Fault Diagnosis of Reciprocating Compressors Using Revelance Vector Machines with A Genetic Algorithm Based on Vibration Data

M. Ahmed, A. Smith, F. Gu, A.D. Ball
University of Huddersfield
Huddersfield, UK
u0870325@hud.ac.uk

Abstract—**This paper focuses on the development of an advanced fault classifier for monitoring reciprocating compressors (RC) based on vibration signals. Many feature parameters can be used for fault diagnosis, here the classifier is developed based on a relevance vector machine (RVM) which is optimized with genetic algorithms (GA) so determining a more effective subset of the parameters. Both a one-against-one scheme based RVM and a multiclass multi-kernel relevance vector machine (mRVM) have been evaluated to identify a more effective method for implementing the multiclass fault classification for the compressor. The accuracy of both techniques is discussed correspondingly to determine an optimal fault classifier which can correlate with the physical mechanisms underlying the features. The results show that the models perform well, the classification accuracy rate being up to 97% for both algorithms.**

Keywords-Reciprocating Compressor; Relevance Vector Machine; Fault Diagnosis; Genatic Algorithms; multiclass multi-kernel relevance vector machine.

I. INTRODUCTION

Reciprocating compressors play a major part in many industrial systems and faults occurring in them can degrade performance, consume additional energy, cause severe damage to the machine and possibly even result in system shut-down. The vibration signal from a reciprocating compressor contains non-linear characteristics (e.g. due to the impacts resulting from the movement of the suction and discharge valves), features extracted from the time, frequency and envelope domains of these signals can be used to assess the health of the system. Unfortunately, not all the extracted features are equally useful in trouble-shooting and experience has shown that even the most useful features are seldom used in the most effective way. In particular the interactions between and among features are not fully considered or even ignored [1] which may undermine the accuracy of diagnosis when the features employed are synergetic.

There have been many attempts made to diagnose and classify earlier faults from reciprocating compressors based on vibro-acoustic measurements. Gu and Ball [2] presented the use of a smooth pseudo-Wigner–Ville distribution for interpretation of machinery vibration data from RC for diagnosing various valve faults.

Support vector machine (SVM) methods were employed in order to classify faults of reciprocating refrigeration compressors through the application of wavelet transform and statistical methods. Significant features were extracted from both acoustic signals and vibration signals. The selection of relevant radial basis function (RBF) kernel parameters was carried out through iteration [1] for more accurate classification of healthy and faulty compressors. In a similar application, SVM methods were applied to reciprocating compressor butterfly valves to classify cavitation faults[3]. Comparable research was performed on reciprocating compressor valves to classify faults through vibration signals alone. Data for this purpose was gathered from the surface of the valve and the resulting vibration signals were decomposed by applying local wave methods[4].

However, studies employing SVM methods have investigated only a small number of faults, which is not sufficient for practical applications. Moreover, recent studies show RVM performs better in separating multiple classes. Compared to SVM, the RVM method required fewer kernel functions and less learning time while demonstrating comparable performance [5]. Therefore, this paper examines the performance using RVM to classify more diverse faults on a compressor.

II. VIBRATION CHARACTERSTICS OF RECIPROCATING COMPRESSORS AND DIAGNOSTIC FEATURES

A. Dataset aquistion

The vibration signals were collected from the accelerometers on the two-stage, single-acting Broom Wade TS9 reciprocating compressor, which has its two cylinders in the form of a "V" (see Fig 1), and which delivers compressed air at up to 0.8 MPa (120 psi) to a horizontal air receiver tank with a maximum working pressure of about 1.38MPa (200psi), sampled at a rate of 55.56 kHz (this enables the high frequencies associated with transient events to be collected); the data length is set at 118833 samples. The time duration of data points = number of samples $\div$ sampling frequency so the real time duration of the samples is 2.14 sec. These operations repeated 3 times for 12 different discharge pressures. Each segment of data includes more than four working cycles of the compressor.

