Original Article

Institution of MECHANICAL ENGINEERS
Journal of MECHANICAL ENGINEERING SCIENCE

Proc IMechE Part C:
J Mechanical Engineering Science
0(0) 1–14

DOI: 10.1177/0954406217740929
journals.sagepub.com/home/pic
SAGE

Single and combined fault diagnosis of reciprocating compressor valves using a hybrid deep belief network

Van Tung Tran[1], Faisal AlThobiani[2], Tiedo Tinga[1], Andrew Ball[3] and Gang Niu[4]

Abstract

In this paper, a hybrid deep belief network is proposed to diagnose single and combined faults of suction and discharge valves in a reciprocating compressor. This hybrid integrates the deep belief network structured by multiple stacked restricted Boltzmann machines for pre-training and simplified fuzzy ARTMAP (SFAM) for fault classification. In the pre-training procedure, an algorithm for selecting local receptive fields is used to group the most similar features into the receptive fields of which top values are the units of each layer, and then restricted Boltzmann machine is applied to these units to construct a network. Unsupervised learning is also carried out for each restricted Boltzmann machine layer in this procedure to compute the network weights and biases. Finally, the network output is fed into SFAM to perform fault classification. In order to diagnose the valve faults, three signal types of vibration, pressure, and current are acquired from a two-stage reciprocating air compressor under different valve conditions such as suction leakages, discharge leakages, spring deterioration, and their combination. These signals are subsequently processed so that the useful fault information from the signals can be revealed; next, statistical features in the time and frequency domains are extracted from the signals and used as the inputs for hybrid deep belief network. Performance of hybrid deep belief network in fault classification is compared with that of the original deep belief network and the deep belief network combined with generalized discriminant analysis, where softmax regression is used as a classifier for the latter two models. The results indicate that hybrid deep belief network is more capable of improving the diagnosis accuracy and is feasible in industrial applications.

Keywords

Fault diagnosis, reciprocating compressor, hybrid deep belief networks, local receptive fields, simplified fuzzy ARTMAP

Date received: 1 June 2017; accepted: 13 October 2017

Introduction

Reciprocating compressors (RCs) are of vital importance in gas transmission pipelines, petrochemical plants, refineries, and general industry processes. Due to a high pressure ratio achievement, RCs are approximately installed three times higher in comparison with the other types such as centrifugal compressor in spite of its higher maintenance costs.[1] In an RC, several components operate in critical conditions such as high temperature and high pressure that lead to frequently failure occurrences. Components that mostly cause forced outages and are closely linked to the RC performance in terms of life and efficiency are suction and discharge valves. In fact, as indicated in the Leonard's[2] study, compressor valves can be the most serious single cause of unscheduled RC shutdowns (36%), and half of total repair cost is related to the valve faults. In some cases, a compressor valve failure can cause other mechanical-related problems. This emphasizes the need to develop accurate and reliable fault diagnostic methodologies for RC valves to minimize the shutdowns and the maintenance costs.

Several approaches using analytical or machine learning techniques in association with conventional signals, e.g. vibration, acoustic emission, and cylinder pressure, have been proposed in the literature to

[1]Faculty of Engineering Technology, Dynamics Based Maintenance, Applied Mechanics, University of Twente, The Netherlands
[2]Faculty of Maritime Studies–Marine Engineering, King Abdulaziz University, Kingdom of Saudi Arabia
[3]School of Computing and Engineering, University of Huddersfield, UK
[4]Institute of Rail Transit, Tongji University, China

Corresponding author:
Van Tung Tran, Faculty of Engineering Technology, Dynamics Based Maintenance, Applied Mechanics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
Email: vtung.tran@gmail.com

diagnose the RC and its valve faults. The principle of these approaches is based on applying advanced signal processing methods to deal with the transient phenomena and noise existing in the measured signals so that the fault patterns can be easily extracted. Then, data analyses or machine learning models such as artificial neural networks (ANNs) and support vector machines (SVMs) are employed to identify the fault occurrences. In case of analytical approaches, the analyses are carried out in various domains to compare the current valve state with reference states so that the fault type is highlighted. For instance, Pichler et al.[3] transformed vibration signals into high-dimensional vector space by using short-time Fourier transform and defined a metric in this space. Then, the distance between the actual compressor states and a reference state computed based on the defined metric was used to indicate the faults. Lin[4] used ensemble empirical mode decomposition to decompose the vibration signal into different frequency intrinsic mode functions and Hilbert spectrum to extract the different fault frequencies, respectively. Elhaj et al.[5] developed a mathematical model for a two-stage RC to simulate normal and different faulty conditions including valve leakage and valve spring deterioration. These faults were detected and identified by comparing the simulated and measured waveforms of cylinder pressure and crankshaft instantaneous angular speed fluctuation. Principal component analysis (PCA) was also applied for compressor fault diagnosis in the studies by Ahmed et al.[6] and Potočnik and Govekar.[7] These authors used PCA for dimensionality reduction of the statistical parameters extracted from the vibration data, and then, the faults were identified by T^2 and Q statistics in the study by Ahmed et al.[6] or by discriminant analysis in the study by Potočnik and Govekar.[7] Other studies using analytical approaches could be found in literature.[8–10] In general, fault diagnosis based on analytical approaches seems to be a challenge to practitioners when in-depth knowledge is required to accurately categorize the fault occurrences. Furthermore, they are inapplicable to automated decision-support systems.

In comparison to analysis-based fault diagnosis, machine learning-based fault diagnosis has been increasingly attracting attention from researchers. The main reason is that the fault diagnostic models can be flexibly generated from data without any prior knowledge or human experts about faulty characteristics of RC valves. Numerous outstanding research studies on valve fault diagnosis using machine learning techniques have been published in literature. Yang et al.[11] employed discrete wavelet transform (WT) with Daubeches-10 as a basic function to decompose the vibration and acoustic signals into different frequency sub-bands. Then, statistical measures were extracted from the first four levels, and ANNs together with SVMs were utilized as classifiers to identify the faults of small RCs used in refrigerators. Lin et al.[12] processed the vibration signals by various time–frequency analysis techniques to describe the dynamic characteristics of an RC, extracted the characteristics of the distribution map, and applied ANN for classification. Some researchers proposed alternative approaches to reveal the fault patterns in the measured signals instead of using signal processing techniques that could lead to costly computation or complexity. Cui et al.[13] extracted different information entropy features from vibration signals, e.g. singular spectrum entropy, spectrum entropy, probability density entropy, wavelet packet entropy, etc., and used these entropies as inputs of SVM classifier to diagnose the valve faults. In the study by Qin et al.,[14] SVM was also applied to recognize the valve fault types where its inputs were the parameterized waveform, decayed characteristics, and damped natural frequency extracted by the wave-matching method. Alternatively, a combination of machine learning models and PV diagram has received much consideration in the field of RC valve fault diagnosis due to the accurate reflection of the PV diagram on the valve conditions. Once any fault occurs, the shape of the PV diagram would change accordingly and the fault types can be recognized by learned models. Thus, instead of processing pressure signals, approaches based on the PV diagram have to deal with images to extract the useful fault patterns. For example, Feng et al.[15] used SVM and features extracted from digital images of the PV diagram by curvelet transform to classify five faulty types of an RC. Similarly, Wang et al.[16] combined image processing techniques to obtain invariant moments from the PV diagram with SVM in a framework to identify the faults of RC valves. Another study investigating the PV diagram for this research field can be found in the studies by Wang et al.[17] and Pichler et al.[18]

Although the combination of machine learning models with signal/image processing techniques provides potential and successful applications for fault diagnosis of RC valves, they still have deficiencies[19]: (a) the accurate results heavily depend on signal/image processing techniques, quality of features extracted from measured signals, and applied dimensionality reduction methods such as principle component analysis or independent component analysis; (b) machine learning models have shallow architecture where only one hidden layer is included. Such architectures limit the capacity of models to learn the complex nonlinear relationships in fault diagnosis. Recently, deep learning has gained much consideration in the machine learning area due to its ability to overcome the shortcomings of traditional learning models. One of the most attractive advantages of deep learning methods is that they can adaptively capture the representative information from raw data through multiple nonlinear transformations and approximate

complex nonlinear functions with a small error.[19–21] Different deep learning methods have currently been presented in the literature such as deep belief network (DBN),[22] deep Boltzmann machines (DBMs),[23] and deep convolutional neural networks.[24] They have been engaged in fault diagnosis of bearings,[20,25–27] industrial chemical processes,[28] RC and RC valves,[29,30] and rotors.[31] For further improving the accuracy of fault diagnosis, several approaches that hybridize the original deep learning with other methods have been proposed for text classification or speech recognition. However, a few approaches have been found in the fault diagnosis field. For instance, Li et al.[32,33] focused on spur and helical gearbox diagnosis. They either extracted the statistical parameters or separated the different modalities from the measured signals, and then DBM was employed to learn the pattern information in each parameter/modality. Finally, SVM or random forest was used to fuse the results.

