

Feature Mapping Techniques for Improving the Performance of Fault Diagnosis of Synchronous Generator

R. Gopinath¹, C. Santhosh Kumar², K. Vishnuprasad³, and K. I. Ramachandran⁴

^{1,2,3,4} *Machine Intelligence Research Laboratory,
Department of Electronics and Communication Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham Coimbatore, Tamil Nadu, India*

r_gopinath@cb.amrita.edu

cs_kumar@cb.amrita.edu

mail2vpk@gmail.com

ki_ram@cb.amrita.edu

ABSTRACT

Support vector machine (SVM) is a popular machine learning algorithm used extensively in machine fault diagnosis. In this paper, linear, radial basis function (RBF), polynomial, and sigmoid kernels are experimented to diagnose inter-turn faults in a 3kVA synchronous generator. From the preliminary results, it is observed that the performance of the baseline system is not satisfactory since the statistical features are non-linear and does not match to the kernels used. In this work, the features are linearized to a higher dimensional space to improve the performance of fault diagnosis system for a synchronous generator using feature mapping techniques, sparse coding and locality constrained linear coding (LLC). Experiments and results show that LLC is superior to sparse coding for improving the performance of fault diagnosis of a synchronous generator. For the balanced data set, LLC improves the overall fault identification accuracy of the baseline RBF system by 22.56%, 18.43% and 17.05% for the R, Y and B-phase faults respectively.

1. INTRODUCTION

Condition based maintenance (CBM) is the most preferred technique in many industrial applications for its reduced maintenance costs and improved safety operations. CBM reduces the downtime and increases the productivity (Jardine, Lin, & Banjevic, 2006). Data acquisition is the primary step in CBM wherein mechanical and electrical signals are collected from the machines to monitor its health. Feature extraction is an important process in CBM which maps the measured signal into the feature space. The performance of the fault diagnosis algorithm is also dependent on the features (Saxena, Wu, &

Vachtsevanos, 2005; Wu et al., 2004; Wu, Saxena, Patrick, & Vachtsevanos, 2005). Signal processing based feature extraction methods such as time-domain (Samanta & Al-Balushi, 2003), frequency-domain (Chen, Du, & Qu, 1995), wavelet (Peter, Peng, & Yam, 2001; Lin & Zuo, 2004; Yan, Gao, & Wang, 2009), and empirical mode decomposition (Yan & Gao, 2008; He, Liu, & Kong, 2011) have been widely used in machine condition monitoring applications. Many feature selection algorithms have been developed for effective fault diagnosis (Chiang, Kotanchek, & Kordon, 2004; Casimir, Boutleux, Clerc, & Yahoui, 2006; Verron, Tiplica, & Kobi, 2008; Y. Yang, Liao, Meng, & Lee, 2011; K. Zhang, Li, Scarf, & Ball, 2011) which are used to select the fault discriminative features from the feature space for better classification. Feature transformation approaches are also used to improve the fault identification performance (Widodo, Yang, & Han, 2007; Widodo & Yang, 2007; Y. Zhang, 2009).

Choosing an appropriate classification algorithm for a particular application is a difficult task. It also depends on the characteristics of extracted features from the raw data. SVM is an important supervised machine learning algorithm widely used in various applications including machine fault diagnosis (Nayak, Naik, & Behera, 2015). The performance of the SVM classifier could be affected by the kernel functions, training sample size, and kernel parameters. Zhou et al. investigated the effects of the training sample size, SVM order, and kernel parameters using least squares SVM for the linear, polynomial and Gaussian kernels (J. Zhou, Shi, & Li, 2011). Wang et al. reviewed SVM for uncertain data. Robust optimization is used when the direct model could not guarantee a good performance on uncertainty data set (X. Wang & Pardalos, 2015). Khang et al. proposed genetic algorithm based kernel discriminative features to improve the performance of multi-class SVM for low speed bearing fault diagnosis (Kang et al., 2015). Fu et al. made a comparative study on grid

Gopinath R., et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

search (GS), genetic algorithm (GA), and particle swarm optimization (PSO) to optimize the penalty factor λ and γ in a kernel function in a training process (Fu, Tian, & Wu, 2015).

In this work, first the baseline system is developed using SVM kernels (linear, radial basis function (RBF), polynomial, and sigmoid) to identify inter-turn faults in a 3kVA synchronous generator. It is observed that the performance of the classifier is not satisfactory as the statistical features are non-linear and does not match with the kernels used. Performance of the SVM classifier can be improved by the following approaches: a) Select the features matching a particular kernel. b) Choose an appropriate kernel that fits the features. c) Linearize the features to a higher dimensional space and match with linear kernel. In this paper, the third approach is experimented to improve the performance of fault diagnosis of a synchronous generator using feature mapping techniques.

Sparse coding is an unsupervised machine learning algorithm used to represent feature vectors as a linear combination of basis vectors. Liu et al. used adaptive sparse features and classified the faults using multi-class linear discriminant analysis for machine fault diagnosis (Liu, Liu, & Huang, 2011). Liu et al. also pointed out that sparse coding requires high computational cost for dictionary learning for which faster algorithms are to be developed (Liu et al., 2011). Zhu et al. proposed an automatic and adaptive feature extraction technique via K-SVD. Faults are diagnosed using the reconstruction error of the sparse representation (Zhu et al., 2014). Further, fusion sparse coding technique was proposed to extract impulse components from the vibration signals effectively (Deng, Jing, & Zhou, 2014). For a good classification performance, coding algorithm should generate similar codes for similar feature vectors. However, sparse coding might select different bases for similar feature vectors to support sparsity, thus fails to capture the correlations between codes (J. Wang et al., 2010).

Local coordinate coding (LCC) overcomes the drawbacks of sparse coding by explicitly encouraging the bases to be local and consequently requires the codes to be sparse (Yu, Zhang, & Gong, 2009). However, sparse coding and LCC algorithms have the computational complexity of $O(M^2)$, where M represents the total number of vectors in the basis set. Wang et al. proposed locality constrained linear coding (LLC) method for the image classification applications (J. Wang et al., 2010) to represent the non-linear features that improve the performance using linear classifiers (J. Yang, Yu, Gong, & Huang, 2009; X. Zhou, Cui, Li, Liang, & Huang, 2009). LLC is a fast implementation of LCC which uses locality constraint to select k nearest bases for each feature vector, therefore it reduces the computational complexity from $O(M^2)$ to $O(M + k^2)$. Recent studies show that LLC have been applied to many image processing applications such as video summarization (Lu et al., 2014), human action recognition (B. Wang et al.,

2014; Rahmani, Mahmood, Huynh, & Mian, 2014), magnetic resonance (MR) imaging (P. Zhang et al., 2013), and colorization for gray scale facial images for improved performance (Liang et al., 2014). In this paper, the performance of the sparse coding and locality constrained linear coding (LLC) are compared for improving the performance of fault diagnosis of a synchronous generator. The details of the experimental setup, data collection, feature extraction, sparse coding, locality constrained linear coding (LLC) and support vector machine (SVM) are discussed in section 2. Experiments and results are discussed in section 3 and finally, section 4 concludes this paper.