To characterise vibrations for different faults, seven faults were separately seeded into the compressor: a leaky valve in the high pressure cylinder, a leaky intercooler, a loose drive belt and three more combined faults:

discharge valve leakage combined with intercooler leakage, suction valve leakage combined with intercooler leakage and discharge valve leakage combined with suction valve leakage. These faults produce little noticeable influence on the performance of generating pressures but do need to use more electrical energy than that of a healthy compressor.

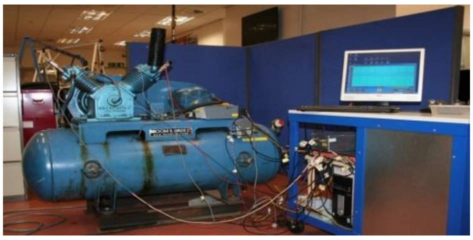

Figure 1 Broom-Wade TS9 reciprocating compressor

B. *Envelope Spectrum Features*

Rather than using the time domain data and its spectrum directly, the frequency based envelope spectrum is produced and used for feature selection. As shown in Fig 2 envelope spectra for different cases clearly exhibit a number of discrete components (mainly due to the compressors fundamental working frequency 7.3Hz and associated harmonics), in contrast, the spectra from the raw data show continuous spectral features from which it is more difficult to select a small number of feature components. Nevertheless, it can be seen in the envelope spectra that the amplitudes vary slightly but significantly between fault cases, making it difficult to establish a simple set of features for accurate separation. Thus the amplitudes of these components were included as candidate features with different harmonics used for each trial run. The resultant feature dataset being an $(n \times (sc))$ matrix (n the number of harmonics, s the number of samples and c the number of cases). Here, with 30 harmonics for each of 40 samples from eight different cases the dataset array has size (30 ×320).

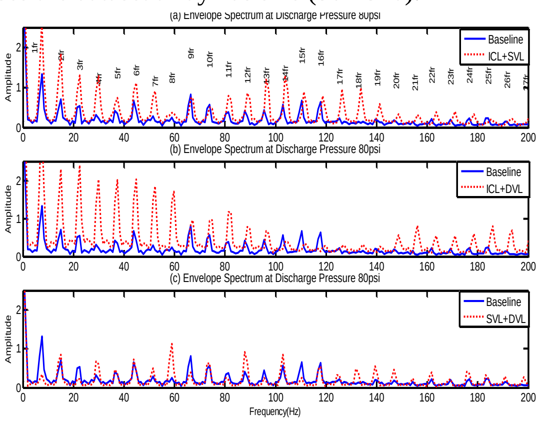

Figure 2 Envelope spectra of compressor vibration for healthy case and three seeded faults

III. Relevance Vector Machines (RVM)

The Relevance Vector Machine (RVM), introduced by M E. Tipping [6], is a probabilistic sparse kernel model and is analogous to support vector machines (SVMs). It adopts a Bayesian approach to learning, by introducing a prior density over the weights, governed by a set of hyper parameters, whose most probable values are iteratively estimated from the data. Sparsity is achieved because in practice the posterior distributions of many of the weights are sharply peaked around zero. Furthermore, unlike the support vector classifier, the non-zero weights in the RVM are not associated with examples close to the decision boundary, but rather appear to represent prototypical examples of classes, the relevance vectors. The most compelling feature of the RVM is that it typically utilises dramatically fewer kernel functions, whilst being generally capable of performance comparable to an equivalent SVM. Furthermore, the RVM suffers from none of the other limitations of the SVM, as it is a probabilistic model.[7].

RVM has also been used for classification. Consider a two-class problem with training points $X = [x_1, \dots x_N]$ and corresponding class labels $t = [t_1, \dots t_N]$ with $t_i = [0,1]$. Based on the Bernoulli distribution, the likelihood (the target conditional distribution) is expressed as:

$$p(t \backslash W) = \prod_{i=1}^{N} \sigma((y(x_i)))^{t^i} [1 - \sigma((y(x_i)))]^{1-t_i} \qquad (1)$$

Where $\sigma(y)$ the sigmoid function:

$$\sigma(y(X)) = \frac{1}{1+\exp(-y(x))} \qquad (2)$$

Unlike the regression case, however, the marginal likelihood $p(t \backslash \alpha)$ can no longer be obtained analytically by integrating the weights from "(1)" an iterative producer has to be used [8].