To date, the hybrid deep learning approaches, especially hybrid deep belief network (HDBN), for fault diagnosis of RC valves have not been considered in the published literature. Furthermore, most of the fault diagnosis studies on RC in general and its valves in particular deal with single isolated faults. In real industrial operations, multi faults can occur at the same time causing the fault identification challenging due to the combined effect. Therefore, in this study, an HDBN based on local receptive fields,[34] DBN, and simplified fuzzy ARTMAP (SFAM)[35] are originally proposed not only to handle diagnosis of combined faults but also to enhance the diagnostic performance of RC valve faults. In addition, a comparative study of accuracy obtained from the HDBN, original DBN, and original DBN combined with generalized discriminant analysis (GDA),[36] for which softmax regression is used as a classifier for the latter two models, is carried out to appraise the improvement of HDBN.

Background

Deep belief networks

Architecture of RBMs. Restricted Boltzmann machine (RBM), whose architecture is shown in Figure 1, is a special type of Markov random field and is structured by one layer of binary stochastic hidden units $h \in \{0,1\}^H$ and one layer of binary stochastic visible units $v \in \{0,1\}^V$. RBM can be represented as a bipartite undirected graphical model where all visible units are connected to all hidden units, but there are no visible-visible or hidden-hidden connections. The weights between the visible layer and the hidden layer are undirected and are denoted by $\mathbf{w}$. The visible and hidden units have their biases represented by vectors $\mathbf{b}$ and $\mathbf{a}$, respectively.

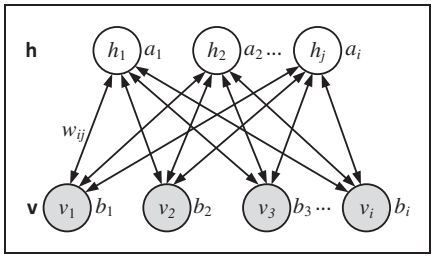

Figure 1. Architecture of the restricted Boltzmann machine.

In binary RBM, the weights of the connections and the biases of the individual units define a probability distribution over the joint states of the visible and hidden units via an energy function. This energy is given[37,38]

$$E(\mathbf{v},\mathbf{h};\theta) = -\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij}v_i h_j - \sum_{i=1}^{V} b_i v_i - \sum_{j=1}^{H} a_j h_j \tag{1}$$

where v_i and h_j are the binary states of visible unit i and hidden unit j;$\theta = \{\mathbf{w},\mathbf{b},\mathbf{a}\}$ is the model parameter; w_{ij} is the weight between visible unit i and hidden unit j; b_j and a_i are their biases; and V and H are the number of visible and hidden units. The joint distribution over the visible and hidden units is defined via the energy function as follows.

$$p(\mathbf{v},\mathbf{h};\theta) = \frac{1}{Z(\theta)} exp(-E(\mathbf{v},\mathbf{h};\theta)) \tag{2}$$

$$Z(\theta) = \sum_{\mathbf{v}}\sum_{\mathbf{h}} exp(-E(\mathbf{v},\mathbf{h};\theta)) \tag{3}$$

The probability that the model assigns to a visible vector $\mathbf{v}$ is given by summing over all possible hidden vectors

$$\begin{aligned} p(\mathbf{v};\theta) &= \sum_{\mathbf{h}} p(\mathbf{v},\mathbf{h};\theta) = \frac{1}{Z(\theta)}\sum_{\mathbf{h}} exp(-E(\mathbf{v},\mathbf{h};\theta)) \\ &= \frac{1}{Z(\theta)}\sum_{\mathbf{h}} exp\left(\mathbf{v}^T\mathbf{w}\mathbf{h} + \mathbf{b}^T\mathbf{v} + \mathbf{a}^T\mathbf{h}\right) \end{aligned} \tag{4}$$

Due to the special bipartite structure of the RBMs, the hidden units can be explicitly marginalized out

$$p(\mathbf{v};\theta) = \frac{1}{Z(\theta)} exp\left(\mathbf{b}^T\mathbf{v}\right)\prod_{j=1}^{H}\sum_{h_j\in\{0,1\}} exp\left(a_j h_j + \sum_{i=1}^{V} w_{ij}v_i h_j\right) \tag{5}$$

The conditional probability of hidden units $\mathbf{h}$ and visible vector $\mathbf{v}$ can be given

$$\begin{aligned} p(\mathbf{h}|\mathbf{v};\theta) &= \frac{p(\mathbf{v},\mathbf{h};\theta)}{p(\mathbf{v};\theta)} \\ &= \frac{\frac{1}{Z(\theta)} exp\left(\mathbf{b}^T\mathbf{v}\right)\prod_j exp\left(a_j h_j + \sum_{i=1}^{V} w_{ij}v_i h_j\right)}{\frac{1}{Z(\theta)} exp\left(\mathbf{b}^T\mathbf{v}\right)\prod_j \sum_h exp\left(a_j h_j + \sum_{i=1}^{V} w_{ij}v_i h_j\right)} \end{aligned}$$

$$= \prod_j \frac{exp\left(a_j h_j + \sum_{i=1}^{V} w_{ij} v_i h_j\right)}{\sum_h exp\left(a_j h_j + \sum_{i=1}^{V} w_{ij} v_i h_j\right)} = \prod_j p(h_j|\mathbf{v}) \tag{6}$$

It is similar for the probability assigning to a hidden vector $\mathbf{h}$ and the conditional probability of visible units $\mathbf{v}$

$$p(\mathbf{h};\theta) = \sum_{\mathbf{v}} p(\mathbf{v},\mathbf{h};\theta) = \frac{1}{Z(\theta)} \sum_{\mathbf{v}} exp(-E(\mathbf{v},\mathbf{h};\theta)) = \frac{1}{Z(\theta)} \sum_{\mathbf{v}} exp\left(\mathbf{v}^T \mathbf{w}\mathbf{h} + \mathbf{b}^T\mathbf{v} + \mathbf{a}^T\mathbf{h}\right) \tag{7}$$

$$p(\mathbf{v}|\mathbf{h};\theta) = \frac{p(\mathbf{v},\mathbf{h};\theta)}{p(\mathbf{h};\theta)} = \prod_i p(v_i|\mathbf{h}) \tag{8}$$

Because there are no hidden-hidden or visible-visible connections, the conditional probabilities $p(\mathbf{v}|\mathbf{h})$ and $p(\mathbf{h}|\mathbf{v})$ are factorial and are given by

$$p(h_j = 1|\mathbf{v}) = \sigma\left(a_j + \sum_i w_{ij} v_i\right) \tag{9}$$

$$p(v_i = 1|\mathbf{h}) = \sigma\left(b_i + \sum_j w_{ij} h_j\right) \tag{10}$$

where $\sigma = 1/(1 + exp(-x))$is the logistic function.

Gaussian–Bernoulli RBM. Originally, RBM was developed by using binary stochastic units for both the visible and hidden layers. However, such binary RBM is very inconvenient and only provides poor representation for modeling real-value data. To deal with this issue, Gaussian–Bernouli RBM (GBRBM) is considered in this study, and the visible units are replaced by linear units with independent Gaussian noise. The energy function of GBRBM becomes[38,39]

$$E(\mathbf{v},\mathbf{h};\theta) = \sum_{i=1}^{V} \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^{H} a_j b_j - \sum_{i=1}^{V}\sum_{j=1}^{H} \frac{v_i}{\sigma_i} h_j w_{ij} \tag{11}$$

where σ_i is standard deviation of visible unit v_i.

