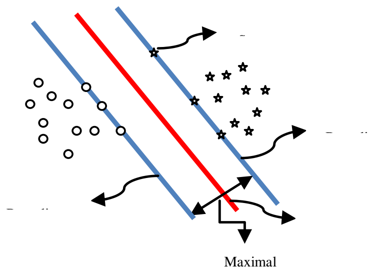


Fig. 1: Linear separation of data points using SVM.

In Fig. 1, Blue denotes the bounding planes which are parallel to the classifier, the red colored one. The distance between the bounding planes is known as 'margin'. SVM tries to find a hyperplane with maximal margin. From the Figure 1, it can be seen that few data points fall on the bounding planes. These data points are called support vectors. SVM can be understood in the simple 2 class scenario. Let $x_1$ and $x_2$ be the two variables. The aim is to find a hyperplane of the form $w_1x_1 + w_2x_2 - \gamma = 0$ and the two bounding planes are given as $w_1x_1 + w_2x_2 - \gamma \geq 1$ and $w_1x_1 + w_2x_2 - \gamma \leq -1$. Here $\gamma$ is a scalar term, known as 'bias' term. During the training stage, SVM finds the appropriate $w_{i's}$ and $\gamma$. Once these parameters are found then the decision boundary is obtained as $w^T x - \gamma = 0$

where $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$. The decision can be made using the function $f(x) = sign(w^T x - \gamma)$, so when a new point come, the sign value decides the class.

## 5. Proposed methodology

Five classes of signals with different power quality problems are considered such as swell, sag, harmonics, sag with harmonics and swell with harmonics. Also a class with pure sinusoidal signal with no disturbances (normal) is included. All the six classes are denoted by C1, C2, C3, C4, C5, and C6 respectively. These disturbances waveforms are synthesized using MATLAB/SIMULINK software. Fig. 2 and Fig. 3 show the block diagram of the system for generating signals. A three phase 380V, 50Hz source is connected to a linear load through an 80km transmission line which is divided into two sections. In between these sections a nonlinear load is connected and also a single line to ground fault is generated. When a single line to ground fault occurs, voltage in that particular line sags and voltage in other two lines swell. The presence of non linear load gives rise to harmonics in the bus voltage. By creating a fault at different positions with non linear load disconnected, bus voltage waveforms with sag and swell can be synthesized. In case if the non linear load is connected, bus voltages will have sag/swell with harmonics also. Pure sinusoidal signals are synthesized with no fault condition with non linear load disconnected.

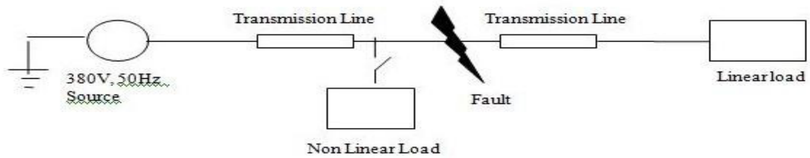


Fig. 2: Block diagram for generating swell with harmonics and sag with harmonics

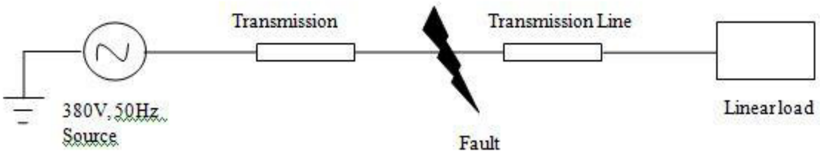


Fig. 3: Block diagram for generating swell and sag signals

Each class consists of 31 cases. Fig. 4 and Fig. 5 show one such generated signals from all the different classes considered. Each PQ signal is decomposed into eight IMF candidates using VMD. Decomposing in to higher number of modes cases to increase the dimension of feature vector which further increases the computational cost. So here we choose eight modes. Modes extracted by VMD for an harmonic signal is shown in Fig. 6. Sine and cosine of variance of extracted modes are used as feature vector. The extracted feature vector contains main useful details of given waveform [14]. For a given PQ signal the dimension of feature vector is $1\times16$. The total size of training data is $168\times17$. Where 168 comes from 28 cases per class multiplied by 6 classes and 17 is the feature size for each signal including label vector. Testing data dimension is $18\times17$. Three signals from each class is used for testing purpose. So 18 comes from three cases per class multiplied by 6 classes.