Let α_i^* denote the maximum a posterior (MAP) estimate of the hyper parameter α_i. The MAP estimate for the weights, W_{MAP}, can be obtained by maximising the posterior distribution of the class labels given the input vectors.

$$J(w_1, \dots w_N) = \sum_{i=1}^{N} \log p(t_i \backslash w_i) + \sum_{i=1}^{N} \log p(w_i \backslash \alpha_i^*) \qquad (3)$$

The first summation term denotes the likelihood of class labels and the second term denotes the prior parameters, w_i. In the resulting solution, only those samples associated with nonzero coefficients w_i (relevance vectors) will contribute to the decision function.

The gradient of the objective function J with respect to w is: $\nabla J = -A^* W - \varphi^T (f - t)$ (4)

where $f = [\sigma(y(x_1)) \dots \sigma(y(x_N))]^T$, matrix φ has elements $\emptyset_{i,j} = K(x_i, x_j)$.

The Hessian of J is: $H = \nabla^2(J) = -(\varphi^T B \varphi + A^*)$ (5)

where $B = \mathrm{diag}(\beta_1, \dots \beta_N)$ is a diagonal matrix $\beta_1 = \sigma(y(\beta x_i))[1 - \sigma(y(\beta x_i))]$.

The posterior is approximated around W_{MAP} by a Gaussian approximation with covariance

$$\sum -H \backslash W_{MAP}{}^{-1} \qquad (6)$$

and mean $\mu = \sum \varphi^T Bt$ (7)

As can be seen from Fig 3 and Fig 4, SVM and RVM respectively. In classifying samples of the same problem, the SVM has selected 24 important samples that lie on the class boundaries while the RVM has selected only 4 samples which are prototypical of the class they represent.

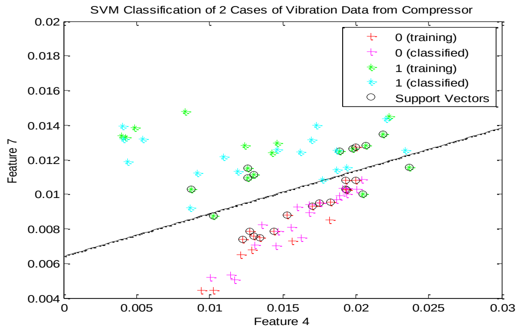

Figure 3 SVM binary classifier: Healthy class and DVL with two input features

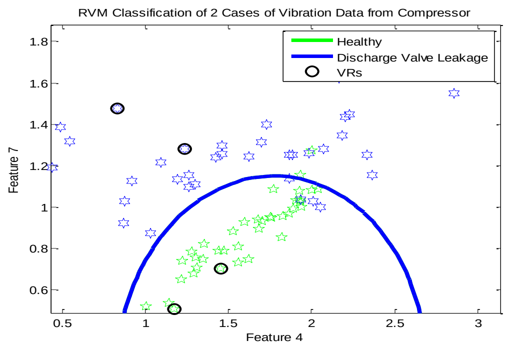

Figure 4 RVM binary classifier: Healthy class and DVL with two input features

The most compelling feature of the RVM is that it achieves comparable recognition accuracy to the SVM yet provides a full predictive distribution and requires dramatically fewer kernel functions. The RVM, therefore, consumes much less test time, a most important consideration in practice, for example in on-line fault detection [9].

IV. Results and Discussion

As shown in Table I there are eight classes of data sets corresponding to different healthy and faulty cases of the compressor.