The probability function can be driven as follows.

$$p(\mathbf{v},\mathbf{h};\theta) = \frac{1}{Z(\theta)} exp(-E(\mathbf{v},\mathbf{h};\theta)) \tag{12}$$

$$Z(\theta) = \int \sum_{\mathbf{h}} exp(-E(\mathbf{v},\mathbf{h};\theta))\mathrm{d}v \tag{13}$$

The probability that the model assigns to a visible vector $\mathbf{v}$ in this case becomes

$$p(\mathbf{v};\theta) = \frac{1}{Z(\theta)} exp\left(\frac{(\mathbf{b}-\mathbf{v})^T(\mathbf{b}-\mathbf{v})}{2\sigma^2}\right) \times \prod_{j=1}^{H}\left(1 + exp\left(a_j + \sum_{i=1}^{V} w_{ij}\frac{v_i}{\sigma_i}\right)\right) \tag{14}$$

Also, the conditional probabilities are expressed.

$$p(v_i = x|\mathbf{h};\theta) = \frac{1}{\sigma_i\sqrt{2\pi}} exp\left(-\frac{\left(x - b_i - \sigma_i \sum_j h_j w_{ij}\right)^2}{2\sigma_i^2}\right) \tag{15}$$

$$p(h_j = 1|\mathbf{v};\theta) = g\left(a_j + \sum_i w_{ij}\frac{v_i}{\sigma_i}\right) \tag{16}$$

where g is the logistic function and x is a real number.

DBN architecture. DBN is a probabilistic generative model that is formed by stacking layer-by-layer a number of RBMs as shown in Figure 2. The learning scheme of DBN consists of pre-training and fine-tuning procedures. In the pre-training procedure, DBN is trained through bottom-up unsupervised fashion in a sequence of RBMs by CD algorithm:

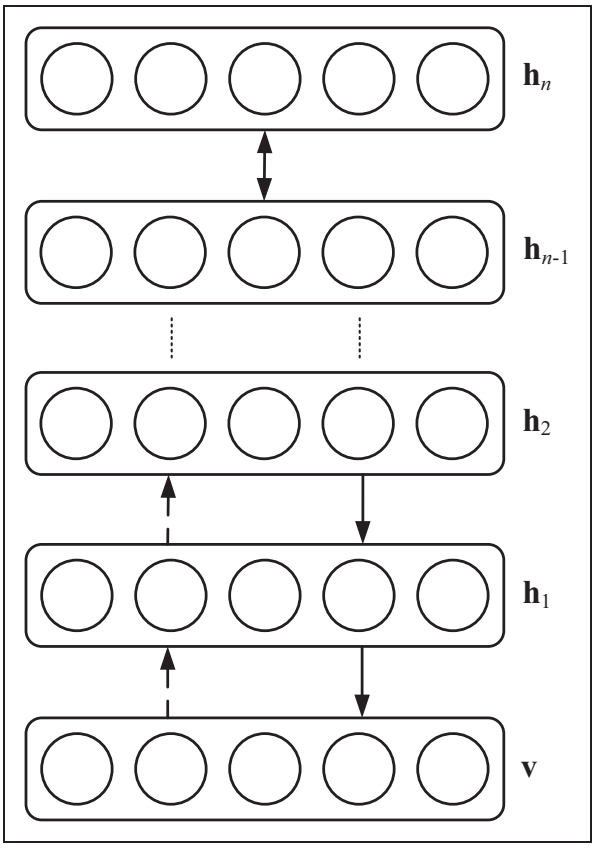

Figure 2. Architecture of deep belief network.

each RBM layer is trained by using the activation probabilities of the lower layer RBM as the input features and its output will be the input for the next RBM layer up. Note that the DBN architecture is structured by GBRBM at the first RBM layer and binary or Bernoulli-Bernoulli RBMs at the others for real-valued data input. The pre-training is followed by the fine-tuning procedure that is top-down supervised leaning using the back propagation algorithm. The classification task in this procedure is performed by softmax regression, which is a generalization of logistic regression model for multiple classes, at the top layer of the network. As a result, the back-propagation algorithm can adjust the network weights in the same way as the standard feed-forward neural network to improve the performance.

Let **h**, **w**, and N, respectively, be the activation of the penultimate layer units, the weight connecting the penultimate layer to the softmax layer, and the number of classes, the total input into a softmax layer **a** can be given as

$$a_i = \sum_k h_k w_{ki} \tag{17}$$

The discrete probability of each softmax layer unit is computed as

$$p_i = \frac{exp(a_i)}{\sum_j^N exp(a_j)} \text{ and } \sum_i^N p_i = 1 \tag{18}$$

The classified class $\hat{i}$ would be

$$\hat{i} = \underset{i}{argmax}\, p_i = \underset{i}{argmax}\, a_i \tag{19}$$

Local receptive fields

The drawback in such DBN architecture is how the hidden units in each layer connect to the ones in the layers beneath. To overcome this problem, Coates and Ng[34] proposed an algorithm for selecting local receptive fields (SLRF) to limit the fan-in of each unit by connecting its extractor to a small receptive field of inputs. Each receptive field is constructed by using a greedy selection scheme to group together the most similar low-level features to each other according to a pairwise similarity metric. This method has been successfully applied to image recognition[34] and financial distress prediction.[40]

As structured, each RBM in DBN contains a visible and a hidden layer. The units of the former are usually known and represent the lower level features, whilst the units of the latter represent features learned from the former. In order to build the units of the hidden layer, the SLRF algorithm groups the similar feature of the visible layer into a receptive field that is also a unit of the hidden layer. Let X denote the data set of feature vectors $x^{(i)},\ i \in \{1, \ldots, m\}$ with element $x_j^{(i)}$. If the data set X consists of linearly uncorrelated features, then a measure of the higher order dependence between the two features can be obtained by looking at the correlation of their energies (squared responses). In particular, if $E[X] = 0$ and $E[XX^T] = I$, then the similarity between the feature x_j and x_k is defined by squared responses[34]

$$S[x_j, x_k] = \text{corr}\left(x_j^2, x_k^2\right) = \frac{E\left[x_j^2 x_k^2 - 1\right]}{\sqrt{E\left[x_j^4 - 1\right] E\left[x_k^4 - 1\right]}} \tag{20}$$

The pairwise similarity between all of the features is given as

$$S_{j,k} \equiv S_X[x_j, x_k] \equiv \frac{\sum_i x_j^{(i)^2} x_k^{(i)^2} - 1}{\sqrt{\sum_i \left(x_j^{(i)^4} - 1\right) \sum_i \left(x_k^{(i)^4} - 1\right)}} \tag{21}$$

After obtaining the matrix of pairwise similarities between features $S_{j,k}$, the receptive fields are constructed by the greedy procedure: one feature is randomly chosen as a seed, then it is grouped with its nearest neighbors according to the similarity $S_{j,k}$. In detail, N rows of matrix S that corresponds to a random seed x_{jn} are randomly selected. Then, a receptive field is constructed that contains the features x_k corresponding to the top values of $S_{j_n k}$. This procedure is repeated for every seed.

SFAM network

SFAM is a simplified version of fuzzy ARTMAP[41] by reducing the complicated and redundant architectures, which is the main drawback of the original model for classification task. As a result, the SFAM is faster than fuzzy ARTMAP and easier to understand. The details of this network could be found in the study by Kasuba.[35]

HDBN architecture

The architecture of HDBN is similar to the original DBN where RBM is used at each layer. Despite having the same architecture, HDBN is different in constructing the network and the learning scheme, thanks to the SLRF algorithm: (a) the SLRF algorithm plays a role as feature selector in defining the hidden units for each layer; (b) Only the unsupervised learning (the pre-training procedure) of RBN is used and the fine-tuning procedure is entirely eliminated. This elimination enables not only reducing the training time of the fault diagnosis system but also free

usage of any classifier at the top of the HDBN to perform the classification task.

