Article citation info:
Al-Bugharbee H, Samaka H, Zubaidi SL. Diagnosis of air compressor condition using minimum redunduncy maximum relevance (MRMR) algorithim and distance metric based classification. Diagnostyka. 2021;22(4):25-32. https://doi.org/10.29354/diag/143762

DIAGNOSTYKA, 2021, Vol. 22, No. 4

ISSN 1641-6414
e-ISSN 2449-5220
DOI: 10.29354/diag/143762

DIAGNOSIS OF AIR COMPRESSOR CONDITION USING MINIMUM REDUNDUNCY MAXIMUM RELEVANCE (MRMR) ALGORITHIM AND DISTANCE METRIC BASED CLASSIFICATION

Hussein AL-BUGHARBEE [1], Hatam SAMAKA [1], Salah L ZUBAIDI [2]

Mechanical Engineering Department, Wasit University, Iraq. hrazzaq@uowasit.edu.iq, hsamaka@uowasit.edu.iq
Civil Engineering Department, Wasit University, Iraq, salahlafta@uowasit.edu.iq

Abstract

Finding a reliable machines condition monitoring technique has been attracted many researchers to avoid the sudden failure in machines and the unexpected consequences. This work proposes a fault diagnosis of air compressors using frequency-based features and distance metric-based classification. The analyzed experimental datasets contain one healthy condition and seven different fault conditions. Features are extracted from the frequency spectrum, then the best feature sets are selected using MRMR algorithm and eventually the classification is conducted using a distance metric classifier. The results demonstrated the automatic classification with more than 97% correct classification rate. The effect of selected feature set size, training sample size on the classification accuracy is also investigated. From the results, this method of analysis can be used for early detection of faults with very great accuracy.

Keywords: condition monitoring, Mahalanobis distance, signal processing

1. INTRODUCTION

Reciprocating compressors are commonly used in many industries such as chemical and petroleum ones. The condition of the compressor must be monitored to maintain a safe working environment and avoid economic losses because of the unplanned shutdown. Fault diagnosis of compressors still attracts the interest of many researchers who keep developing techniques that efficiently and automatically detect the compressor fault.

Various condition monitoring methods have been developed for fault diagnosis in reciprocating compressors. A simulation study of a two-stage reciprocating compressor has been conducted in [1]. Five various physical processes are involved in the mathematical modelling to represent the speed-torque characteristics of induction motor, crankshaft rotational movement, variation of cylinder pressure, vibration of the valves' plates and the flow characteristics through valves. In this study, the valve spring defect and valve leakage modelling are achieved. It was found that pressure variation and the crankshaft instantaneous angular speed are clearly sensitive for the presence of valve leakage and spring deterioration. The autocorrelation of spectrogram difference is also utilised for the detection of broken valve diagnosis under variable load conditions [2]. In Reference [3] the acoustic emission recordings of the valve motion are used for distinguishing the valve healthiness. The reference motion of normal valve is used for comparison. Two faults conditions, namely valve delayed closing and valve flutter and as well as the normal condition are considered in the analysis.

The authors of [4] compare the sensitivity of feature sets obtained from different signal transforms for fault detection. These feature sets are used for classification of three different compressor states namely, healthy, leakage inlet valve fault and leakage outlet fault. Generally, it was found the feature sets obtained from time-frequency transformation have better preformation in comparison to others.

A research was conducted in [5] to investigate the effectiveness of sensor position on the accuracy of fault detection in air compressor. More recently, the hybrid deep neural network is also incorporated for the purpose of fault diagnosis in air compressor [6]. Teager Kaiser energy operator is combined with deep belief classifier to identify the condition of the air compressor valve [7]. The researchers in [8] utilise the Local Mean Decomposition (LMD) in combination with Stack denoising Autoencoder for diagnosing faults in reciprocating compressors using vibration signals. Signal features are extracted using LMD and then classified using the SADE model. A forward neural network is also utilised for diagnosing of high-pressure compressor valve fault. Thirty-two various features are acquired from the compressor including voltages, pressure..etc. These most important features are then selected based on the scattering matrices [9]. In Reference [10] the performance of an optimised SVM in compressor fault diagnosis is compared to other SVM techniques. In that work, ninety-two features are extracted from the compressor pressure readings and

divided equally into a training and testing samples. In reference [11], the simulated motion of the suction and discharge valves as well as the measured AE signals are studied for the purpose of the fault diagnosis of valves in air compressors. It was found in the study that the valve fault can be identified as an increase in the AE amplitude.