TABLE I Fault Situation in the Condition Process

Classes	*Class Description*
H	Healthy
DVL	Discharge Valve leakage
SVL	Suction Valve leakage
LB	Loose Drive Belt
IL	Intercooler Leakage
DVL+SVL	Discharge Valve leakage with Suction Valve leakage combined fault
SVL+IL	Suction Valve leakage with Intercooler combined fault
DVL+IL	Discharge Valve leakage with Intercooler combined fault

Based on previous studies, the harmonics of compressor operation frequencies obtained from envelope spectra are the most effective features for differentiating between the cases. So the same harmonic components are adopted for RVM based classification.

A preliminary application of the RVM algorithm from M E Tipping [6, 10] has found that the algorithms are not stable when the feature size (number of harmonics) is larger than 16. In particular, there are often errors with ill conditioned Hessian matrices due to poor conditioning of numbers in the dataset. In addition, a previous study identified that the 1st harmonic component performs less well in separating the cases. Therefore, only the harmonic components from 2 to 15 are used in the RVM application.

80 samples per case were collected, covering the rated operating pressure range from 70psi to 120psi (4.83 to 6.90 bar). In total the data matrix is 15x80x8. For RVM training a random sample of 40 is selected from each class. The remaining 40 data values being used to validate the trained RVMs.

In addition, One against One (OAO) binary classifiers are used for the multiclass classification because the training sets are smaller and the problems to be learned are usually easier. Since the classes have less overlap if k is large and we need to evaluate the $k(k-1)/2$ classifiers, then the resulting system may be slower than the corresponding one-against-all RVMs [11].

For instance in this study, if $k = 8$, one needs to train 28 binary classifiers rather than 8 classifiers as in the method above. Although training time increases, the individual problems that need to be trained are significantly smaller. Furthermore, if the training algorithm scales super linearly with the training set size, it is possible to save processing time. This is related to the runtime execution speed. To classify a test pattern all 28 binary classifiers need evaluating and classifying according to the classes which get the highest number of votes. A vote for a given class is defined as a classifier putting the pattern into that class. The individual classifiers, however, are usually smaller in size (they have fewer RVs) than they would in the one-against-all approach.

A. RVM without GA feature selection

In this section classification results are presented from the straight use of RVMs without feature selection. However, as RVM is virtually a binary classifier, usually two classification schemes are adopted to perform multiclass problems. The first one is OAO and the second one is one against all (OAA). According to previous research OAO outperforms OAA in that it can be trained more efficiently and subsequently obtain better classification results [11, 12] . Therefore, this study only examines the performance of OAO.

The classification performance of the RVM based on the OAO scheme (OAO-RVM) is shown in Fig 5. To further examine performance the results are also presented in Table II. Fig 5 shows the performances of RVM classification rate on extracted features of the vibration signal (%) of RVM-OAO using Gaussian kernel for testing datasets for different classes. An overall accuracy of 95.95% was achieved using the technique developed.

TABLE II AVERAGE CLASSIFICATION RATES OF THE TRAINED CLASSIFIERS

Class no	***Cases***	***Input Harmonics***	***Classification Rate (test data) (%)***
Overall	Overall	2-15	95.95
C_1	Healthy	2-15	90.00
C_2	DVL	2-15	100.00
C_3	SVL	2-15	97.50
C_4	LB	2-15	95.00
C_5	IL	2-15	95.00
C_6	DVL+SVL	2-15	100.00
C_7	SVL+IL	2-15	90.00
C_8	DVL+IL	2-15	100.00

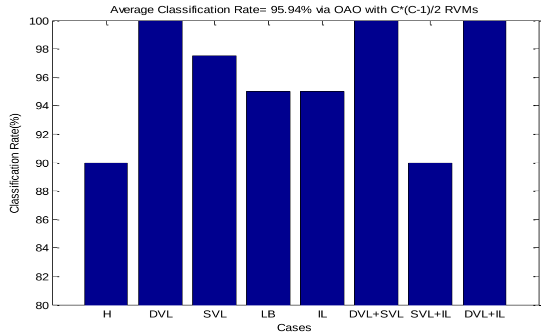

Figure 5 RVM–OAO Classification rates of different cases

The errors occurred in classification of data between Classes 1(Healthy), 3(SVL), 4(LB), 5(IL) and 7(SVL+IL) due to proximity of classes and some of them having the same fault characteristic(s). Table III displays the misclassification between classes. It is confirmed that the suggested RVM-OAO algorithm has high classification accuracy with small failure in some cases.