The process of building and training HDBN network is summarized as follows:

1. At the last bottom RBM layer, the inputted features or raw data are grouped into the receptive fields by SLRF. The receptive fields that have the highest values of correlation are collected to be the top candidates. These candidates present the units of the hidden layer.
2. RBM is built based on the input and the hidden units. Then, the RBM pre-training procedure is executed over these units to compute the weights and biases. The output of the current RBM is the input for the next RBM layer up.
3. The steps from 1 to 2 are iteratively carried out until the hidden units of the last RBM layer are reached.
4. SFAM classifier is connected to the last layer to perform the diagnosis results.

Proposed framework for fault diagnosis

The proposed framework for diagnosing single and combined valve faults of two-stage RC is shown in Figure 3. The vibration and pressure signals are acquired from the two stages, while the current signal is collected from a phase of a driving induction motor. For the vibration signals, the transient impact always exists due to the multitude of component moving in both rotary and linear directions. To handle these transient phenomena, the Teager-Kaiser energy operator (TKEO)[42,43] with DESA-1 algorithm[44] is used, thanks to its high time resolution and good adaptability to the instantaneous energy changes of signals. As a result, TKEO transforms the vibration signals into Teager-Kaiser energies (TKEs). For the pressure and current signals, the random noise contained in these signals are eliminated by the WT denoising method. Next, the statistical measures in the time domain and the frequency domain are extracted from the amplitude envelope of TKEs as well as the denoised pressure and current signals to represent the characteristics of the valve conditions.

In order to classify the faults of the RC valves, the extracted features are inputted into HDBN where the process of forming the DBN's structure, pre-training, and classification are carried out. As mentioned, the SLRF algorithm and the pre-training procedure are respectively executed to group the similar features layer by layer into receptive fields from the bottom to the top layer and train the network. Then, the output of the last hidden units is randomly split into a training set to build SFAM classifier and a test set to validate the accuracy of classification.

Valve fault experiments

To evaluate the proposed framework, our experiments were carried out on a Broom & Wade RC (model TS9) designed to compress air between 0.55 and 0.8 MPa to the receiver working at maximum pressure of 1.38 MPa. This compressor consists of two stages where the air pressed at the first stage (low pressure) is passed to the second stage (high pressure) via an air-cooled intercooler and then is stored in the air receiver. The crankshaft operated at a speed of 440 r/min and was driven through two V-belts by a three-phase induction motor running at speed of 1420 r/min. The layout of this RC including sensor locations, experimental system, and other specifications are respectively shown in Figure 4 and Table 1.

Different conditions of suction and discharge valves in the two stages working under different loads were used to generate the experimental data. One of these was a normal condition, whilst the others were suction leakages, discharge leakages,

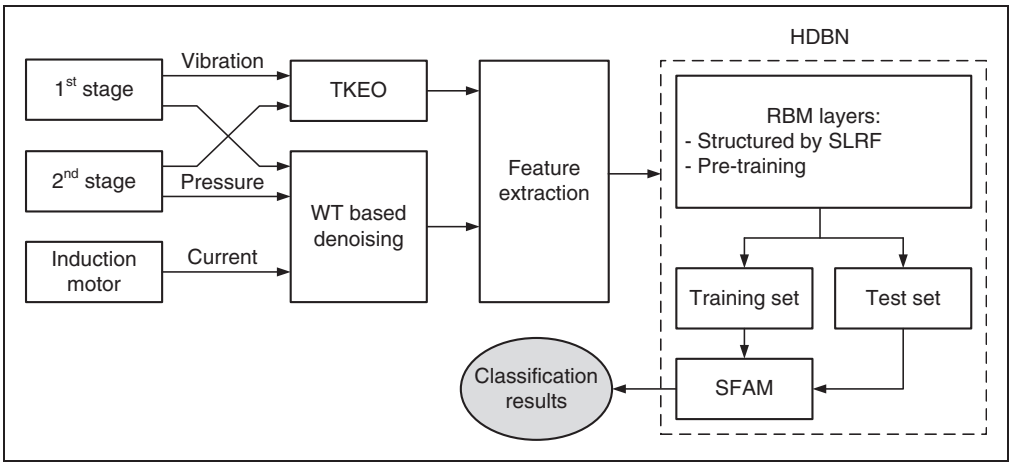

Figure 3. Proposed framework for reciprocating compressor valve fault diagnosis.
RBM: restricted Boltzmann machine; SFAM: simplified fuzzy ARTMAP; SLRF: selecting local receptive field; TKEO: Teager-Kaiser energy operator; WT: wavelet transform.

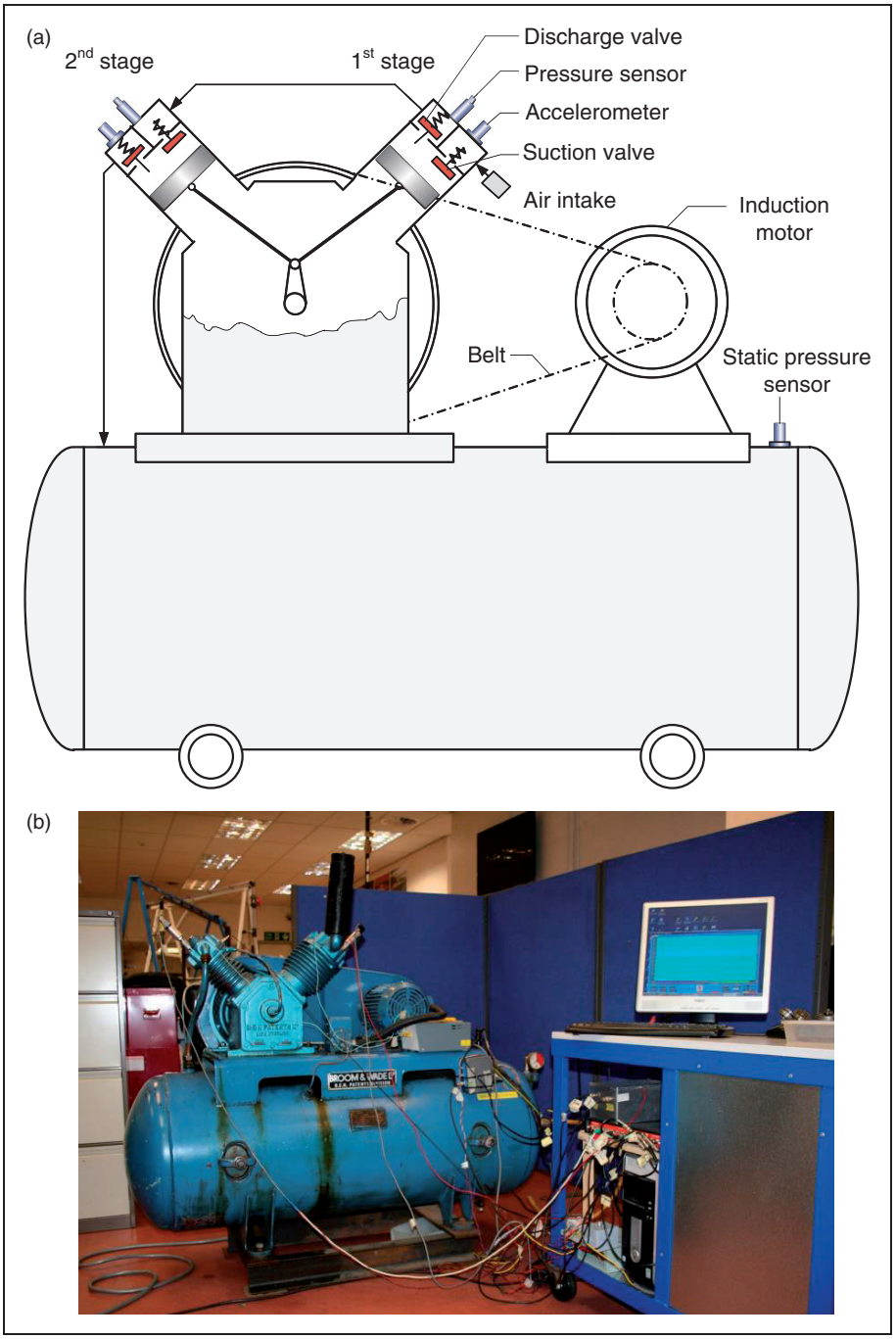

Figure 4. Reciprocating air compressor: (a) Layout and (b) Experimental system.

Table 1. The specifications of a two-stage reciprocating compressor.