2. SYSTEM DESCRIPTION

2.1. Experimental Setup and Data Collection

In synchronous generators, short circuit faults may happen in stator winding and field winding coils. Generally, stator and field winding terminal of synchronous generator has taps at 0% and 100% of the coil windings. The three phase 3kVA synchronous generator is customized to inject faults in different magnitude. For example, 30%, 60%, and 82 % of the total number of turns in the stator winding leads are made available to the front panel for injecting short circuit faults between any of these points.

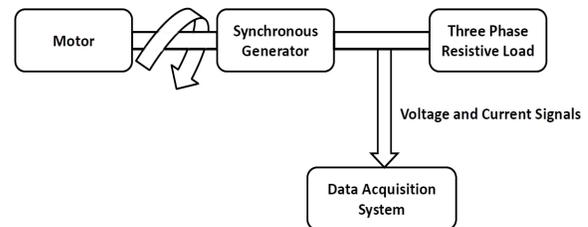


Figure 1. Block diagram of the experimental setup

Each phase has 18 taps, making the total number of taps across three phases to 54, to inject different short circuit faults in the generator. All terminal leads from the stator winding are taken to the front panel board. Block diagram of the experimental setup is shown in Fig. 1. Photograph of the customized synchronous generator and experimental facility is shown in Figs. 2 and 3. Design details of the customized synchronous generator can be found in (Gopinath et al., 2013).

In this work, inter turn short circuit faults are injected in a controlled manner. Generator is connected to the three phase resistive load. Data acquisition system NI-PXI 6221¹ is used to interface the current sensors. Each experiment is conducted for 10 seconds and current signals are sampled at 1 kHz. This makes the number of samples available for each trial to be 10,000. The process is then repeated for different

¹<http://sine.ni.com/ds/app/doc/p/id/ds-15/lang/en>

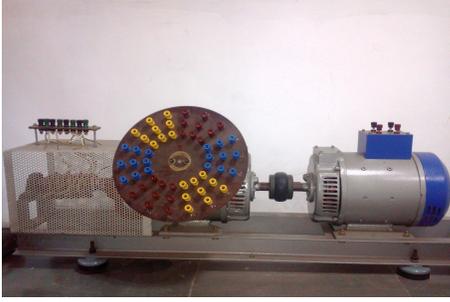


Figure 2. Customized 3 kVA Synchronous Generator



Figure 3. Experimental Facility

fault conditions to make the data collection process complete. Fig. 4, illustrates the current signatures acquired from the 3kVA generator. Specifications of the synchronous generator is listed in Appendix A. Short circuit faults are injected at 30% (8 turns), 60% (16 turns), and 82% (22 turns) of the total number of turns (27 turns). The data from each trial is divided into multiple frames with a window size of 512 samples. The time domain signal is converted into frequency domain using Fast Fourier Transform (FFT). Statistical frequency domain features are used to extract the fault information from the raw data. The details of the frequency domain features (Lei, He, & Zi, 2008) are listed in Table 1.

2.2. Sparse Coding

Sparse coding is an unsupervised learning method, learns the set of over-complete bases to represent the data efficiently (Olshausen & Field, 1997). Sparse coding finds the basis vectors such that, the input vectors $X = [x_1, x_2, \dots, x_N]^T \in \mathbb{R}^{D \times N}$, with D dimension can be expressed as a linear combination of these bases. The input vector can be expressed as (Olshausen & Field, 1997):

$$x_i = \sum_{i=1}^N b_i c_i \quad (1)$$

where $B = [b_1, b_2, \dots, b_M] \in \mathbb{R}^{D \times M}$ is basis vectors or codebooks. Over-complete basis identifies the patterns in the

input data. However, coefficients or codes c_i are not uniquely determined by the input vectors, with an over-complete basis. This necessitates to add the sparsity criterion for better representation (J. Yang et al., 2009), i.e., most of the coefficients c_i are zero or nearly zero and only a few coefficients are non-zero to represent the input data efficiently. Sparse coding can be expressed as (J. Yang et al., 2009):

$$\min_c \sum_{i=1}^N \|x_i - Bc_i\|^2 + \lambda \|c_i\|_{l^1} \quad (2)$$

where $\lambda \|c_i\|_{l^1}$ is the sparse regularization term and it is determined as l^1 norm of c_i . Sparse regularization ensures that the codebook is over-complete and unique solution for the under-determined system, hence it captures the patterns in the input data.

2.3. Locality Constrained Linear Coding (LLC)

Locality constrained linear coding (LLC) is a feature mapping technique used to represent non-linear features as linear features (J. Wang et al., 2010). LLC codes are sparse and high dimensional. In LLC, each input vector is represented as a linear combination of k -nearest basis vectors. The basis vectors are computed from the data set using k -means clustering algorithm. Basis vectors are called as codebooks in the context of coding schemes algorithm (J. Wang et al., 2010). The LLC coding process is described in Fig. 5. Let X be a set of D - dimensional input database $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{D \times N}$. Given a codebook, $B = [b_1, b_2, \dots, b_M] \in \mathbb{R}^{D \times M}$, coding algorithm converts each input data into a M - dimensional code, where $[b_1, b_2, \dots, b_M]$ are basis vectors and these basis vectors captures the patterns in the input data. Criteria for the LLC code is expressed as (J. Wang et al., 2010):

$$\min_c \sum_{i=1}^N \|x_i - Bc_i\|^2 + \lambda \|d_i \odot c_i\|^2 \quad (3)$$

$$s.t \ 1^T c_i = 1, \forall i$$

where \odot represents element-wise multiplication. Locality adaptor $d_i \in \mathbb{R}^M$ provides freedom for each basis vector proportional to its similarity to the input feature x_i . Locality adaptor can be expressed as (J. Wang et al., 2010):

$$d_i = \exp\left(\frac{\text{dist}(x_i, B)}{\sigma}\right) \quad (4)$$

where $\text{dist}(x_i, B)$ is the Euclidean distance between x_i and b_j and it can be written as (J. Wang et al., 2010):