## 6. Results and discussion

The classification results obtained using SVM and RKS are shown in Table 1 and Table 2. The table represents confusion matrix for different types of kernel functions. The diagonal elements represent the number of correctly classified PQ disturbances and off-diagonal elements represent the misclassified PQ types. An optimum selection of kernel function and parameters gives good classification results [15]-[17]. From the result obtained we can observe that RKS based classification provides high accuracy. Form the confusion matrix it can be observed that RKS algorithm, when compared to SVM gives the highest accuracy of 94.44% for radial basis function (RBF) kernel. In

SVM the RBF kernel gave an accuracy of 83.33%. At the same time, linear kernel for both methods gave an accuracy of 88.89%.

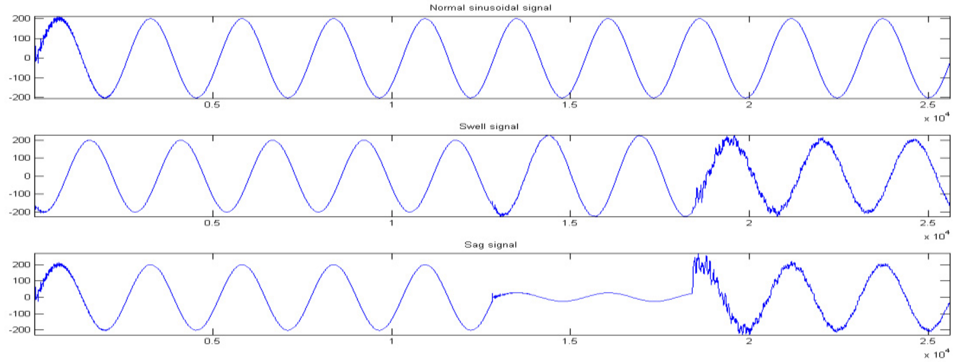


Fig. 4: PQ disturbances waveforms

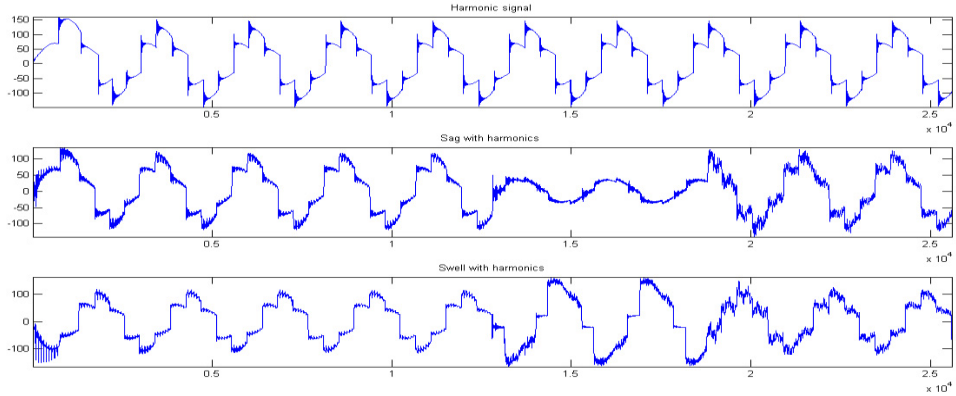


Fig. 5: PQ disturbances waveforms

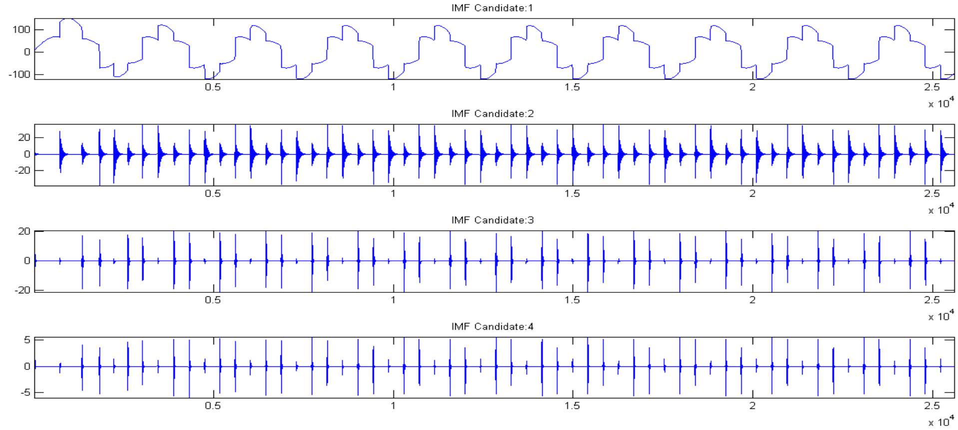


Fig. 6: Four IMF components extracted from harmonic signal

**Table 1:** Classification results using SVM

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | **SVM** | | | |
| | Class | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
| | **C1** | 3 | 0 | 0 | 0 | 0 | 0 |
| **Linear** | **C2** | 0 | 3 | 0 | 0 | 0 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Kernel Function** | **C3** | 0 | 0 | 3 | 0 | 0 | 0 |
| | **C4** | 0 | 0 | 0 | 2 | 1 | 0 |
| | **C5** | 1 | 0 | 0 | 0 | 2 | 0 |
| | **C6** | 0 | 0 | 0 | 0 | 0 | 3 |
| | **Overall Accuracy** | **88.89%** | | | | | |
| | Class | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
| **Polynomial Kernel Function** | **C1** | 3 | 0 | 0 | 0 | 0 | 0 |
| | **C2** | 0 | 3 | 0 | 0 | 0 | 0 |
| | **C3** | 0 | 1 | 1 | 1 | 0 | 0 |
| | **C4** | 0 | 0 | 0 | 3 | 0 | 0 |
| | **C5** | 0 | 0 | 1 | 1 | 1 | 0 |
| | **C6** | 0 | 0 | 0 | 0 | 0 | 3 |
| | **Overall Accuracy** | **77.78%** | | | | | |