Compressor		Induction motor	
Max working pressure	1.38 MPa	Power	2.5 kW
Number of cylinders	2	Voltage	380 V
Piston stroke	76.2 mm	Phase	3
Crank speed	440 r/min	Speed	1420 r/min
Tank description	Horizontal	Current	4.1 A

and spring deterioration. The valve leakages were artificially made by drilling a 1- and 3-mm diameter holes on the valve plates for small and large leakages, respectively. Totally apart from the normal condition, eight faulty conditions involving small discharge leakage in the first stage, small discharge leakage in the second stage, large discharge leakage in the second stage, discharge spring fault in the second stage, small suction leakage in the second stage, large suction leakage in the second stage, suction spring fault in the second stage, and a combined fault that is small

discharge leakage in the first stage and small suction leakage in the second stage were created.

To collect the vibration and pressure signals from the two stages, accelerometers and pressure sensors were vertically installed at the top of the stages. In detail, two accelerometers, model YD-5-2 with the frequency range 0–15 kHz and sensitivity 45 mv/ms^{-2}, were mounted on the valve coverings. The pressure sensors that were GEMS-22CSs with the range up to 4 MPa (600 psi) were installed in a thread hole made into the head of each cylinder. For the current signal acquisition, a Hall-effect-based current transducer having a frequency response from DC to 1.5 kHz was mounted on the phase A of the induction motor. Another static pressure sensor, SensorTechnics-PS2000 covering from 0 to 1.35 MPa (200 psi) was also installed to continuously monitor the pressure (load) in the air receiver during the experiments. This sensor was used not only to switch off the driving induction motor but also to trigger data collection at predefined pressures which began from 0.0689 MPa (10 psi) to 0.827 MPa (120 psi) with 0.0689 MPa (10 psi) intervals. Generally, the pressures at every moment were increasingly changed during the experiments. However, the data collection process was short enough that the increase of the pressure in the air receiver was insignificant in this period. Therefore, the pressures could be considered as constants at this acquisition. Once each of the predefined pressures were reached, data were acquired by a 16-bit resolution ADC system (model CED Power1401) with the sampling rate of 57,142 Hz. The software for the data acquisition was built based on LabWindows/CVI, National Instruments, that provides powerful function libraries and a comprehensive set of software tools for data acquisition, analysis, and presentation. The data set for each predetermined pressure consisted of four segments in which each segment contained 118,833 samples and covered approximately 15 working compressor cycles. Thus, 12 data sets according to 12 predefined pressure values in total were collected for each condition and then stored to the PC for analysis and classification.

Results and discussion

TKEO implementation and feature extraction

Due to transients, the features extracted from raw time-waveforms of the vibration signals or their fast fourier transform (FFT) spectra are not able to provide proper information for fault diagnosis. For instance, approximately a quarter-segment of the time-waveform vibrations at the second stage of the large discharge leakage (3 mm), the small discharge leakage (1 mm), the discharge spring fault, and the normal conditions at the load of 0.276 MPa are respectively shown in Figure 5(a). It can be seen that various transient impacts and noise by the motion of RC are observable. These time-waveforms further show that they seem to be indistinguishable.

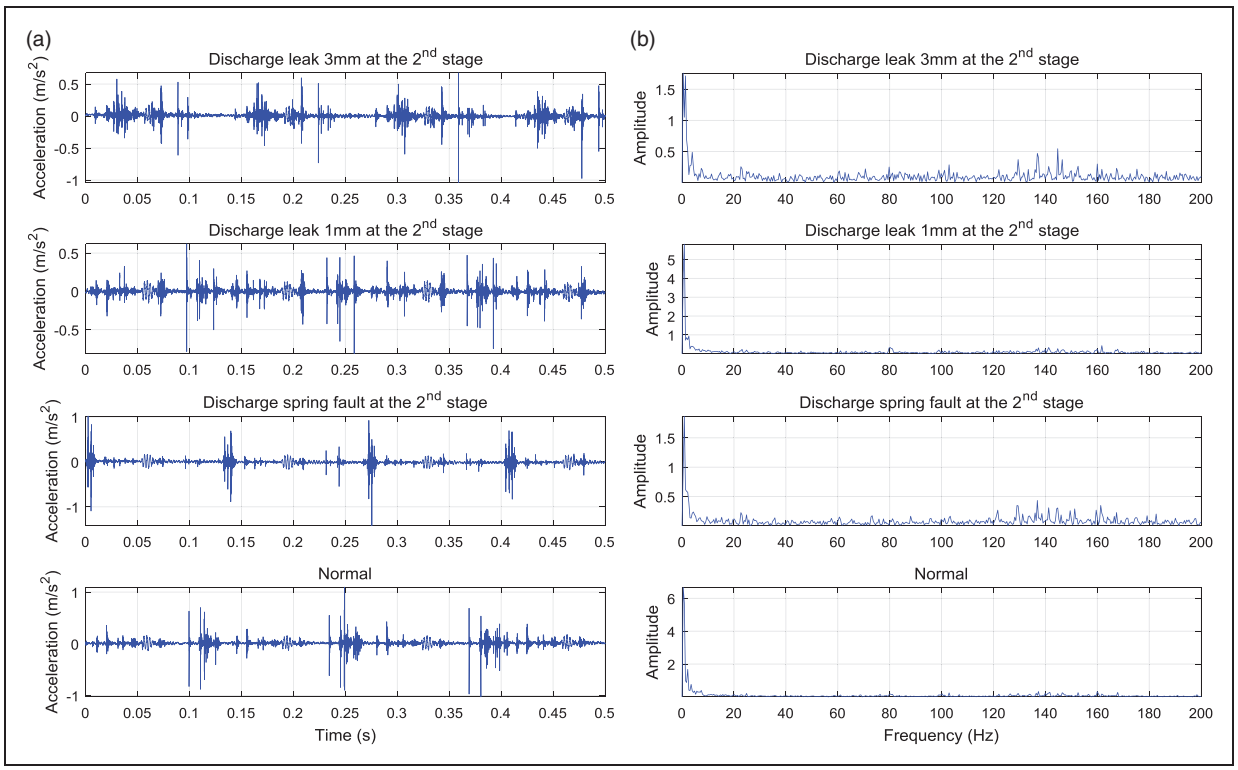

Figure 5. Vibration signals at the second stage of different conditions: (a) Time-waveform and (b) Frequency spectra.

When the load increases, more impacts and noise arise in the vibration signals leading them to be indistinctive. In the frequency spectra which are shown in Figure 5(b), the main frequency (7.3 Hz) cannot be revealed by FFT for all the vibration signals of the conditions. Therefore, TKEO is applied to transform the vibration signals into the TKEs. As observed in Figure 6(a) where the TKE time-waveforms of the same conditions above are depicted, the characteristic impacts are highlighted, and the noise is almost eliminated. Moreover, the main frequency 7.3 Hz and its harmonics are clearly discriminated in the frequency spectra, depicted in Figure 6(b). Obviously, TKEO can effectively uncover the characteristics of the vibration signal and remove the noise without using any filtering techniques. The advantages of TKEO application to the vibration signals can be realized as follows: (a) band-pass filter is unnecessary, so the appropriate estimation of the central frequency and the bandwidth of the band-pass filter is avoided; (b) the implementation of TKEO is quite simple and computationally efficient due to the fact that only two sample differences are used to compute in DESA-1 algorithm.

Some preprocessing steps are carried out for the experimental data before extracting the statistical features. As mentioned in the experimental section, for each predefined pressure level, the data contain four segments and each segment has 118,833 data points. The data set of each condition is partitioned into four subsets, thus 48 data samples (12 predefined pressures × 4) are obtained. Next, TKEO is applied to subsets of the vibration signal to convert them into TKEs. For the pressure and current signals, the random noise in the signals from the measurement process is removed by using the discrete WT-based denoising technique. The parameters for this denoising technique in this study consist of soft thresholding, Daubechies' least asymmetric as basic wavelet function (sym9), and three levels for the multiresolution decomposition. Lastly, the statistic features consisting of ten features in the time domain, three features in the frequency domain, and eight features of regression estimation are extracted from the amplitude envelopes of TKEs, pressure, and current signals, respectively. These features are correspondingly described from Tables 2 to 4. Totally, in the case of the vibration and pressure signals, we obtain a feature set containing 432 samples (9 valve conditions × 48 data samples) and 42 features (21 features × 2 compressor stages). In the case of current, the feature set is of 432 × 21.