The present paper has the following structure. The first section presents the state of art of techniques used for fault detection in reciprocating compressors. The second section briefed the proposed methodology including feature extraction, feature selection and classification. The third section explains the details of the experimental datasets used in the present paper. The fourth section presents and discusses the most important results obtained in the current work. The fifth section concludes the paper.

2. METHODOLOGY

This paper suggests a new methodology for fault detection of faults in air compressor. The methodology can be divided into three parts: features extraction, features selection and classification.

2.1 Feature extraction

The frequency representations of readings are obtained using Fourier transform and then the frequency spectrum are divided into a number of frequency bandwidths for every frequency spectrum. The mean value of every bandwidth's amplitude is then arranged and used to form the feature vector (f_v) as in equation 1.

$$\begin{bmatrix} fv_1 \\ fv_2 \\ fv_3 \\ \vdots \\ fv_m \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} & \cdots & f_{1n} \\ f_{21} & f_{22} & f_{23} & \cdots & f_{2n} \\ f_{31} & f_{32} & f_{33} & \cdots & f_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & f_{m3} & \cdots & f_{mn} \end{bmatrix} \quad (1)$$

2.2 Feature selection

The use of all extracted features is not always the right choice for a classifier input. It might negatively affect the accurateness of the classifier as well as causing additional computational burdens. The most effective and important features that provide best classification can be selected using the Minimum Redundancy Maximum Relevance (MRMR) algorithm. This algorithm sorting the extracted features according to their importance for the classification process. The feature importance of a given feature f_{ij} can be expressed according to the following equation 2[12, 13]

$$f^{MRMRM}(f_{ij}) = I(Y_i, f_{ij}) - \frac{1}{|S|}\sum_{F_S \in S} I(F_S, f_{ij}) \quad (2)$$

Where

f_{ij}	is the j^{th} feature for the i^{th}
I (.,.)	The mutual information.
S	The set of selected features.
$\|S\|$	The number of features in S.

The mutual information I (.,.) measures the "amount of information" obtained about one random variable through observing the other random variable [14].

2.3 Classification

The selected feature sets are divided into a training and testing sample. The training sample is used for building a reference space for every compressor condition while the testing sample is used to evaluate the accuracy of the classification model. The feature vectors of the testing sample are used as input to the Mahlanobis distance-based classifier. In this classifier, the distance between each testing feature vector to every training sample category is measured. Eventually, the feature vector can be assigned to the category which the measured distance is a minimum. The accuracy of the classification can be evaluated using the so-called confusion matrix. The matrix main diagonal represents the rates of the correct classification while the off diagonal elements represent the mistaken classification.

3. EXPERIMENTAL DATA SETS

The proposed methodology was applied to a single stage reciprocating air compressor acoustics datasets. These datasets were acquired from a compressor which has air pressure range 0-35 Kg/cm^3, driven by 5HP, 5 Am, 415V, 50 Hz and 1440 rpm induction motor. The obtained datasets cover 8 conditions of the air compressor (one healthy condition and seven fault conditions). These faults contain the Leakage Outlet Valve (LOV) fault, Non-Return Valve (NRV) fault, Leakage Inlet Valve (LIV) fault, Flywheel fault, Piston ring fault, Rider belt fault, and Bearing fault [12]. These datasets were acquired at 50kHz sampling rate using a microphone and an NIDAQ. The total number of the dataset was 1800 (i.e. 225 sample x 8 compressor conditions). The figure 1 shows the air compressor from which acoustic datasets were acquired.

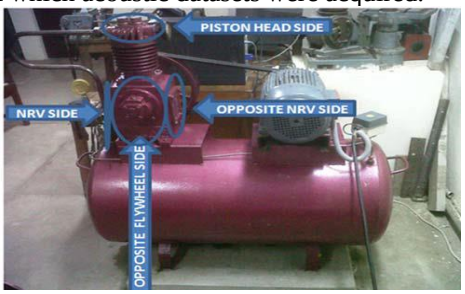
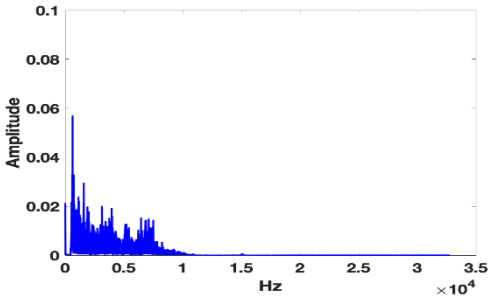