B. RVM with GA feature selection

A GA-based parameter method was used to automatically obtain the optimal features of the RVM classifier. The classification rate of the training data is the fitness function of the GA. Therefore, the optimum features are achieved when a minimum error is detected by the classifier to complete the generation. Also for the GA-RVM-OAO method, the population size is considered to be equal to 20. The initial range is taken to be within [0, 2] for all individuals.

TABLE III MISCLASSIFICATION BETWEEN CLASSES

Class no	***Cases***	***Misclassification classes***	***No of misclassified Samples***	***Misclassific-ation Rate (%)***
1	Healthy	IL, SVL	4	10
2	DVL	-	0	0
3	SVL	Healthy	1	2.5
4	LB	DVL+SVL	2	5
5	IL	Healthy	2	5
6	DVL+SVL	-	0	0
7	SVL+IL	DVL+SVL	4	10
8	DVL+IL	-	0	0

In this section, two kinds of experiments were carried out. The first objective being the accuracy of classification rate, the second objective the output of feature selection subset with a high rate of classification.

Fig 6 shows the results of GA classification performance. It is remarkable the influence of the classification using GA-RVM-OAO approach, with an accuracy rate close to 97.00%.

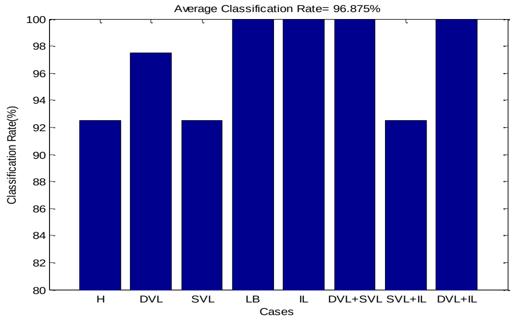

Figure 6 RVM–OAO Classification rates of different cases based on GA-based parameter selections

Moreover, it should be noted that a significant reduction of the number of features is achieved by GA optimisation. As shown in Fig 7, an initial set of 15 variables is reduced to feature subsets with no more than 10 features in all RVMs training. This allows the online implementation of RVM to be more efficient.

Furthermore, the relation between the final feature subset selected and the classifier used in the fitness function is also taken into account. Best results are achieved when using the same function in the classifier as in the feature selection and classification stage.

As shown in Fig 8, the proposed indexing method makes use of a knowledge-base consisting of relationships between features and concept classes. The relationships are not necessarily trivial, thus the key of this method lies in acquiring distinction. This figure demonstrates empirically that the set of relevant features is individual for each class. Indeed whilst there are features shared by all the classes there are other features which are entirely absent from a class. In particular, feature 11 is not selected for class 1(healthy case), feature 12 is not used in class 2(DVL) while features 5, 11 and 13 are not selected in class number 6 (DVL+SVL). On the other hand all classes utilise feature 10.