Classification results and comparative studies

The feature sets obtained from the previous step are fed into the HDBN where its number of hidden layers is chosen as two. As introduced in the HDBN architecture section, the SLRF algorithm is firstly carried out to define the hidden units, and then the pre-training is applied to compute the weights and biases as well as the output of each layer. As a result, the number of hidden units of the two layers for the

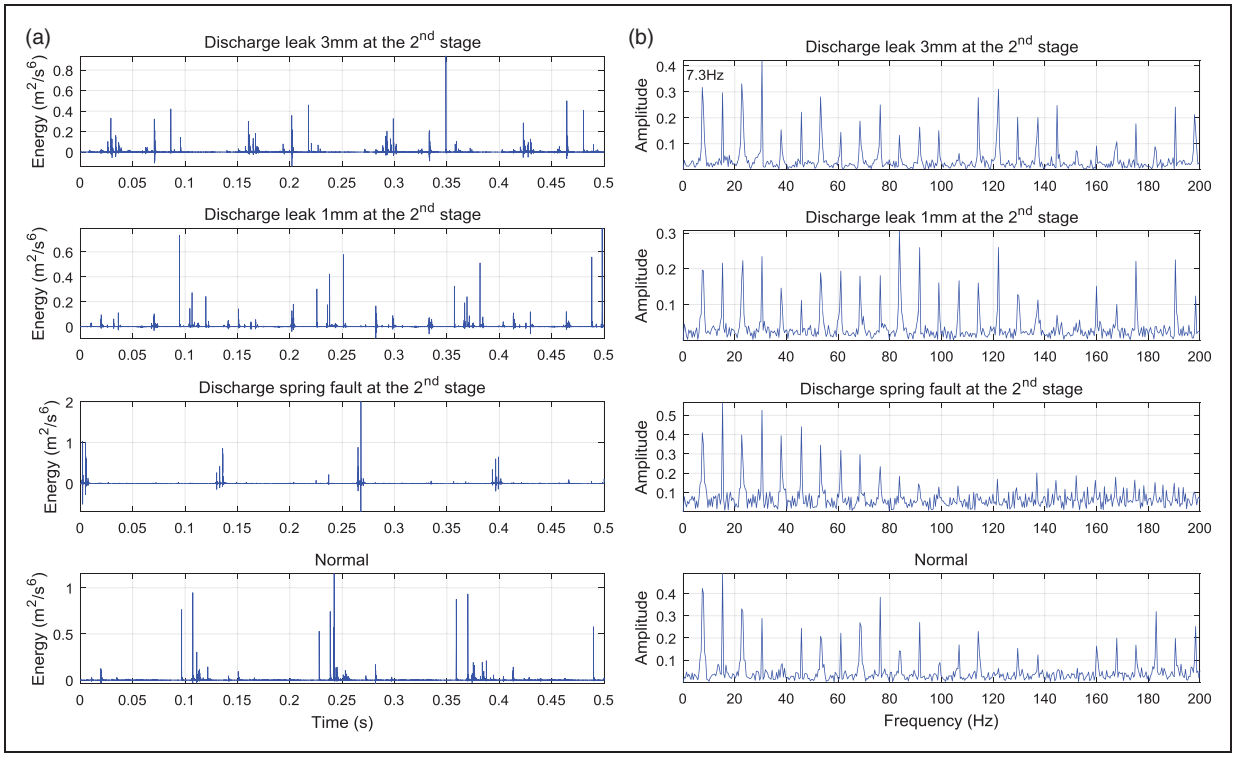

Figure 6. Teager-Kaiser energies of different conditions: (a) Time-waveform and (b) Frequency spectra.

Table 2. The statistical features in the time domain.

Features	Expression	Notes
Mean	$m = \frac{1}{N}\sum_{i=1}^{N} x_i$	x_i is the ith value of signal x, N is the number of data points
RMS	$x_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	
Shape factor	$SF = x_{rms}/\frac{1}{N}\sum_{i=1}^{N} \lvert x_i \rvert$	
Skewness	$Sk = \frac{1}{N}\sum_{i=1}^{N} (x_i - m)^3 \Big/ \left(\sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - m)^2}\right)^3$	
Kurtosis	$Kur = \frac{1}{N}\sum_{i=1}^{N} (x_i - m)^4 \Big/ \left(\frac{1}{N}\sum_{i=1}^{N} (x_i - m)^2\right)^2$	
Crest factor	$CF = \max(x)/x_{rms}$	
Entropy error	$E_e(x_i) = \sum p(x_i) \ln p(x_i)^2$	$p(x_i)$ is the distribution on the whole signal
Entropy estimation	$E_s(x_i) = -\sum p(x_i) \ln p(x_i)$	
Histogram lower	$h_L = \max(x_i) - \frac{\Delta}{2}$	$\Delta = \frac{\max(x_i) - \min(x_i)}{N-1}$
Histogram upper	$h_U = \max(x_i) + \frac{\Delta}{2}$	

RMS: root mean square.

Table 3. The statistical features in the frequency domain.

Features	Expression	Notes
Frequency center	$FC = \sum_{i=2}^{N} \dot{x}_i x_i / 2\pi \sum_{i=1}^{N} x_i^2$	$\dot{x} = (x_i - x_{i-1})/\Delta$
RMS variance frequency	$RMSF = \sqrt{\sum_{i=2}^{N} \dot{x}_i^2 / 4\pi^2 \sum_{i=1}^{N} x_i^2}$	
Root variance frequency	$RVF = \sqrt{\left(\sum_{i=2}^{N} \dot{x}_i^2 / 4\pi^2 \sum_{i=1}^{N} x_i^2\right) - FC^2}$	

RMS: root mean square.

Table 4. The statistical features of regression estimation.

Features	Expression	Notes
Auto-regression coefficients: a_1–a_8	$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \varepsilon_t$	y_t is the signal under investigation, n is the order of auto-regression model, and ε_t is Gaussian white noise.

features of the vibration, pressure, and current signals is obtained as shown in Table 5. As observed, the number of features is shrunk for each higher RBM layer due to the feature similarity grouped by SLRF.

Next, the output of the last hidden layer is partitioned into a training set and a test set by using the holdout validation method, where the ratio between these sets is chosen as 0.5. As a result, 216 samples are randomly selected for the test set and the rest for the training set. For the SFAM configuration, the learning rate β, conservative mode α, and number of epochs are used as 1, 0.001, and 4, respectively. Due to the randomized selection of samples for the training set and test set in the holdout validation method, the process of partitioning features, training, and evaluating the diagnostic model was repeated five times.

Table 5. The parameters of hybrid deep belief network.

	Values		
Parameters	Vibration	Pressure	Current
Number of hidden layers	2	2	2
Number of hidden units in layer 1	33	31	15
Number of hidden units in layer 2	27	24	11

In each replication, the training and testing process of SFAM were repeated five times to account for the stochastic nature, and then the results were averaged. Finally, the results of five-time validation were computed to show the accuracy of the models.