Fig. 1. The air compressor from which acoustic datasets were acquired [12]

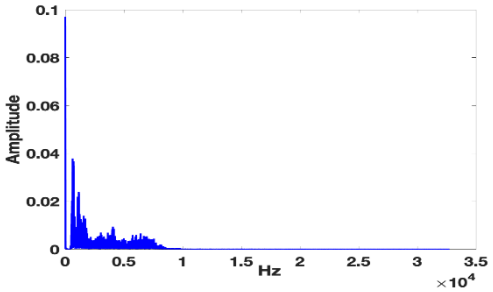
4. RESULTS AND DISCUSSION

4.1 Feature extraction

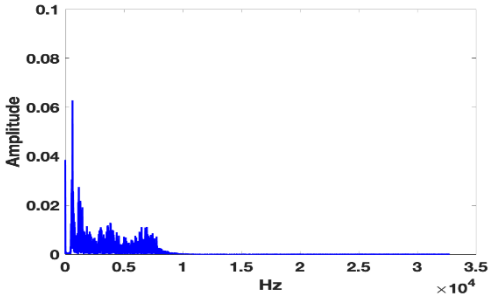
Figure 2 illustrates the frequency spectrum samples of all the compressor conditions considered in the current analysis.



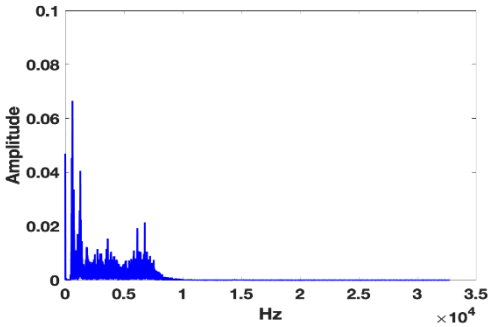
(a) Healthy



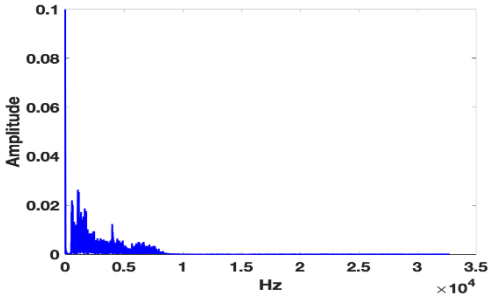
(b) Bearing fault



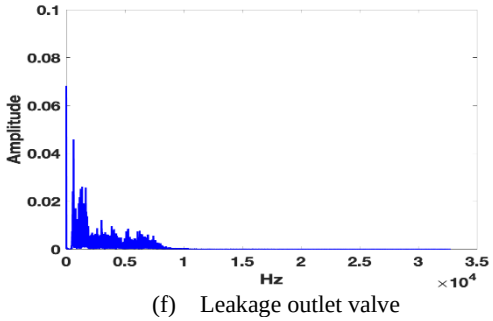
(c) Fly wheel fault



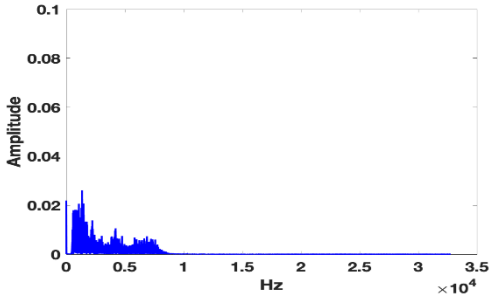
(d) Piston fault



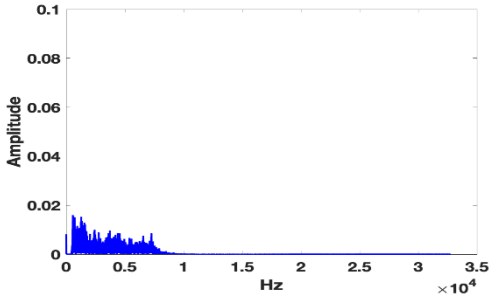
(e) Leakage inlet valve



(f) Leakage outlet valve



(g) Non-return valve fault



(h) Ride belt fault

Fig. 2. Spectrogram plot acoustic signals obtained from the compressors at different health conditions.