| | Class | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
| **RBF Kernel Function** | **C1** | 3 | 0 | 0 | 0 | 0 | 0 |
| | **C2** | 0 | 3 | 0 | 0 | 0 | 0 |
| | **C3** | 0 | 0 | 3 | 0 | 0 | 0 |
| | **C4** | 0 | 0 | 0 | 2 | 1 | 0 |
| | **C5** | 1 | 0 | 1 | 0 | 1 | 0 |
| | **C6** | 0 | 0 | 0 | 0 | 0 | 3 |
| | **Overall Accuracy** | **83.33%** | | | | | |
| | Class | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
| **Sigmoid Kernel Function** | **C1** | 3 | 0 | 0 | 0 | 0 | 0 |
| | **C2** | 0 | 3 | 0 | 0 | 0 | 0 |
| | **C3** | 0 | 0 | 3 | 0 | 0 | 0 |
| | **C4** | 1 | 0 | 0 | 1 | 1 | 0 |
| | **C5** | 1 | 0 | 0 | 1 | 1 | 0 |
| | **C6** | 0 | 0 | 0 | 0 | 0 | 3 |
| | **Overall Accuracy** | **77.78%** | | | | | |

**Table 2:** Classification results using RKS

| | **RKS** | | | | | | |
|---|---|---|---|---|---|---|---|
| | Class | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
| **Linear Kernel With Hold Out cross validation** | **C1** | 3 | 0 | 0 | 0 | 0 | 0 |
| | **C2** | 0 | 3 | 0 | 0 | 0 | 0 |
| | **C3** | 0 | 0 | 2 | 0 | 1 | 0 |
| | **C4** | 0 | 0 | 0 | 2 | 1 | 0 |
| | **C5** | 0 | 0 | 0 | 0 | 3 | 0 |
| | **C6** | 0 | 0 | 0 | 0 | 0 | 3 |
| | **Overall Accuracy** | **88.89%** | | | | | |
| | Class | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
| **RBF Kernel with Leave One Out cross** | **C1** | 3 | 0 | 0 | 0 | 0 | 0 |
| | **C2** | 0 | 3 | 0 | 0 | 0 | 0 |
| | **C3** | 0 | 0 | 2 | 1 | 0 | 0 |
| | **C4** | 0 | 0 | 1 | 2 | 0 | 0 |
| | **C5** | 0 | 0 | 0 | 0 | 3 | 0 |
| | **C6** | 0 | 0 | 0 | 0 | 0 | 3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **validation** | **Overall Accuracy** | **88.89%** | | | | | |
| | Class | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
| **RBF Kernel with Hold Out cross validation** | **C1** | 3 | 0 | 0 | 0 | 0 | 0 |
| | **C2** | 0 | 3 | 0 | 0 | 0 | 0 |
| | **C3** | 0 | 0 | 3 | 0 | 0 | 0 |
| | **C4** | 0 | 0 | 1 | 2 | 0 | 0 |
| | **C5** | 1 | 0 | 0 | 0 | 3 | 0 |
| | **C6** | 0 | 0 | 0 | 0 | 0 | 3 |
| | **Overall Accuracy** | **94.44%** | | | | | |