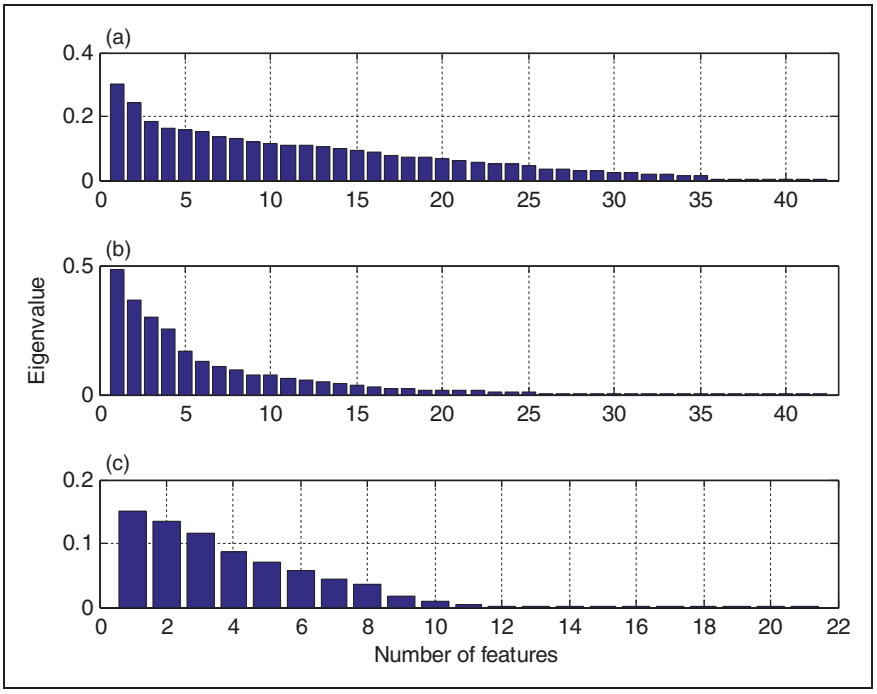

Figure 7. Eigenvalue of covariance matrix of feature sets: (a) vibration, (b) pressure, and (c) current.

Additionally, in order to appraise the improvement of the proposed framework, a comparative study of classification results are obtained from (a) this framework, (b) the framework of our previous study[30] that combines GDA for increasing the separation among the feature clusters and original DBN with softmax classifier for classification (GDA-DBN), (c) the framework that has only original DBN with softmax classifier to use DBN's ability in mining the fault information through pre-training. In case of GDA-DBN, the eigenvalue of the covariance matrix was used to estimate the sufficient amount of necessary features. As a result, the number of features in the feature sets retained after mapping into a new space was 35 for vibration, 25 for pressure, and 11 for the current signals as shown in Figure 7. Moreover, apart from the reused parameters of HDBN, e.g. the number of hidden layers and hidden units as shown in Table 5, other parameters for the fine-tuning procedure involving learning rate, number of training epochs, and mini-batch size are respectively 0.001, 100, and 9 to configure the network of GDA-DBN and original DBN.

The averaged classification results of five-time replications in the testing phase of DBN, GDA-DBN, and HDBN applied for the vibration, pressure, and current signals are shown in Table 6. Noting that the accuracy of HDBN given in this table is the maximum accuracy of which the vigilance parameter (VP) values vary from 0 to 0.9 with 0.1 intervals. As seen, HDBN provides better performance in comparison with GDA-DBN and significantly higher accuracy than the original DBN. In detail, HDBN achieves, respectively, 97.96%, 97.11%, and 91.18% classification accuracy for vibration, pressure, and current signals whilst GDA-DBN accuracy is slightly lower (97.07%, 95.52%, and 72.39%). The accuracy of the original DBN is the lowest, which is 85.41% for vibration signal, 82.09% for pressure signal, and 68.07% for current signal. Reasons for high performance of HDBN in comparison with the others are as follows: (a) SLRF and GDA are vastly superlative to DBN itself in clustering fault features among different valve conditions to increase the separation between them, which improves the classification accuracy; (b) The discrimination performed by SFAM used for HDBN is superior to the softmax regression used for GDA-DBN. The classification results presented in Table 6 also indicate that the current signal provides less accuracy due to the similarity of current waveforms of different valve conditions. This similarity leads to difficulty in extracting the fault characteristics from these waveforms. In other words, the current signal could insufficiently be sensitivity to highlight the differences among the valve conditions, especially in the case of multiple combined faults occurring in the machine.

Table 6. Averaged classification results (%).

Model	Vibration	Pressure	Current
DBN	85.41	82.09	68.07
GDA-DBN	97.07	95.52	72.39
HDBN	97.96	97.11	91.18

DBN: hybrid deep belief network; GDA: generalized discriminant analysis; HDBN: hybrid deep belief network.

Figure 8 shows the details of HDBN performance for the VP values from 0 to 0.9. As observed, the

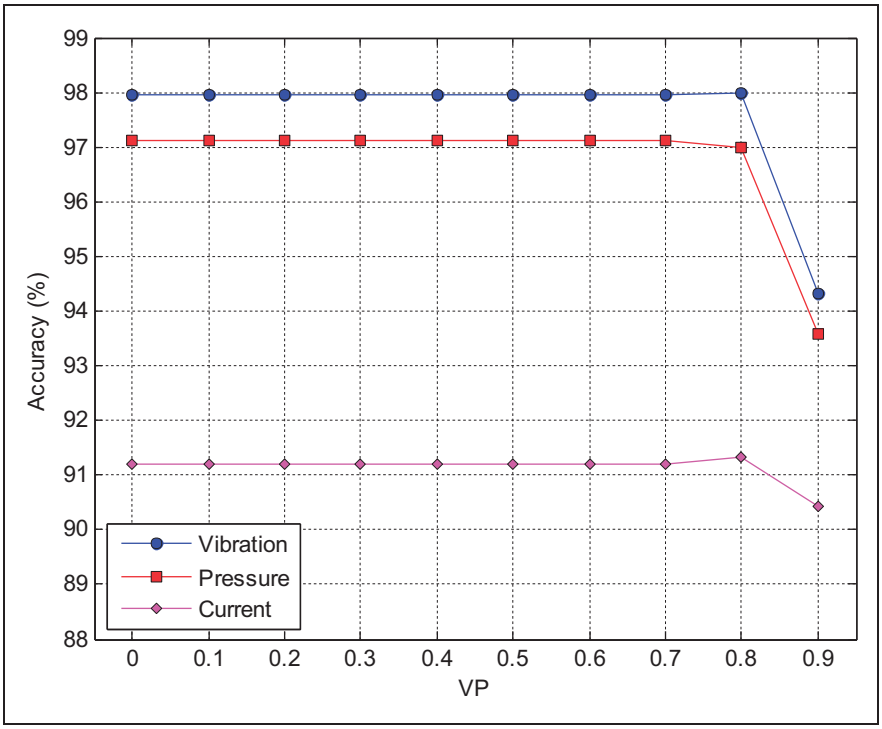

Figure 8. Classification results of hybrid deep belief network for varying vigilance parameters.

classification accuracy for the vibration signal stably reaches to 97.96% for the VP values ranging from 0 to 0.8. Similarly, the classification achieves 97.11% and 91.18% of accuracy for pressure and current signals, respectively. It is noted that the accuracy slightly decreases in case of pressure (96.98%) or less increases in case of vibration (97.98%) and current signal (91.3%) where the VP is of 0.8; however, these variations are not significant. At the VP of 0.9, all the accuracy is lessened. Obviously, the VP values do not strongly affect the HDBN performance and have a wide range from 0 to 0.8 for usage.

Summary and conclusions

This paper has presented HDBN in the fault diagnosis framework for identifying single and combined faults occurring in the suction and discharge valves of RC. This HDBN integrates DBN for pre-training and SFAM for classification. In the pre-training procedure, the SLRF algorithm is used to structure each RBM of DBN. Then, the pre-training procedure is carried out to compute the network parameters. Finally, the output of the last hidden layer is fed into the SFAM classifier without the fine-tuning procedure as for the original DBN. In order to validate the proposed approach, three signal types involving vibration, pressure, and current of experiments under eight single and one combined valve conditions are acquired. These signals are preprocessed by suitable methods and their statistical measures are extracted as the inputs for HDBN.

From the results, HDBN achieves high classification accuracy for the three signals (97.96% for vibration, 97.11% for pressure, and 91.18% for current). In comparison with the others that are original DBN and original DBN combined with GDA, HDBN has shown a remarkable improvement in accuracy. This confirms that HDBN in the proposed framework is eminently applicable for real fault diagnosis of RC valves.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is associated to the MaSeLMA project (Integrated Maintenance and Service Logistics Concepts for Maritime Assets), funded by Dinalog (Dutch Institute for Advanced Logistics).