It can be easily seen that direct comparison of frequency spectrum is not always helpful neither practical. In addition, using the whole spectrum is not practical in classification of compressor condition. For this reason, features are formed by dividing the frequency spectrum into a number of bandwidths. Then the average amplitude value of every bandwidth is considered as a feature.

Figure (3) illustrates the average value of the frequency bands at 25 Hz respectively.

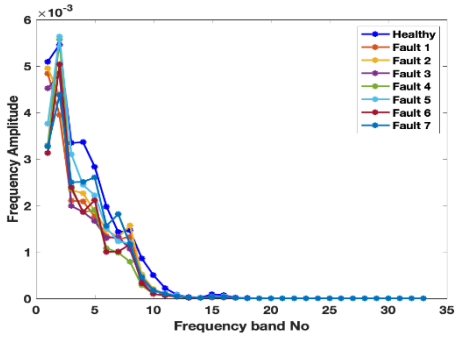


Fig. 3. The average value of frequency bands for the available conditions

4.2 Feature selection

The representation of these spectrums by a number of compact information still need for further processing in order to alleviate the computational burdens and to select only useful features for the purpose of compressor condition classification. As it was mentioned in section (II), the MRMR algorithm is used to rank the extracted features according to their importance in classification process. Figure 4 illustrates the rank of frequency bandwidths according to their importance score in the classification.

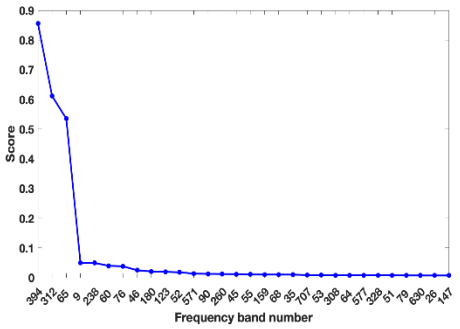


Fig. 4. of frequency bandwidths according to their importance score in the classification

Figure 5 visualizes the spatial distribution for the first three features of raw features and MRMR based selected features.

4.3 Classification

The automatic selection of features is important in reducing the burdens of forming sets of features randomly and it also helps in reducing the time consuming for calculations. The classification performance is investigated at different training sample length, different feature vectors and different frequency bandwidths.

For the detection process, the classifier is firstly investigated for the two-class healthy – faulty purpose. The table (1) illustrates the percentages of the correct classification rate for the testing sample in the detection process. These rates are calculated for two cases namely raw and MRMR based feature selected (i.e. bold and between brackets). The features used for the detection of faults are only the first three features. These results are also shown for different training sample size (i.e. 50, 100 and 125 features vector). The significant effect of the MRMR based feature selection is clearly seen in improving the correct classification rate percentage value for both bandwidths. For example, for the case where the frequency bandwidth is 25Hz, the accuracy of the detection rose to %97.7 at small training sample size (i.e. 50) while it raised to %99.4 at training sample size of 100. In addition, it rose to %99.6 for training sample size of 125.

For the fault diagnosis (i.e. identification of fault types), more longer feature vectors are required in comparison with the detection process as more information is required for distinguishing among different compressor condition.

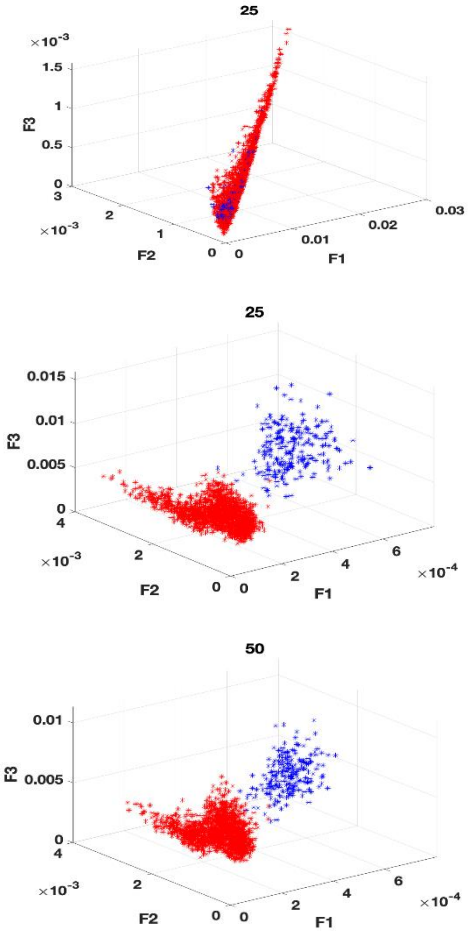


Fig. 5. Three diemnsional visuilisation for the first three features. Top (first three raw features), middle (first three selected features based on MRMR algorithm) at 25 Hz bandwith case, and bottom (first three selected features based on MRMR algorithm) at 50 Hz bandwith case