Table 1. Frequency Domain Features

Description	Feature	Description	Feature
Mean	$Feat_1 = \frac{\sum_{k=1}^K s(k)}{K}$	Root Mean Square (RMS) frequency	$Feat_7 = \sqrt{\frac{\sum_{k=1}^K f_k^2 s(k)}{\sum_{k=1}^K s(k)}}$
Variance	$Feat_2 = \frac{\sum_{k=1}^K (s(k) - Feat_1)^2}{K-1}$	Spectrum power convergence	$Feat_8 = \sqrt{\frac{\sum_{k=1}^K f_k^4 s(k)}{\sum_{k=1}^K f_k^2 s(k)}}$
Skewness	$Feat_3 = \frac{\sum_{k=1}^K (s(k) - Feat_1)^3}{K(\sqrt{Feat_2})^3}$	Stability factor	$Feat_9 = \frac{\sum_{k=1}^K f_k^2 s(k)}{\sqrt{\sum_{k=1}^K s(k) \sum_{k=1}^K f_k^4 s(k)}}$
Kurtosis	$Feat_4 = \frac{\sum_{k=1}^K (s(k) - Feat_1)^4}{K(Feat_2)^2}$	Coefficient variability	$Feat_{10} = \frac{Feat_6}{Feat_5}$
Frequency centre	$Feat_5 = \frac{\sum_{k=1}^K f_k s(k)}{\sum_{k=1}^K s(k)}$	Skewness frequency	$Feat_{11} = \frac{\sum_{k=1}^K (f_k - Feat_5)^3 s(k)}{K(Feat_6)^3}$
Standard deviation frequency	$Feat_6 = \sqrt{\frac{\sum_{k=1}^K (f_k - Feat_5)^2 s(k)}{K}}$	Kurtosis frequency	$Feat_{12} = \frac{\sum_{k=1}^K (f_k - Feat_5)^4 s(k)}{K(Feat_6)^4}$
Spectrum power positional factor	$Feat_{13} = \frac{\sum_{k=1}^K (f_k - Feat_5)^{\frac{1}{2}} s(k)}{K\sqrt{Feat_6}}$		

where $s(k)$ is the spectrum for $k = 1, 2, \dots, K$, K is the number of spectrum lines f_k is the frequency value of the k^{th} spectrum.

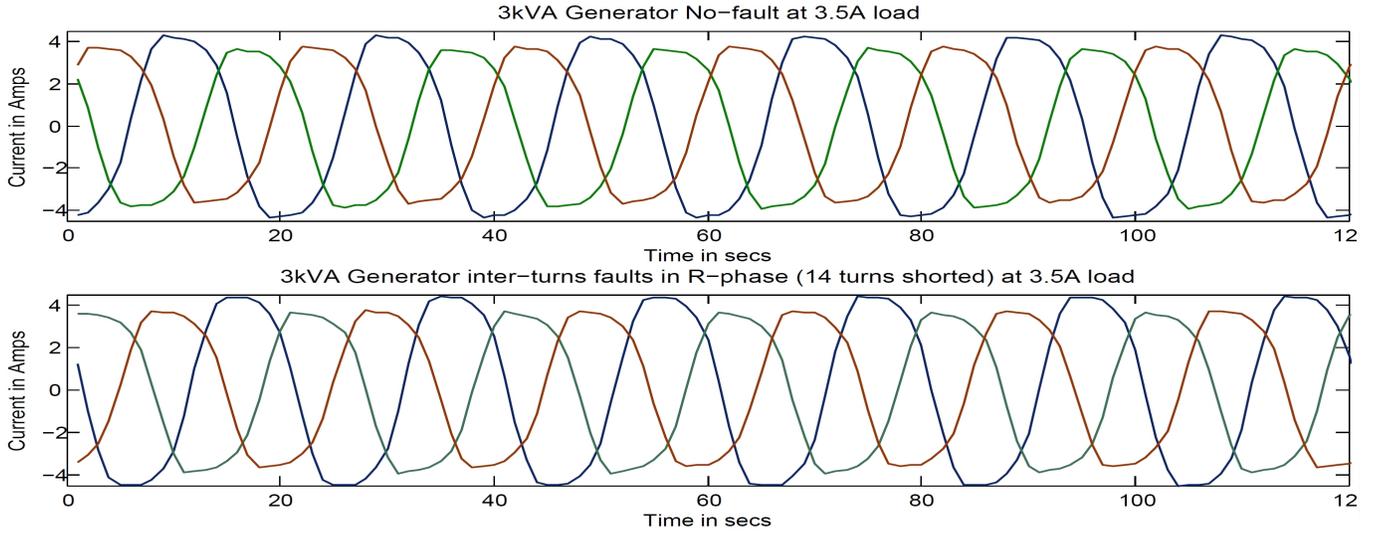


Figure 4. Current signature captured during no fault and inter turn fault conditions for the 3 kVA generator

$$dist(x_i, B) = [dist(x_i, b_1), \dots, dist(x_i, b_M)]^T$$

d_i is normalized to (0,1] by subtracting $max(dist(x_i, B))$ from $dist(x_i, B)$. σ adjusts the weight decay speed for the locality adaptor. The constraints $1^T c_i = 1$ indicates the shift invariant requirements of the LLC code. LLC selects the local bases from the basis set for each feature vector to form a local coordinate system using Eq. (3). LLC encoding process can be speeded up by using the k -nearest neighbors of x_i as the local bases B_i instead of using all the bases in the Eq. (3). This approach is called as **fast approximation LLC method** and it uses the following criteria (J. Wang et al., 2010):

$$\min_C \sum_{i=1}^N \|X_i - \tilde{c}_i B_i\|^2 \quad (5)$$

$$s.t. 1^T \tilde{c}_i = 1, \forall i$$

Fast approximation LLC method reduces the computation complexity from $O(M^2)$ to $O(M+k^2)$. In this paper, fast approximation LLC is used for its reduced computational complexity and fast encoding process (J. Wang et al., 2010).