References

1. Griffith WA and Flanagan EB. Online continuous monitoring of mechanical condition and performance for critical reciprocating compressors. In: *Proceeding of the 30th turbo-machinery symposium*, Houston, TX, Texas A&M University 2001.
2. Leonard SM. Increase reliability of reciprocating hydrogen compressors. *Hydrocarbon Process* 1996; 75: 67–74.
3. Pichler K, Schrems A, Buchegger T, et al. Fault detection in reciprocating compressor valves for steady-state load

conditions. In: *IEEE international symposium on signal processing and information technology*, Bilbao, Spain, 14–17 December 2011, pp.224–229. Piscataway, NJ: IEEE.
4. Lin J. Fault diagnosis of natural gas compressor based on EEMD and Hilbert marginal spectrum. In: *IEEE 2nd international conference on information science and engineering*, Hangzhou, China, 4–5 December 2010, pp.3701–3704. Piscataway, NJ: IEEE.
5. Elhaj M, Gu F, Ball A, et al. Numerical simulation and experimental study of a two-stage reciprocating compressor for condition monitoring. *Mech Syst Signal Process* 2008; 22: 374–389.
6. Ahmed M, Baqqar M, Gu F, et al. Fault detection and diagnosis using principal component analysis of vibration data from a reciprocating compressor. In: *Proceedings of the 2012 UKACC international conference on control*, Cardiff, UK, 3–5 September 2012, pp.461–466. IEEE.
7. Potočnik P and Govekar E. Vibration-based condition monitoring of compressors by principal component analysis and discriminant analysis. In: *The 23rd international conference on sound and vibration*, Athens, Greece, 10–14 July 2016.
8. Gu F, Shao Y, Hu N, et al. Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment. *Mech Syst Signal Process* 2011; 25: 360–372.
9. Wang Y, Xue C, Jia X, et al. Fault diagnosis of reciprocating compressor valve with the method integrating acoustic emission signal and simulated valve motion. *Mech Syst Signal Process* 2015; 56–57: 197–212.
10. Sim HY, Ramli R, Saifizul AA, et al. Empirical investigation of acoustic emission signals for valve failure identification by using statistical method. *Meas* 2014; 58: 165–174.
11. Yang BS, Hwang WW, Kim DJ, et al. Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines. *Mech Syst Signal Process* 2005; 19: 371–390.
12. Lin YH, Wu HC and Wu CY. Automated condition classification of a reciprocating compressor using time-frequency analysis and an artificial neural network. *Smart Mater Struct* 2006; 15: 1576–1584.
13. Cui H, Zhang L, Kang R, et al. Research on fault diagnosis for reciprocating compressor valve using information entropy and SVM method. *J Loss Prev Process Indust* 2009; 22: 864–867.
14. Qin Q, Jiang ZN, Feng K, et al. A novel scheme for fault detection of reciprocating compressor valves based on basis pursuit, wave matching and support vector machine. *Meas* 2012; 45: 897–908.
15. Feng K, Jiang Z, He W, et al. A recognition and novelty detection approach based on curvelet transform, nonlinear PCA and SVM with application to indicator diagram diagnosis. *Expert Syst Appl* 2011; 38: 12721–12729.
16. Wang F, Song L, Zhang L, et al. Fault diagnosis for reciprocating air compressor valve using P-V indicator diagram and SVM. In: *IEEE international symposium on information science and engineering*, Shanghai, China, 24–26 December 2010, pp.255–258. IEEE.
17. Wang GW, Zhuang J and Yu DH. Research and application of manifold learning to fault diagnosis of reciprocating compressor. In: *IEEE seventh international conference on fuzzy systems and knowledge discovery*, Yantai, China, 10–12 August 2010, pp.2652–2656. IEEE.
18. Pichler K, Lughofer E, Pichler M, et al. Detecting cracks in reciprocating compressor valves using pattern recognition in the pV diagram. *Pattern Anal Appl* 2015; 18: 461–472.
19. Jia F, Lei Y, Lin J, et al. Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech Syst Signal Process* 2016; 72–73: 303–315.
20. Shao H, Jiang H, Wang F, et al. An enhancement deep feature fusion method for rotating machinery fault diagnosis. *Knowl Base Syst* 2017; 119: 200–220.
21. LeCun Y, Bengio Y and Hinton GE. Deep learning. *Nature* 2015; 521: 436–444.
22. Hinton GE, Osindero S and Teh YW. A fast learning algorithm for deep belief nets. *Neural Comp* 2006; 18: 1527–1554.
23. Salakhutdinov R and Hinton GE. Deep Boltzmann machines. In: *Proceedings of the 12th international conference on artificial intelligence and statistics*, Clearwater Beach, Florida, USA, 2009, pp.448–455. PMLR.
24. Krizhevsky A, Sutskever I and Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, (ed F Pereira, CJC Burges, L Bottou and KQ Weinberger), 2012, pp.1106–1114. NIPS.
25. Chen Z, Zeng X, Li W, et al. Machine fault classification using deep belief network. In: *IEEE international instrumentation and measurement technology conference*, Taipei, Taiwan, 23–26 May 2016, pp.1–6. IEEE.
26. Shao H, Jiang H, Zhang X, et al. Rolling bearing fault diagnosis using an optimization deep belief network. *Meas Sci Technol* 2015; 26: 115002.
27. Jie T, Yi-Lun L, Da-Lian Y, et al. Fault diagnosis of rolling bearing using deep belief networks. In: *International symposium on material, energy and environmental engineering*, 2015, pp.566–569. Atlantis Press.
28. Lv F, Wen C, Bao Z, et al. Fault diagnosis based on deep learning. In: *IEEE American control conference*, Boston, MA, USA, 6–8 July 2016, pp.6851–6856. IEEE.
29. Keerqinhu, Qi G, Tsai WT, et al. Fault diagnosis for reciprocating compressors using big data. In: *IEEE second international conference on big data computing service applications*, Oxford, UK, 29 March–1 April 2016, pp.72–81. IEEE.
30. Tran VT, AlThobiani F and Ball A. An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks. *Expert Syst Appl* 2014; 41: 4113–4122.
31. Jeong H, Park S, Woo S, et al. Rotating machinery diagnostics using deep learning on orbit plot images. In: *44th North American manufacturing research conference*, Blacksburg, Virginia, USA, 2016, pp.1107–1118. Elsevier Procedia.
32. Li C, Sanchez RV, Zurita G, et al. Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing* 2015; 168: 119–127.

33. Li C, Sanchez RV, Zurita G, et al. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mech Systems Signal Process* 2016; 76–77: 283–293.
34. Coates A and Ng AY. Selecting receptive fields in deep neural network. In: *Advances in neural information processing systems*, (ed J Shawe-Taylor, RS Zemel, PL Bartlett, F Pereira and KQ Weinberger), 2011, pp.2528–2536. NIPS.
35. Kasuba T. Simplified fuzzy ARTMAP. *AI Expert* 1993; 8: 19–25.
36. Baudat G and Anouar F. Generalized discriminant analysis using a kernel approach. *Neural Comp* 2000; 12: 2385–2404.
37. Mohamed AR, Dahl G and Hinton GE. Acoustic modeling using deep belief networks. *IEEE Trans Audio Speech Lang Process* 2012; 20: 14–22.
38. Hinton GE. A practical guide to training restricted Boltzmann machines. Technical Report UTML TR 2010-003, Department of Computer Science, University of Toronto, 2010.
39. Senoussaoui M, Dehak N, Kenny P, et al. First attempt of Boltzmann machines for speaker verification. In: *Odyssey 2012-the speaker and language recognition workshop*, Singapore, 25-28 June 2012, pp.117–121. ISCA.
40. Lanbouri Z and Achchab S. A hybrid deep belief network approach for financial distress prediction. In: *IEEE 2015 10th international conference on intelligent systems: theories and applications*, Rabat, Morocco, 20–21 October 2015, pp.1–6. IEEE.
41. Carpenter CA, Grossberg S, Markuzon N, et al. Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans Neural Netw* 1992; 3: 698–713.
42. Kaiser JF. On a simple algorithm to calculate the 'energy' of a signal. In: *IEEE international conference on acoustics, speech, and signal processing*, 1990, pp.381–384. IEEE.
43. Kaiser JF. Some useful properties of Teager's energy operators. In: *IEEE international conference on acoustics, speech, and signal processing*, 1993, pp.149–152. IEEE.
44. Maragos P, Kaiser JF and Quatieri TF. Energy separation in signal modulations with application to speech analysis. *IEEE Trans Signal Process* 1993; 41: 3024–3051.