From the figure above, the improvement of separation between healthy and faulty categories are easily observed due to the selection algorithm.

Figure 6 shows a three-dimensional bar representation for the minimum correct classification rate percentage for the raw (left) and MRMR-Mahalanobis distance (right). As the figure 7 represents the minim value of the correct classification rate, this means the average value of the correct classification rates will be higher. It is also shown that increasing the training sample size as well as the feature vector lengths improve the correct classification rates.

The same discussion and findings can also be seen for the results obtained for the case of 50Hz bandwidth (see Figure (5) and Table (1).

Figure 8 illustrates the confusion matrix for the correct classification percentage rate obtained using feature vectors length of 15 and for a different training sample size. The improvement of classification accuracy is seen clearly when the off diagonal becomes less while the main diagonal of the matrix becomes higher. The same improvement can also be seen for the case where frequency band is 50Hz.

Table I. Percentages of correct (two classes) rate of testing sample. % Raw (MRMR)

Size of healthy training sample	**Raw (25Hz)**	**idx**	**Raw (50Hz)**	**idx**
50	% 45.6(**97.7**)	394 312 65	% 42.9(**98.4**)	197 156 33
100	% 40.2(**99.4**)	394 312 65	% 36.3 (**99.4**)	197 156 33
125	% 31.4 (**99.6**)	394 312 65	% 30.8 (**99.3**)	197 156 33

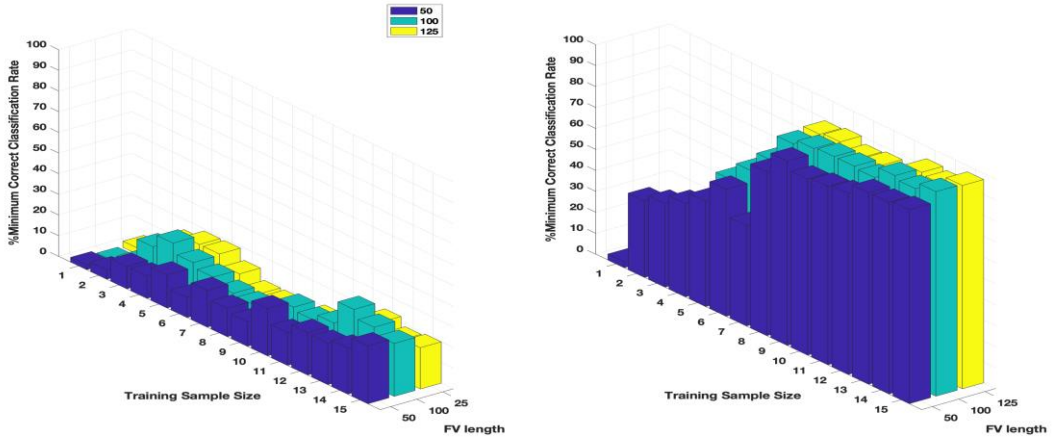


Fig. 6. The 3D bar representation of the minimum correct classification percentage rate. (left) raw features and (right) MRMR-based feature at 25Hz.