## 7. Conclusion

This paper treats PQ classification approach based on VMD and RKS algorithm based on regularized least square approach. VMD is used to construct the feature vector. The main advantage of VMD comes from its ability to accurately separate harmonics present in the signal independent of how close their frequency components [7]. The variance of decomposed modes by VMD is used as feature vector.From the comparison between RKS algorithm and SVM for classification, concluding that RKS algorithm using RBF kernel with hold-out cross validation outperforms over SVM. High classification accuracy of the given method is the result of the use of RKS algorithm for classification. It is the purpose of this present work to introduce recent adaptive signal processing method called VMD and an excellent classification algorithm called RKS for PQ disturbances classification. This work can be extended for the classification of real time detection and identification of PQ disturbances generated in distribution systems together with expanding the disturbance classes such as notch, flicker, outage, impulses, glitches, spikes, etc.

## References

[1] G. Rata, M. Rata, C. Filote, Theoretical and experimental aspects concerning Fourier and wavelet analysis for deforming consumers in power network, ElektronikairElektrotechnika (Electronics andElectrical Engineering), no. 1, p. 62–66, 2010

[2] T. K. A. Galil, M. Kamel, A. M. Youssef, E. F. E. Saadany, M. M. A.Salama, Power quality disturbance classification using the inductive inference approach, IEEE Trans. Power Delivery, vol. 19, no. 10, p. 1812–1818, 2004

[3] H. He, J. A. Starzyk, A self-organizing learning array system for power quality classification based on wavelet transform, IEEETrans. Power Delivery, vol. 21, no. 1, p. 286–295, 2006

[4] M. Uyara, S. Yildirima, M. T. Gencoglub, An effective wavelet based feature extraction method for classification of power quality disturbance signals", Journal Electric Power System Research, vol.78, p. 1747–1755, 2008

[5] A.M. Gaouda, M.M.A. Salama, M.R. Sultan, A.Y. Chikhani, Power quality detection and classification using wavelet-multiresolution signal decomposition,IEEE Trans. Power Deliv. 14 (4) ,1999,p. 1469–1476

[6] Z.L. Gaing, Wavelet-based neural network for power disturbance recognitionand classification, IEEE Trans. Power Deliv. 19(4) ,2004, p.1560–1568

[7] Konstantin Dragomiretskiy and Dominique Zosso, Variational Mode Decomposition, IEEE transactions on signal processing.

[8] N.E. Huang, Z. shen, S.R. Long, M.L. Wu, H.H Shih, Q. Zheng, N.C. Yen, C.C. Tung and H.H. Liu, The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. Roy. Soc. London A, Vol 454, 1998, p. 903-995.

[9] Jerome Gilles. Empirical wavelet transform. IEEE transactions on signal processing, Vol. 61, No, 16, 2013

[10]T. Kijewski-Correa and A. Kareem,Performance of Wavelet Transform and Empirical Mode Decomposition in Extracting signals Embedded in Noise, Journal of Engineering Mechanics, Vol. 133, No. 7, July 1, 2007.

[11] Dr. K. P. Soman, Ajav. V, Loganathan R, Machine Learning with SVM and other Kernel Methods, PHI Learning Pvt. Ltd, 2009.

[12] Thorsten Joachims. 1999.,ransductive Inference for Text Classification using Support Vector Machines, Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99)

[13] A. Rahimi and B. Recht. Random features for large-scale kernel machines - Advances in Neural Information Processing Systems (NIPS), 2007.

[14] A. Sengur, I. Turkoglu, M.C. Ince, Wavelet packet neural networks for texture classification, Expert Syst. Appl. ,2007,,p. 527–533

[15] B. Scholkopf, A. Smola, R. C.Williamson, and P. L. Bartlett, New support vector algorithms, Neural Computation, vol. 12, 2000, p. 12071245.

[16] V. Vapnik, Statistical Learning Theory. New York: Wiley, 1998

[17] C. Cortes and V. Vapnik, Support-vector network, Mach. Learning,vol. 20, 1995, p. 273–297.