2.4. Support Vector Machine (SVM)

SVM (Vapnik & Vapnik, 1998) is a supervised learning algorithm which is widely used for classification problems. SVM

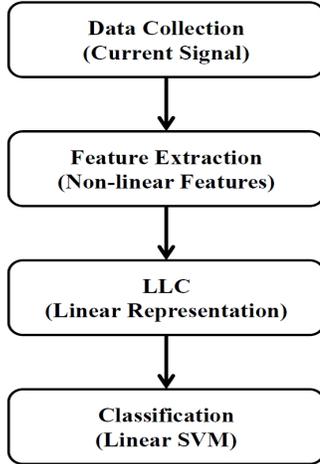


Figure 5. LLC coding process

constructs a hyperplane that separates two classes, and it can be extended to multi-class classification problems. Fig. 6, pictorially explains how an SVM classifier finds the separating hyper-plane. The planes parallel to the hyperplane are called bounding planes. The data points that lie on the bounding planes are called support vectors. The distance between the two bounding planes is the margin. The central hyperplane or classifier maximizes the margin and it can be obtained by finding the distance (d_1 and d_2) between each bounding plane to the origin respectively, and subtracting between the distances ($d = d_2 - d_1$). Margin of the classifier can be expressed as (Vapnik & Vapnik, 1998):

$$\frac{2}{\|w\|^2} \quad (6)$$

In order to maximize the margin, SVM learning is formulated by rewriting the Eq. (6) as a minimization problem, and it can be expressed as (Vapnik & Vapnik, 1998):

$$\min_{w,b} \frac{\|w\|^2}{2} \quad (7)$$

Subject to the constraint:

$$y_i[(w^T x_i) + b] \geq 1$$

where $x_i = [x_1, x_2, \dots, x_n]$ is the data set and $y_i \in \{1, -1\}$ be the class label of x_i . b is the bias. Using the Lagrange multipliers, the optimal solution can be computed by using the following equation:

$$\min_{w,b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i (w^T x_i - b) - 1] \right\} \quad (8)$$

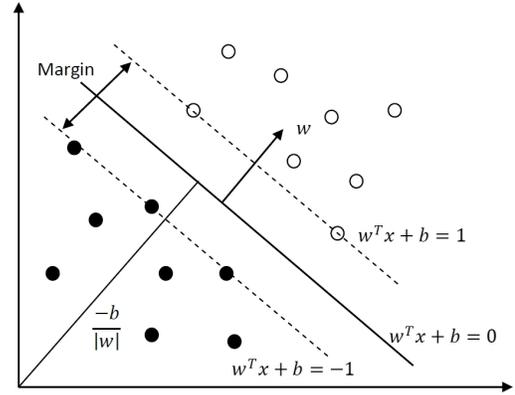


Figure 6. Linear separating hyperplanes for a separable case

where α_i is a Lagrange multiplier. Using Karush Kuhn Tucker (KKT) conditions (Kuhn & Tucker, 1951), solution can be expressed as the following:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (9)$$

$$b = \frac{1}{N_{SV}} \sum_{i=1}^N w^T x_i - y_i \quad (10)$$

Using the definition for w as defined in Eq. (9), the problem can be written in the dual form as:

$$\max_{\alpha_i} \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \right\} \quad (11)$$

$$\text{subject to } \alpha_i \geq 0 \text{ and } \sum_{i=1}^n \alpha_i y_i = 0.$$

where $k(x_i, x_j)$ is a kernel function. α_i can be obtained by solving the Eq. (11), and the decision function can be expressed as (Vapnik & Vapnik, 1998):

$$f(x) = \text{sign}(w^T x + b) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i x_i^T x - b\right) \quad (12)$$

When the data is not linearly separable, the non-linear kernel is applied for the classification, which transforms the features into a higher dimensional space, where it is linearly separable. In this paper, linear and non-linear SVM are experimented for the fault diagnosis. The complete process of the proposed approach is shown in the Fig. 7.

3. EXPERIMENTS AND RESULTS

In this work, the following experiments have been carried out for diagnosing the inter-turn faults in the 3 kVA synchronous generator.

1. Develop baseline system using linear, polynomial, radial basis function (RBF), and sigmoid SVM kernels for the fault classification
2. Improve the performance of the baseline system using feature mapping techniques, sparse coding and LLC.
3. Performance comparison of the feature mapping techniques using overall classification accuracy and receiver operating characteristic (ROC) curve.

Inter-turn faults are injected in the R, Y, and B phases of a 3 kVA generator stator winding. For every trial, current signals are acquired from all the three phases. Frequency domain statistical features are extracted from the raw data. Then, the experiments with the R, Y, and B phase inter-turn faults are treated as independent two class classification problems, i.e., no-fault or fault in R phase, no-fault or fault in Y phase and no-fault or fault in B phase, of a 3 kVA generator. It may be noted that an N class problem may be realized as N two class problems.

Experiments are conducted at different load conditions such as 0.5 A, 1 A, 1.5 A, 2 A, 2.5 A, 3 A, and 3.5 A loads. Data from these loads are combined together for fault classification. However, training and test data are collected separately, and no data is shared between the training and test sets. In this work, the experiments have been performed using balanced and unbalanced data sets to check the effectiveness of the proposed approach. Further, k-fold cross validation technique is also experimented using balanced data set separately. In addition, experiments using unseen load condition is performed to check the effectiveness of the proposed approach in removing load dependencies of the features. Experiments are carried out on a IBM X3100 M4² (Intel Xeon E3-1220v2 series) server, with 8 GB memory and 3.1 GHz Quad-core processor. MATLAB³ toolboxes are used for computing sparse⁴ and LLC⁵ codes.

²<http://www-03.ibm.com/systems/x/hardware/tower/x3100m4/>

³<http://in.mathworks.com/products/matlab/>

⁴http://stanford.edu/~boyd/l1_ls/

⁵<http://www.ifp.illinois.edu/~jyang29/LLC.htm>

3.1. Experiments using balanced data set

In this experiment, the data set used for training and test are nearly balanced (no-fault data: 55% and fault data: 45%). Details of the data sets used in our experiments are presented in Table 2.