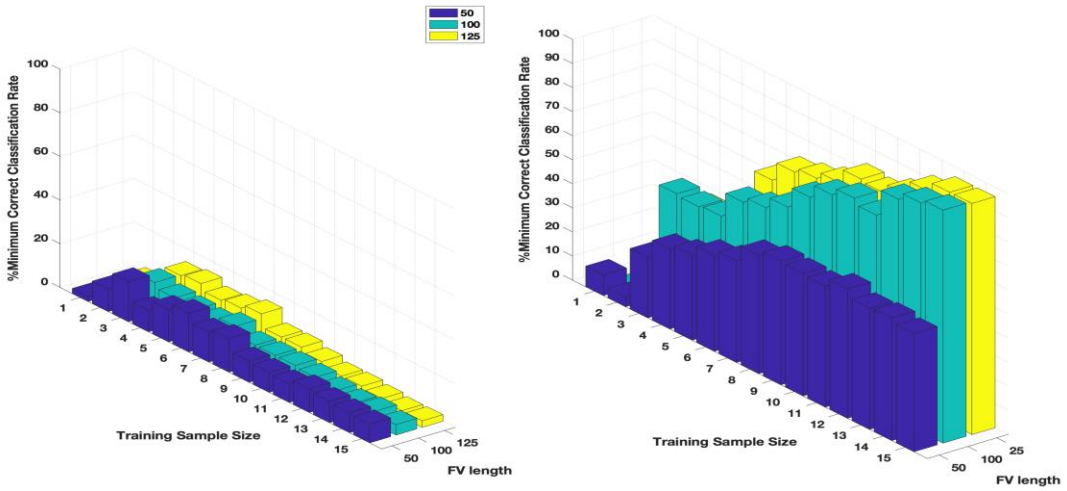
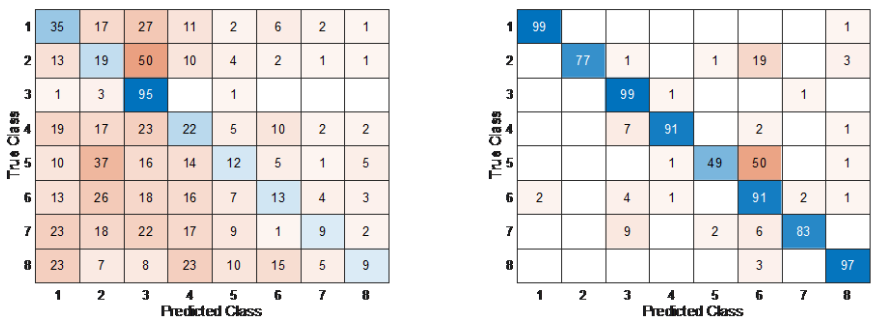
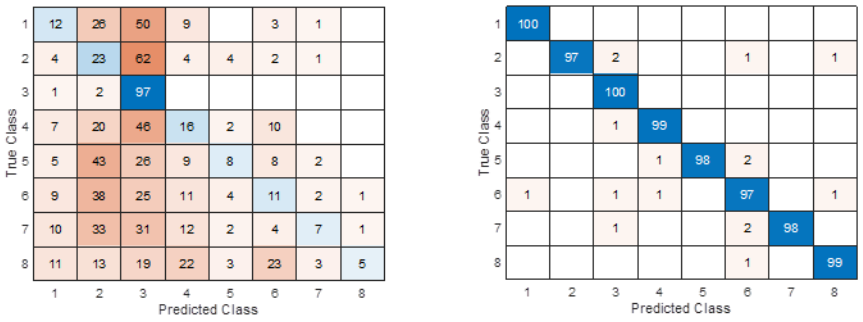


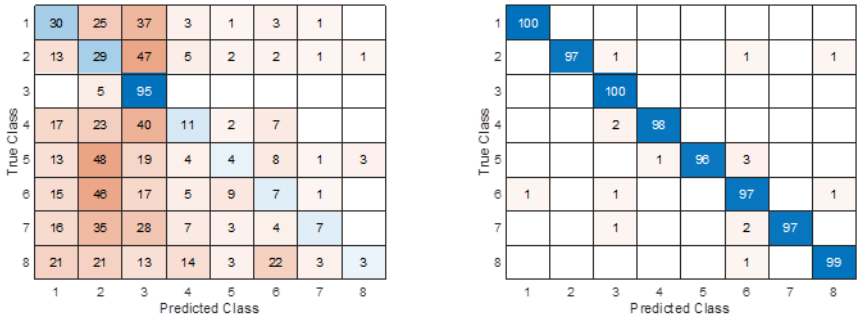
Fig. 7. The 3D bar representation of the minimum correct classification percentage rate. (left) raw features and (right) MRMR-based feature at 50Hz



(a) Training sample of 50 signals.