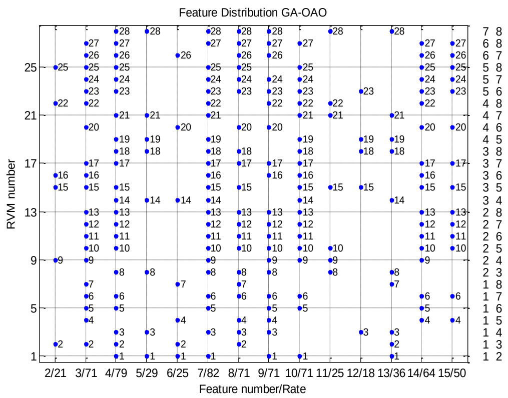

Figure 7 Features Selection with GA-RVM

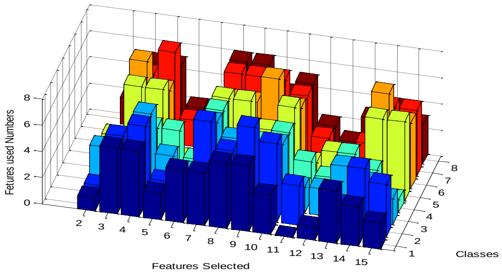

Figure 8 Relating features with classes

Fig 9 shows the mutual information between each feature and diagnosis. The estimation is based on all available cases of the dataset. Clearly, features 4, and 7 stand out as the most informative ones for the diagnosis of reciprocating compressors accounting for 85% and 80% of the total selected, respectively. Features 3, 8, 9, 10 and 14 are reasonably predictive of reciprocating compressor diagnosis with 70%. The remaining features number 2, 6, 11, 12 and 15 showed close to chance behaviour ranging between 20% and 30%.

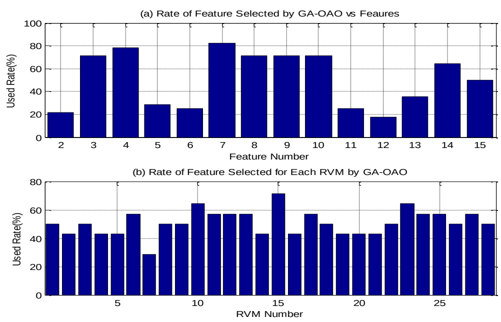

Figure 9 Performance of GA- OAO-RVMs for different number of selected features

TABLE IV RELATING FEATURES WITH CLASSES

	Features													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C_1	1	5	5	2	4	4	5	5	3	0	1	4	3	2
C_2	1	5	6	2	1	7	5	7	6	3	0	2	5	4
C_3	3	4	6	3	2	6	5	3	5	2	2	4	3	2
C_4	2	3	4	4	2	6	2	4	5	3	2	4	3	1
C_5	2	6	6	1	1	6	6	4	6	2	3	1	6	6
C_6	1	7	5	0	2	5	5	7	5	0	1	0	7	5
C_7	0	5	7	2	1	6	6	6	5	2	0	2	5	5
C_8	2	5	5	2	1	6	6	4	5	2	1	3	4	3

TABLE V THE MOST FEATURES USING IN EACH CLASS

Class	**Cases**	**Features used**
C_1	Healthy	3,4,6,7,8,9,13
C_2	DVL	3,4,7,8,9,14
C_3	SVL	3,4,7,8,10,13
C_4	LB	4,5,7,9,10,13
C_5	IL	3,4,7,8,9,14,15
C_6	DVL+SVL	3,4,7,8,8,10,14
C_7	SVL+IL	3,4,7,8,9,10,14,15
C_8	DVL+IL	3,4,7,8,9,10,14

From Table IV, it is apparent that the RVM-OAO based on GA techniques selected features very different from the ones previously selected in each case. The different selections affected the diagnosis performance to achieve separation of the classes. The RVM training utilised two important features, 4 and 7, selected based on GA for all classes separation, feature subsets per class are displayed in Table V.

C. Multi-class Relevance Vector Machine Classification

The main purpose of implementing multi-class relevance vector machine (mRVMs) algorithms is that they can provide high predictive accuracy while retaining computational efficiency. The mRVMs algorithms produce more sparse solutions both sample and kernel-wise which enables their application to large-scale multi-feature multinomial classification complications [13].

In addition to identifying the key elements of a dataset, another important issue is being able to capture predictive error in a systemic way.

To evaluate the performance of mRVM in classifying compressor vibration data, the mRVM model was trained for four classes of samples: healthy, IL, DVL+IL and SVL+DVL with two input parameters to the model. The generalisation performance and accuracy rate of the classifier model was then verified. The classification results of mRVM in Fig 10 shows that a mRVM classifier model with two input harmonics has good generalization performance and high classification accuracy with low error. See Table VI.