Table 2. Total data collected for no fault and fault conditions (No-fault data:55%; Fault-data:45%)

	Machine condition		Total
	No fault	Inter-turn fault	
Train data	13300	10640	23940
Test data	5852	4683	10535
Total	19152	15323	34475

Baseline system: The baseline system is developed using SVM kernels such as linear, polynomial, RBF, and sigmoid. Table 3 lists the baseline system performance of the SVM kernels for the R, Y, and B phase faults. From the experiments, it is noted that RBF kernel performs better when compared to other kernels for the baseline system. It is observed that the performance of the classifier is not encouraging since the features does not match with the kernels and exhibit non-linear characteristics. The classification accuracies are generally low in the baseline system because the fault characteristics vary largely across the load conditions. The classification accuracy can be improved by removing this load dependencies of the features. In this paper, the objective is to improve the performance of fault diagnosis of generator by linearizing the features in a higher dimensional space using feature mapping techniques.

Table 3. Baseline system classification performance (accuracy in %) of SVM for the balanced data set

Kernel/Inter-turn fault	R phase	Y Phase	B Phase
Linear $c = 10, \gamma = 0.0256$	67.45	61.46	66.68
RBF $c = 10, \gamma = 0.0256$	76.66	81.35	82.31
Polynomial $c = 1, \gamma = 0.0256$	65.33	67.94	72.86
Sigmoid $c = 1, \gamma = 0.0256$	55.54	55.54	55.54

where c is the cost and γ is the kernel parameter.

Sparse coding: First, codebook (basis set or dictionary) is computed from the training data set. Subsequently, the feature vectors from the training and test data set are represented as a linear combination of the basis vectors from the codebook. In this process, codebook is common for the training and test data sets. Then the feature dimension is expanded from 39 (13 features per phase) to 256, 512, and 1024 feature vectors. Linear SVM is then used to classify the sparse represented features. Table 4 lists the classification perfor-

mance of the sparse coding for the inter-turn fault diagnosis of the generator. Experiments are performed for the different codebook sizes of 256, 512, and 1024, and obtained the improved performance for 256 codebooks. Sparse coding improves the performance of the baseline RBF kernel, from 76.66% to 87.35%, 81.35% to 87.37% and 82.31% to 91.14% for the R, Y and B phase faults respectively. Table 5 lists the improved classification performance and its computation time for sparse coding. Since the computation time is very large, it is not suitable for practical applications.

Table 4. Classification performance of sparse coding using linear SVM for the balanced data set

Inter-turn fault	Classification accuracy (%)		
	256 codebooks	512 codebooks	1024 codebooks
R Phase	87.35	87.40	87.23
Y Phase	87.37	86.91	86.82
B Phase	91.14	90.85	91.44

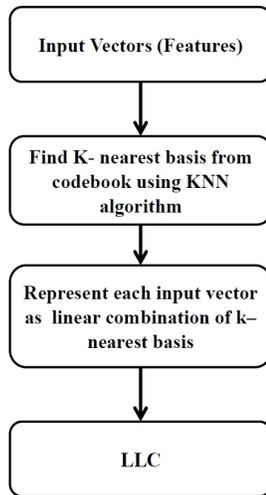


Figure 7. Machine fault diagnosis using locality constrained linear coding (LLC)

LLC: In LLC, finding basis vectors is similar to sparse coding, but the input data is represented as a linear combination of k nearest basis vectors. Experiments are performed with different codebook sizes and k -nearest neighbors (k -NN) to achieve a good classification performance for the fault iden-

Table 5. Improved classification performance and its computation time for sparse coding (256 codebooks)

Inter-turn fault	Accuracy (%)	CPU time (Mins)
R Phase	87.35	36.87
Y Phase	87.37	35.43
B Phase	91.14	39.93

tification system. Table 6 lists the performance of the LLC based linear SVM for the different codebook size and k -NN respectively. Codebook sizes 256, 512, and 1024 are used, and each codebook is experimented with 10, 20, 30, and 40 nearest neighbors for the R, Y, and B phase inter-turn faults respectively. Though codebook sizes 256 and 512 improves the performance, 1024 codebooks achieve the best classification performance. For 1024 codebooks with a selection of 40 nearest neighbors, LLC improves the baseline system (RBF kernel) by 22.56%, 18.43% and 17.05% absolute for the R, Y and B phase faults respectively.

Table 6. Classification performance of LLC using linear SVM for the balanced data set

Inter-turn fault	k -NN	Codebooks & accuracy (%)		
		256	512	1024
R Phase	10	98.33	98.19	98.50
	20	97.88	99.07	99.14
	30	96.43	99.12	99.15
	40	92.74	98.31	99.22
Y Phase	10	98.07	98.88	98.48
	20	98.01	99.34	99.17
	30	97.16	99.17	99.58
	40	95.95	97.96	99.78
B Phase	10	99.41	99.37	98.56
	20	99.20	99.64	99.31
	30	99.17	99.50	99.34
	40	99.01	99.23	99.36

Table 7. Improved classification performance and its computation time for LLC (1024 codebooks)

Inter-turn fault	kNN	Accuracy (%)	CPU time (Sec)
R Phase	40	99.22	70.45
Y Phase	40	99.78	63.90
B Phase	40	99.36	50.13

Table 7 lists the improved classification performance and its computation time for LLC. From our experiments, it is noted that LLC takes less computation time compared to sparse coding, therefore LLC reduces computational complexity and achieves improved performance for the classifier.

3.2. Experiments using unbalanced data set

In practical applications, injecting faults and collecting large number of fault data is not possible, to capture the intelligence about the system. This necessitates to analyze the fault identification system using smaller proportion of fault data, to check the effectiveness of the feature mapping techniques. In this work, the experiments were carried out using unbalanced data (No-fault data:80%; Fault-data:20%) by taking smaller proportion of fault data for the training. However, equal pro-

portion of fault and no-fault data is taken for the testing for a fair comparison of the results with the balanced data set experiments. Details of the data set used in our experiments is listed in Table 8. The performances of the baseline, sparse coding and LLC for the unbalanced data set are listed in the Tables 9 -11. From the results it is observed that, the performance of the baseline system, sparse coding and LLC have reduced for R, Y and B phases with the use of unbalanced data set. For the sparse coding, 256 codebooks performs better than other codebook sizes. However, when compared to the performance of the sparse coding for the balanced and unbalanced data set for 256 codebooks, the performance for the unbalanced data set got decreased by 4.27%, 2.10%, and 2.25% absolute for the R, Y, and B phases respectively. Similarly, the performance of the LLC with 1024 codebooks using balanced and unbalanced data set, the performance for the unbalanced data set got reduced by 3.84%, 1.77%, and 0.76% absolute for the R, Y, and B phases respectively. Though the unbalanced data set affect the performance of the classifier, the overall classification accuracy does not reduce significantly for the sparse coding and LLC, emphasizing these algorithms are suitable when data under fault conditions is scarce.