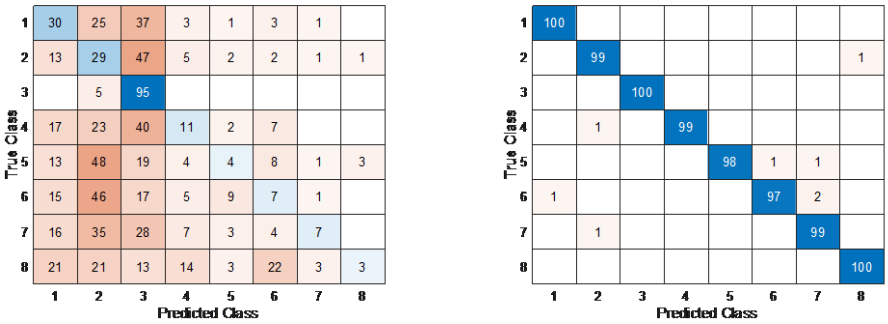


(b) Training sample of 100 signals.

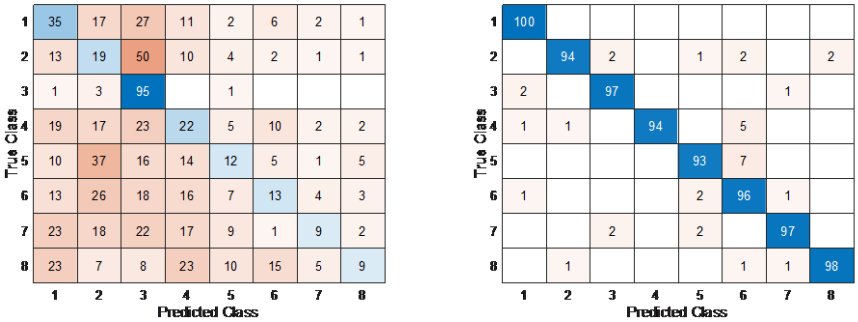


(c) Training sample of 125 signals.

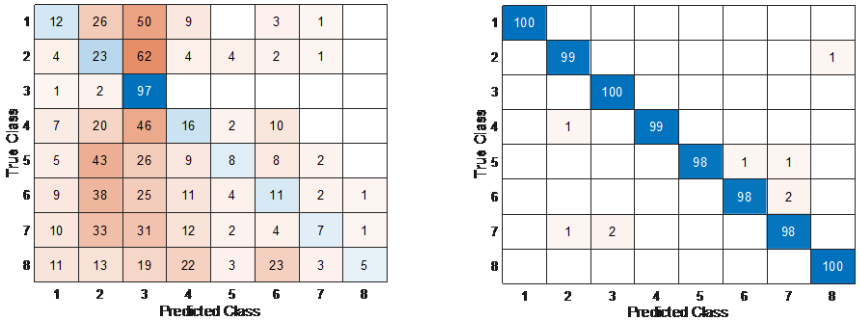
Fig. 8. Confusion matrix of compressor fault diagnosis at 25 Hz frequency band and different training sample size. Left: raw features, right: MRMR based features



(a) Training sample of 50 signals.



(b) Training sample of 100 signals.



(c) Training sample of 125 signals.

Fig. 9. Confusion matrix of compressor fault diagnosis at 50 Hz frequency band and different training sample size

5. CONCLUSION

In the present study, the diagnosis of fault presence and type in air compressor is investigated. The methodology contains linking the MRMR based feature selection and the distance metrics for the purpose of fault diagnosis. The features are extracted from the frequency spectrum of compressor acceleration vibration signals. The MRMR based selected features are then used as input to Mahalanobis distance classifier to identify the compressor health condition. The present methodology is evaluated at different training sample size and at different frequency bandwidth. The results show a very high correct recognition rate and reaches 99% in average.