TABLE VI CLASSIFICATION RESULTS FOR FOUR CLASSES WITH TWO HARMONICS

State	***Number of Samples***	***Correct classification***	***Accuracy***	***Error Rate***
Healthy	40	40	100%	0.0125%
IL	40	40	100%	
SVL+IL	40	39	97.5%	
ICL+DVL	40	39	97.5%	
Overall			98.75%	

Similarly, this approach has also been applied for the defect diagnostics for more than 4-classes considering multiple defects with different numbers of input harmonics that ranged between 2 to 15 as explained in section *A*.

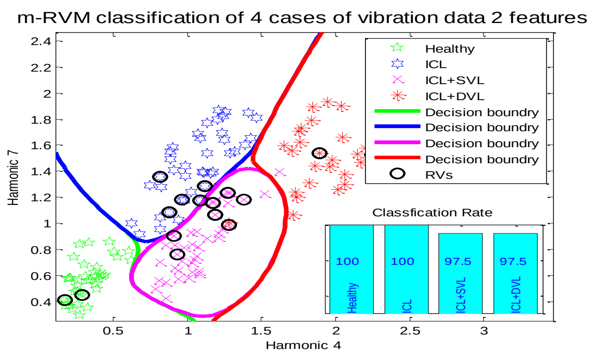

Figure 10 Classification results using mRVM four classes with two input harmonics

Table VII shows results for the mRVM. The harmonic peaks in the spectral analysis provided a total of 14 possible features for use in classification of the seeded faults, the number of peaks in the spectrum which are used for classification. The first column refers to the number of input cases. The second column represents the number of features, ranging between 2 to 15. For example, 12 features are tested in the first row (between 4-15) the peaks which gave the best classification rate were chosen. The correct classification rate was 98.75%, but as the number of features decreased the classification rate dropped quite sharply. The classification rates were 24.16%, obtained with 5 features and 8 cases as shown in row 7. This is due to the bad scaling of the type-II ML procedure with respect to the number of classes [13, 14] and the dimensionality of the Hessian required for the Laplace approximation [14, 15].

TABLE VII CLASSIFICATION RESULTS FOR DIFFERENT CLASSES WITH DIFFERENT HARMONICS ORDER

Classes	***Harmonic Orders***	***Accuracy***	***Error Rate***
1,4,6,7	4-15	98.75%	0.0125
1,4,5,7,8	4-15	86%	0.160
1,2,4,5,7,8	2-15	73.75%	0.2625
1,2,3,4,5	2-15	77.0%	0.23
1,2,3,4,5,6,7	2-15	59.29%	0.4071
1-8	4,5,6,10,15	24.16 %	0.3969
1-8	2-15	50%	0.50

Looking at Fig 10 it can be seen that some faults are easier to identify than others using the mRVM. For example there are large boundaries for healthy, IL, DVL+IL and SVL+DVL, so those four cases could be identified relatively easily. However, as the requirement is that the four faults be correctly identified, a relatively large number of peaks are required for high classification rates. Moreover, the algorithm is very sensitive in that it can only accept a restricted number of harmonics as discussed previously, and needs to improve in future work.

V. CONCLUSION

A procedure is presented for the detection and diagnosis of reciprocating compressors using RVMs-OAO classifiers and RVMs-OAO with GA-based feature selection from the envelope spectrum of vibration signals. The selection of input features and the appropriate classifier parameters have been optimised using a GA-based approach.

The characteristics of vibration signals, obtained under normal operation and operation with various defects have been investigated. The classification accuracy of RVMs with GA was shown to give higher classification rates than of RVMs without GA. (96.875%, 95.95) % respectively. The results show the potential application of GAs for feature selection. This also opens up the potential use of optimized features and classifier parameters for real-time implementation leading to possible development of an automated machine condition monitoring and diagnostic system.