Table 8. Experiments using unbalanced data set (Training data: No-fault data:80%; Fault-data:20%, Test data: No-fault data:55%; Fault-data:45%)

	Machine condition		Total
	No-fault	Inter-turn fault	
Train data	13300	3346	16646
Test data	5852	4683	10535
Total	19152	8029	27181

Table 9. Baseline system classification performance (accuracy in %) of the SVM for the unbalanced data set

Kernel/Inter-turn fault	R phase	Y Phase	B Phase
Linear	62.71	57.30	60.54
$c = 5, \gamma = 0.0256$			
RBF	59.54	59.92	59.49
$c = 10, \gamma = 0.0256$			
Polynomial	59.28	63.78	72.67
$c = 20, \gamma = 0.0256$			
Sigmoid	55.54	55.54	55.54
$c = 1, \gamma = 0.0256$			

where c is the cost and γ is the kernel parameter.

3.3. ROC curve analysis

Receiver operating characteristic (ROC) curve is used to visualize and evaluate the classifier performance (Japkowicz & Shah, 2011). It shows the trade-off between the probability of detection or true positives rate (TPR), and the probability of false alarm or false positives rate (FPR). In this work, the performance of the baseline RBF system, sparse coding

Table 10. Classification performance of sparse coding using linear SVM for the unbalanced data set

Inter-turn fault	Classification accuracy (%)		
	256 codebooks	512 codebooks	1024 codebooks
R Phase	83.08	83.02	82.21
Y Phase	85.27	84.02	84.17
B Phase	88.89	88.66	88.45

Table 11. Classification performance of LLC using linear SVM for the unbalanced data set

Inter-turn fault	k -NN	Codebooks & accuracy (%)		
		256	512	1024
R Phase	10	95.67	94.60	93.76
	20	95.48	94.65	94.71
	30	93.29	94.86	94.93
	40	90.56	96.44	95.38
Y Phase	10	95.87	96.25	96.01
	20	94.69	97.69	97.27
	30	93.59	97.09	98.01
	40	94.30	94.40	97.82
B Phase	10	97.05	97.19	96.77
	20	98.32	97.84	97.86
	30	97.03	98.69	98.15
	40	97.31	98.25	98.60

and LLC are compared using ROC curve for the balanced and unbalanced data set (smaller proportion of fault data is considered for the training). Figure 8 -10 shows the performance comparison of the feature mapping techniques for the R, Y, and B phase faults. It is observed that area under curve (AUC) value becomes closer to 1 for the LLC compared to baseline RBF and sparse coding techniques. However, for the experiments using unbalanced AUC value got reduced for the baseline RBF, sparse coding and LLC.

3.4. k-fold cross validation using balanced data set

The experiments discussed in the subsections 3.1 and 3.2 were performed using one partition of data only. However, the use of fixed data set may over-fit the model. To overcome this problem, k-fold cross validation technique is used to assess the performance of the classifier. The balanced data set is used for evaluating the performance of the classifier. In this experiment, 19152 samples of no-fault and 15323 of fault samples are used for 10-fold cross validation. The performance of the baseline system, sparse coding and LLC for the 10-fold cross validation technique is listed in Table 12-14.

From the experiments it is noted that RBF kernel performs better than other kernels. Sparse coding improves the performance by 7.61%, 1.07% and 5.37% for the R, Y, and B phases respectively for 256 codebooks. Similarly, LLC improves the

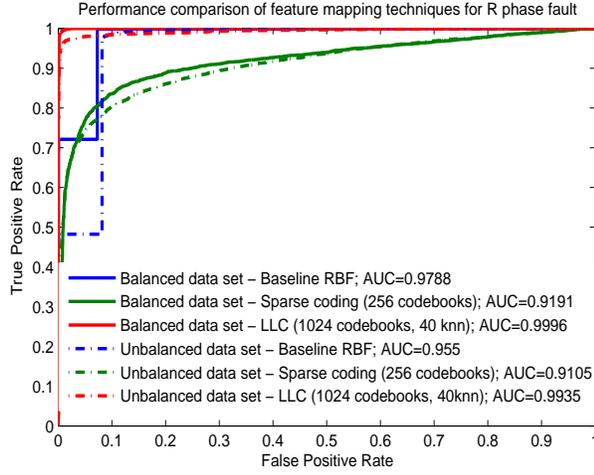


Figure 8. ROC plot for the performance comparison of feature mapping techniques for the R phase fault

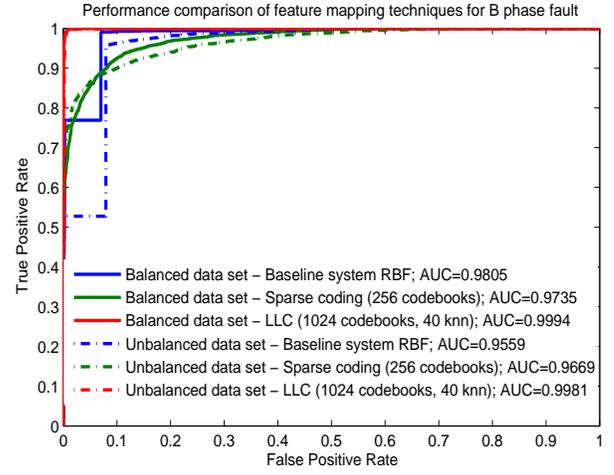


Figure 10. ROC plot for the performance comparison of feature mapping techniques for the B phase fault

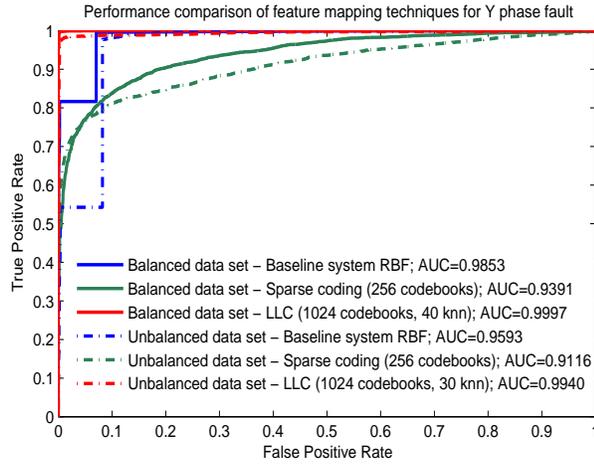


Figure 9. ROC plot for the performance comparison of feature mapping techniques for the Y phase fault