REFERENCES

1. Elhaj M, Gu F, Ball A, Albarbar A, Al-Qattan M, Naid A. Numerical simulation and experimental study of a two-stage reciprocating compressor for condition monitoring. Mechanical Systems and Signal Processing. 2008;22(2):374-389. https://doi.org/10.1016/j.ymssp.2007.08.003
2. Pichler K, Lughofer E, Pichler M, Buchegger T, Klement EP, Huschenbett M. Fault detection in reciprocating compressor valves under varying load conditions. Mechanical Systems and Signal Processing. 2016; 70:104-119. https://doi.org/10.1016/j.ymssp.2015.09.005
3. Wang Y, Xue C, Jia X, Peng X. Fault diagnosis of reciprocating compressor valve with the method integrating acoustic emission signal and simulated valve motion. Mechanical systems and signal Processing. 2015;56:197-212. https://doi.org/10.1016/j.ymssp.2014.11.002
4. Verma NK, Gupta R, Sevakula RK, Salour A. Signal transforms for feature extraction from vibration signal for air compressor monitoring. TENCON 2014-2014 IEEE Region 10 Conference, 2014: 1-6.
5. Verma NK, Kadambari J, Abhijit B, Tanu S, alour A. Finding sensitive sensor positions under faulty condition of reciprocating air compressors. 2011 IEEE Recent Advances in Intelligent Computational Systems, 2011:242-246.
6. Tran VT, AlThobiani F, Tinga T, Ball A, Niu G. Single and combined fault diagnosis of reciprocating compressor valves using a hybrid deep belief network. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 2018;232(20):3767-3780. https://doi.org/10.1177/0954406217740929
7. Tabrizi AA, Al-Bugharbee H, Trendafilova I, Garibaldi L. A cointegration-based monitoring method for rolling bearings working in time-varying operational conditions. Meccanica 2017; 52(4-5): 1201-1217. https://doi.org/10.1007/s11012-016-0451-x
8. Liu Y, Duan L, Yuan Z, Wang N, Zhao J. An intelligent fault diagnosis method for reciprocating compressors based on LMD and SDAE. Sensors 2019; 19(5): 1041. https://doi.org/10.3390/s19051041
9. Li CJ, Yu X. High pressure air compressor valve fault diagnosis using feedforward neural networks. Mechanical Systems and Signal Processing. 1995; 9(5):527-536. https://doi.org/10.1006/mssp.1995.0040
10. Verma NK, Roy A, Salour A. An optimized fault diagnosis method for reciprocating air compressors based on SVM. IEEE International Conference on System Engineering and Technology, 2011: 65-69.
11. Wang Y, Gao A, Zheng S, Peng X. Experimental investigation of the fault diagnosis of typical faults in reciprocating compressor valves. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 2016; 230(13): 2285-2299. https://doi.org/10.1177/0954406215592921
12. Zhang X, Song Z, Li D, Zhang W, Zhao Z, Chen Y. Fault diagnosis for reducer via improved LMD and SVM-RFE-MRMR. Shock and Vibration 2018: 4526970. https://doi.org/10.1155/2018/4526970
13. Tang X, He Q, Gu X, Li C, Zhang H, Lu J. A Novel bearing fault diagnosis method based on GL-mRMR-SVM. Processes. 2020; 8(7): 784. https://doi.org/10.3390/pr8070784
14. Zubaidi SL, Ortega-Martorell S, Al-Bugharbee H, Olier I, Hashim KS, Gharghan SK, Kot P, Al-Khaddar R. Urban water demand prediction for a city that suffers from climate change and population growth: Gauteng province case study. Water 2020; 12(7): 1885. https://doi.org/10.3390/w12071885

Received 2021-08-17
Accepted 2021-11-06
Available online 2021-11-08

Hussein AL-BUGHARBEE was born in Iraq in 1980. HE received his BsC degree in mechanical engineering from University of Baghdad, Iraq, in 2002. Then, he received his M.sc in applied mechanics from University of Baghdad in 2006. In 2016, he received his PhD in Aerospace and mechanical engineering from Strathclyde University, UK. We has worked in the college of engineering since 2006 and his research focus on machinery condition monitoring and vibration analysis.

Hatam SAMAKA Was born in Iraq in 1965. He received his Bachelor's degree in Mechanical Engineering from University of Technology - Iraq, in 1987. Then he received MSc in welding processes from Al-Tahedy University – Libya in 2010 and PhD in car design field in 2015. Since 2016, he is a university teacher in Wasit University – Iraq. His research interests include design of frontal car for pedestrian protection, welding processes, the health effects of collisions on the head and leg of pedestrian with the frontal of car

Salah L. ZUBAIDI was born in Iraq in 1976. He received his BSc degree in Civil engineering from the University of Technology, Baghdad, in 1998. Then, he received his MSc degree in Environmental Engineering from the same university in 2012. His PhD degree in Environmental engineering was awarded from Liverpool John Moores University, UK in 2018. Since 2018, he is a lecturer with the Civil Engineering Department, Wasit University, Wasit. His research interests include water demand prediction, data mining, machine learning, and climate change.