The importance of just two features (4 and 7) in explaining fundamental behaviour in all cases is demonstrated. These two features having the capability to correctly classify binary faults in 100% of cases and (98.75% of combined fault cases were also correctly classified).

mRVM when applied to classification of 4 cases (healthy, IL, DVL+IL and SVL+DVL) using just two input parameters (harmonics 4 and 7) produced extremely high classification rates i.e. maintaining high predictive accuracy whilst retaining computational efficiency. However, on introducing further fault classes as for the RVM model, classification success dramatically decreased proportionate to the number of features incorporated.

Further work is to focus on relevant feature selection assessed by goodness of fit of the model thus ensuring reduced feature sets retain optimum powers of separation. The justification being that features would seem to have particular relevance to a given fault being used repeatedly in that faults presence. For example, feature 9 has a strong association with the DVL cases whilst the SVL relies heavily on features 8 and 10. The IL obviously emulates some of the characteristics of the other faults and in addition to the ‘base’ features (4 and 7) utilises the SVL set (8 and 10) along with features 3, 14 and 15.

REFERENCES

[1] B.-S. Yang, W.-W. Hwang, D.-J. Kim, and A. Chit Tan, "Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines," *Mechanical Systems and Signal Processing*, vol. 19, pp. 371-390, 2005.

[2] G. F. and B. A. D, "Use of the smoothed pseudo-Wigner–Ville distribution in the interpretation of monitored vibration data Maintenance," vol. 10, pp. 16–23, 1995.

[3] B. S. Yang, W. W. Hwang, M. H. Ko, and S. J. Lee, "Cavitation detection of butterfly valve using support vector machines," *Journal of Sound Vibration,* vol. 287, pp. 25–43, 2005.

[4] Q. Ren, X. Ma, and G. Miao, "Application of support vector machine in reciprocating compressor valve fault diagnosis," *Lecture Notes in Computer Science,* vol. 12, pp. 81–84, 2005.

[5] C. He, Y. Li, Y. Huang, C. Liu, and S. Fei, "Relevance vector machine based gear fault detection," in *Chinese Conference on Pattern Recognition*, Nanjing, 2009, pp. 1-5.

[6] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research,* vol. 1, pp. 211-244, 2001.

[7] Y. Hu and P. Luob, "Performance Data Prognostics Based on Relevance Vector Machine and Particle Filter," *CHEMICAL ENGINEERING,* vol. 33, pp. 349-354, 2013.

[8] D. G. Tzikas, L. Wei, A. Likas, Y. Yang, and N. P. Galatsanos, "A tutorial on relevance vector machines for regression and classification with applications," *EURASIP News Letter, 17, 2, 4,* vol. 23, 2006.

[9] D. Tzikas, A. Likas, and N. Galatsanos, "Large scale multikernel relevance vector machine for object detection," *International Journal on Artificial Intelligence Tools,* vol. 16, pp. 967-979, 2007.

[10] A. Tipping and A. Faul, "Analysis of sparse Bayesian learning," *Advances in neural information processing systems,* vol. 14, pp. 383-389, 2002.

[11] B.-S. Yang, T. Han, and W.-W. Hwang, "Fault diagnosis of rotating machinery based on multi-class support vector machines," *Journal of Mechanical Science and Technology,* vol. 19, pp. 846-859, 2005.

[12] V. Jain, G. Pillai, and I. Gupta, "Fault diagnosis in analog circuits using multiclass relevance vector machine," in *Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on*, 2011, pp. 641-643.

[13] T. Damoulas, Y. Ying, M. A. Girolami, and C. Campbell, "Inferring sparse kernel combinations and relevance vectors: an application to subcellular localization of proteins," in

Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on, 2008, pp. 577-582.

[14] I. Psorakis, T. Damoulas, and M. A. Girolami, "Multiclass relevance vector machines: sparsity and accuracy," *Neural Networks, IEEE Transactions on,* vol. 21, pp. 1588-1598, 2010.

[15] C. M. Bishop, *Pattern recognition and machine learning* vol. 1: springer New York, 2006.