Table 12. Baseline system classification performance (accuracy in %) of the SVM using 10-fold cross validation

Kernel/Inter-turn fault	R phase	Y Phase	B Phase
Linear	60.16	58.31	65.70
$c = 10, \gamma = 0.0256$ RBF	80.01	85.77	86.02
$c = 10, \gamma = 0.0256$ Polynomial	63.05	67.16	65.43
$c = 30, \gamma = 0.0256$ Sigmoid	55.55	55.55	55.55
$c = 10, \gamma = 0.0256$			

where c is the cost and γ is the kernel parameter.

performance by 18.97%, 13.83%, and 13.76% for the R, Y, and B phases respectively for 1024 codebooks. It is observed that, the experiments using 10-fold cross validation technique

Table 13. Classification performance of sparse coding using linear SVM for the 10-fold cross validation

Inter-turn fault	Classification accuracy (%)		
	256 codebooks	512 codebooks	1024 codebooks
R Phase	87.62	87.48	87.25
Y Phase	86.84	86.55	86.63
B Phase	91.39	90.88	91.44

Table 14. Classification performance of LLC using linear SVM for the 10-fold cross validation

Inter-turn fault	k -NN	Codebooks & accuracy (%)		
		256	512	1024
R Phase	10	98.47	98.86	98.98
	20	98.10	99.08	98.32
	30	96.31	99.04	98.31
	40	93.47	98.03	98.42
Y Phase	10	98.00	98.61	98.67
	20	98.36	99.32	99.46
	30	98.04	99.29	99.60
	40	99.01	99.23	99.60
B Phase	10	99.41	99.37	99.42
	20	99.20	99.64	99.73
	30	99.17	99.50	99.75
	40	99.01	99.23	99.78

for the balanced data set does not affect the performance of the feature mapping techniques significantly compared to experiments using single partition of data.

3.5. Experiments on unseen load condition

Experiments are also carried out for unseen load condition of 3 kVA synchronous generator to check the effectiveness of LLC in removing the load dependencies of the features. The model is trained using 0.5A, 1A, 2A, 2.5A, 3A, and 3.5A load conditions and tested on 1.5A load. Total samples of 18240 and 1338 are used for training and testing, respectively, with an equal proportion of no-fault and fault data. From the experiments, it is observed that the best performance for the baseline system is obtained using linear kernel with an overall classification accuracy of 59.04%, 72.79% and 80.94% for the R, Y, and B phases respectively. The improved performance is obtained using LLC with an overall accuracy of 73.46%, 78.70%, and 87.66% by selecting 8, 23, and 36 nearest neighbors for the R, Y, and B phases respectively, with the codebook size of 64. Experimental results show that LLC could perform better even if unseen load condition is used for fault diagnosis.

4. CONCLUSIONS

In this paper, feature mapping algorithms, sparse coding and LLC are used to improve the performance of SVM for the inter-turn fault identification of 3kVA synchronous generator. As the features are non-linear, feature mapping techniques are used to linearize the features in a higher dimensional space to improve performance of the fault diagnosis system. Experiments are performed for the balanced and unbalanced data using single partition of data. Sparse coding improves the performance significantly with a high computational cost. Therefore, LLC is used to reduce the computational complexity and enhance the performance of the system. By using the balanced data set, 1024 codebooks with 40 nearest neighbors are selected empirically for its best performance. LLC improves the overall fault identification accuracy of the baseline RBF system by 22.56%, 18.43% and 17.05% absolute for the R, Y and B-phase faults respectively. The performance of the feature mapping techniques are also illustrated through ROC curves. The AUC value becomes closer to one for LLC compared to the sparse coding and baseline RBF system. The performance of the classifier is also assessed using 10-fold cross validation technique. From the experiments, it is observed that LLC outperforms sparse coding in terms of classification performance and computational cost for the inter-turn fault diagnosis of the synchronous generator. Though LLC has been used widely in image classification problems, the reported experimental results show that, LLC could be used for other applications also if the features used in the system does not match with SVM kernels and exhibits non-linear characteristics.

ACKNOWLEDGEMENT

Authors would like to acknowledge the financial support provided by Aeronautical Development Agency (ADA) -National Programme on Micro and Smart Systems (NPMAS), India. Authors thank Dr. Vanam Upendranath and Mr. P. V. R. Sai kiran, Structural Technologies Division, CSIR- National Aerospace Laboratories, Bangalore, India for their help during the course of this work. Authors also thank Mr. Pushparajan, M., Mr. Kuruvachan K. George., and Mr. Sreekumar, K. T., for their support through out this work.

REFERENCES

- Casimir, R., Boutleux, E., Clerc, G., & Yahoui, A. (2006). The use of features selection and nearest neighbors rule for faults diagnostic in induction motors. *Engineering Applications of Artificial Intelligence*, 19(2), 169-177.
- Chen, Y. D., Du, R., & Qu, L. S. (1995). Fault features of large rotating machinery and diagnosis using sensor fusion. *Journal of Sound and Vibration*, 188(2), 227-242.
- Chiang, L. H., Kotanchek, M. E., & Kordon, A. K. (2004). Fault diagnosis based on fisher discriminant analysis and support vector machines. *Computers & chemical engineering*, 28(8), 1389-1401.
- Deng, S., Jing, B., & Zhou, H. (2014). Fusion sparse coding algorithm for impulse feature extraction in machinery weak fault detection. In *Prognostics and system health management conference (phm-2014 human)* (p. 251-256).
- Fu, M., Tian, Y., & Wu, F. (2015). Step-wise support vector machines for classification of overlapping samples. *Neurocomputing*, 155, 159-166.
- Gopinath, R., Nambiar, T. N. P., Abhishek, S., Pramodh, S. M., Pushparajan, M., Ramachandran, K. I., ... Thirugnanam, R. (2013). Fault injection capable synchronous generator for condition based maintenance. In *Intelligent systems and control (isco), 2013 7th international conference on ieee* (p. 60-64).
- He, Q., Liu, Y., & Kong, F. (2011). Machine fault signature analysis by midpoint-based empirical mode decomposition. *Measurement Science and Technology*, 22(1), 015702.
- Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: a classification perspective*. Cambridge University Press.
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition based maintenance. *Mechanical systems and signal processing*, 1483-1510.
- Kang, M., Kim, J., Kim, J., Tan, A., Kim, E. Y., & Choi, B. (2015). Reliable fault diagnosis for low-speed bearings using individually trained support vector machines

- with kernel discriminative feature analysis. *Power Electronics, IEEE Transactions on*, 30(5), 2786-2797.
- Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the second berkeley symposium on mathematical statistics and probability* (p. 481-492). University of California Press.
- Lei, Y., He, Z., & Zi, Y. (2008). A new approach to intelligent fault diagnosis of rotating machinery. expert systems with applications. *Expert Systems with Applications*, 1593-1600.
- Liang, Y., Song, M., Bu, J., & Chen, C. (2014). Colorization for gray scale facial image by locality-constrained linear coding. *Journal of Signal Processing Systems*, 74(1), 59-67.
- Lin, J., & Zuo, M. J. (2004). Extraction of periodic components for gearbox diagnosis combining wavelet filtering and cyclostationary analysis. *Journal of vibration and acoustics*, 126(3), 449-451.
- Liu, H., Liu, C., & Huang, Y. (2011). Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 25(2), 558-574.
- Lu, S., Wang, Z., Mei, T., Guan, G., & Feng, D. D. (2014). A bag-of-importance model with locality-constrained coding based feature learning for video summarization. *Multimedia, IEEE Transactions on IEEE*, 16(6), 1497-1509.
- Nayak, J., Naik, B., & Behera, H. S. (2015). A comprehensive survey on support vector machine in data mining tasks: Applications and challenges. *International Journal of Database Theory and Application*, 8(1), 169-186.
- Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23), 3311-3325.
- Peter, W. T., Peng, Y. H., & Yam, R. (2001). Wavelet analysis and envelope detection for rolling element bearing fault diagnosis their effectiveness and flexibilities. *Journal of Vibration and Acoustics*, 123(3), 303-310.
- Rahmani, H., Mahmood, A., Huynh, D., & Mian, A. (2014). Action classification with locality-constrained linear coding. In *Pattern recognition (icpr), 2014 22nd international conference on ieee* (p. 3511-3516).
- Samanta, B., & Al-Balushi, K. R. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical systems and signal processing*, 17(2), 317-328.
- Saxena, A., Wu, B., & Vachtsevanos, G. (2005, June). A methodology for analyzing vibration data from planetary gear systems using complex morlet wavelets. *Proceedings of IEEE American Control Conference*, 4730-4735.
- Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory* (Vol. 2). New York: Wiley.
- Verron, S., Tiplica, T., & Kobi, A. (2008). Fault detection and identification with a new feature selection based on mutual information. *Journal of Process Control*, 18(5), 479-490.
- Wang, B., Gai, W., Guo, S., Liu, Y., Wang, W., & Zhang, M. (2014). Spatially regularized and locality-constrained linear coding for human action recognition. *Optical Review*, 21(3), 226-236.
- Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., & Gong, Y. (2010). Locality-constrained linear coding for image classification. *IEEE Computer Vision and Pattern Recognition (CVPR) conference*, 3360-3367.
- Wang, X., & Pardalos, P. M. (2015). A survey of support vector machines with uncertainties. *Annals of Data Science*, 1(3-4), 293-309.
- Widodo, A., & Yang, B. S. (2007). Application of non-linear feature extraction and support vector machines for fault diagnosis of induction motors. *Expert Systems with Applications*, 33(1), 241-250.
- Widodo, A., Yang, B. S., & Han, T. (2007). Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert Systems with Applications*, 32(2), 299-312.
- Wu, B., Saxena, A., Khawaja, T. S., Patrick, R., Vachtsevanos, G., & Sparis, P. (2004, September). An approach to fault diagnosis of helicopter planetary gears. *Proceedings of IEEE AUTOTESTCON*, 475-481.
- Wu, B., Saxena, A., Patrick, R., & Vachtsevanos, G. (2005, July). Vibration monitoring for fault diagnosis of helicopter planetary gears. In *Proceedings of 16th IFAC World Congress*.
- Yan, R., & Gao, R. X. (2008). Rotary machine health diagnosis based on empirical mode decomposition. *Journal of Vibration and Acoustics*, 130(2), 21007.
- Yan, R., Gao, R. X., & Wang, C. (2009). Experimental evaluation of a unified time-scale-frequency technique for bearing defect feature extraction. *Journal of Vibration and Acoustics*, 131(4), 041012.
- Yang, J., Yu, K., Gong, Y., & Huang, T. (2009). Linear spatial pyramid matching using sparse coding for image classification. *IEEE Computer Vision and Pattern Recognition conference*, 1794-1801.
- Yang, Y., Liao, Y., Meng, G., & Lee, J. (2011). A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis. *Expert Systems with Applications*, 38(9), 11311-11320.
- Yu, K., Zhang, T., & Gong, Y. (2009). Nonlinear learning using local coordinate coding. *Advances in neural information processing systems*, 2223-2231.
- Zhang, K., Li, Y., Scarf, P., & Ball, A. (2011). Feature selection for high-dimensional machinery fault diagnosis data using multiple models and radial basis function networks. *Neurocomputing*, 74(17), 2941-2952.

- Zhang, P., Wee, C., Niethammer, M., Shen, D., & Yap, P. (2013). Large deformation image classification using generalized locality-constrained linear coding. In *Medical image computing and computer-assisted intervention-miccai 2013* (p. 292-299). Springer.
- Zhang, Y. (2009). Enhanced statistical analysis of non-linear processes using kpca, kica and svm. *Chemical Engineering Science*, 64(5), 801-811.
- Zhou, J., Shi, J., & Li, G. (2011). Fine tuning support vector machines for short-term wind speed forecasting. *Energy Conversion and Management*, 52(4), 1990-1998.
- Zhou, X., Cui, N., Li, Z., Liang, F., & Huang, T. S. (2009, September). Hierarchical gaussianization for image classification. *IEEE Conference on Computer Vision*, 1971-1977.
- Zhu, H., Wang, X., Zhao, Y., Li, Y., Wang, W., & Li, L. (2014). Sparse representation based on adaptive

multiscale features for robust machinery fault diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*.

APPENDIX A

Synchronous generator specifications

Parameter	Value
Rated power	3kVA
Rated voltage	415V
Rated frequency	50Hz
Connection type	star
Number of poles	4
Number of phases	3
Speed	1500 rpm
Current	4.2A
Power factor	0.8