Reliability Engineering and System Safety 115 (2013) 124–135

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Failure diagnosis using deep belief learning based health state classification

Prasanna Tamilselvan, Pingfeng Wang*

Industrial and Manufacturing Engineering Department, Wichita State University, Wichita, KS 67208, USA

ARTICLE INFO

Article history:
Received 22 March 2012
Received in revised form
19 February 2013
Accepted 21 February 2013
Available online 14 March 2013

Keywords:
Fault diagnosis
Artificial intelligence in diagnosis
Classification
Deep belief networks

ABSTRACT

Effective health diagnosis provides multifarious benefits such as improved safety, improved reliability and reduced costs for operation and maintenance of complex engineered systems. This paper presents a novel multi-sensor health diagnosis method using deep belief network (DBN). DBN has recently become a popular approach in machine learning for its promised advantages such as fast inference and the ability to encode richer and higher order network structures. The DBN employs a hierarchical structure with multiple stacked restricted Boltzmann machines and works through a layer by layer successive learning process. The proposed multi-sensor health diagnosis methodology using DBN based state classification can be structured in three consecutive stages: first, defining health states and preprocessing sensory data for DBN training and testing; second, developing DBN based classification models for diagnosis of predefined health states; third, validating DBN classification models with testing sensory dataset. Health diagnosis using DBN based health state classification technique is compared with four existing diagnosis techniques. Benchmark classification problems and two engineering health diagnosis applications: aircraft engine health diagnosis and electric power transformer health diagnosis are employed to demonstrate the efficacy of the proposed approach.

1. Introduction

Effective health diagnosis provides multifarious benefits such as improved safety, improved reliability and reduced costs for operation and maintenance of complex engineered systems. Research on real-time diagnosis and prognosis which interprets data acquired by smart sensors and distributed sensor networks, and utilizes these data streams in making critical decisions advances significantly across a wide range of applications [1–3]. Maintenance and life-cycle management activities, which constitute a large portion of overhead costs in many industries, would greatly benefit from these advances [4–9]. In manufacturing and service sectors, unexpected breakdowns can be prohibitively expensive since they immediately result in lost production, failed shipping schedule, and poor customer satisfaction. In order to reduce and possibly eliminate such problems, it is necessary to accurately assess current stated of system degradation through effective health diagnosis. Two major research areas have tried to address these challenges: reliability and condition monitoring. Although reliability and condition monitoring are seemingly related, reliability focuses on population-wide characteristics while condition monitoring deals with component-specific properties. Reliability analysis is generally performed based on time-to-failure data [10–13] or physics-based models (e.g., fatigue, wear, and corrosion) [14,15]. In contrast, condition monitoring research uses sensory information from functioning systems to assess their degradation states. Continuous condition monitoring and real-time failure diagnosis using sensory data acquired from smart sensors not only notify performance degradation of system components at both early and advanced stages of damages, but enable failure prognosis to facilitate crucial decision-makings on system reliability and safety improvements [16–18]. A wide range of practical applications for condition monitoring and failure diagnosis have been reported in the literature and some of the successful ones include condition monitoring of bearings [19–21], machine tools [22], transformers [23], engines [24], and turbines [25]. Despite the success, effective diagnosis of current health state based on heterogeneous sensory data from multiple sensors is still an intricate problem and remains as a major challenge for condition monitoring techniques to be applied on complex engineered systems, mainly due to high system complexity and sensory data heterogeneity. Thus, one of the most important tasks in multi-sensor health diagnosis is to develop diagnostic approaches which can effectively handle multidimensional heterogeneous sensory signals and accurately classify different health states based on these acquired sensory signals.

Despite the challenges in system health diagnosis, there is another isolated group of research specific to pattern classification in image processing, which mainly specializes on classification processes. System health diagnosis with heterogeneous

Abbreviations: BNN, back-propagation neural network; DBN, deep belief networks; HS, health state; MD, Mahalanobis distance; PHM, prognostics and health management; RBM, restricted Boltzmann machine; SVM, support vector machine; SOM, self-organizing map

* Corresponding author. Tel.: +1 316 978 5910, fax: +1 316 978 3742.
E-mail addresses: pxtamilselvan@wichita.edu (P. Tamilselvan), pingfeng.wang@wichita.edu (P. Wang).

0951-8320/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2013.02.022

Nomenclature

$\mathbf{x}_i$	p-dimensional vector
c_i	ith class label
$\boldsymbol{\mu}_j$	mean vector of the training data
$\mathbf{S}_j$	variance matrix of the training data
w_{ij}	synaptic weight between the ith and the jth neurons
s_i	state of the ith neuron
$P(\cdot)$	probability distribution function
b_i	bias of the ith neuron
h_i	state of the ith neuron in hidden layer
v_i	state of the ith neuron in visible layer
m	momentum
n	epoch number

multidimensional sensory data is similar with the pattern classification problem with a high dimensionality of image data. In both cases, machine learning techniques have been dominant approaches and the learning complexity grows exponentially with the increase in heterogeneity and dimensionality of acquired sensory data [26]. In the past decade, pattern classification techniques have moved into a new platform of learning procedure called deep machine learning [26]. Analysis of the similarity between health diagnosis and pattern classification in these different applications motivates the emergence of a perfect collaboration in their learning techniques. There lies a great potential and also a critical need to utilize the advantages of deep machine learning techniques to address the challenges faced in system health diagnosis. However, advantages of evolving deep machine learning process have not been employed in current condition monitoring and health diagnosis research. Thus, this study proposes an efficient way to utilize the benefit of deep machine learning process to handle the complexity of sensory signals for the application of structural health diagnosis.

This paper presents a novel multi-sensor health diagnosis method using deep belief network (DBN). DBN has recently become a popular approach in machine learning for its promised advantages such as fast inference and the ability to encode richer and higher order network structures. The DBN works based on the restricted Boltzmann machine (RBM) and learns layer by layer of a deep network structure. The proposed diagnosis methodology can be structured in three consecutive stages: first, defining health states and preprocessing sensory data for DBN training and testing; second, developing DBN based classification models for the diagnosis of predefined health states; third, validating DBN classification models with testing sensory dataset. Health diagnosis using DBN based health state classification technique is compared with four existing diagnosis techniques: SVM, BNN, SOM, and MD classifier. Benchmark classification problems and two engineering health diagnosis applications: aircraft engine health diagnosis and electric power transformer health diagnosis are employed to demonstrate the efficacy of the proposed approach. The rest of the paper is organized as follows: Section 2 presents the related work of health diagnosis with existing state of the art classification techniques; Section 3 details the proposed health diagnosis approach with DBN based classification; Section 4 demonstrates the developed diagnosis approach with case studies; and Section 5 summarizes the presented research and the future work.

2. Related work

Due to the complexity of system degradation characteristics and potential heterogeneity of sensory signals, multi-sensor health diagnosis of complex engineered systems remains as an intricate problem. Consequently, machine learning techniques and statistical inference techniques are often employed to solve this problem. Significant advances have been achieved in applying classification techniques based on machine learning [27–34] or statistical inferences [35–37], which resulted in a number of state-of-the-art health state classification methods, such as back-propagation neural network (BNN) [27–30], self-organizing maps (SOM) [31], support vector machine (SVM) [28,32–34], and Mahalanobis distance (MD) [32,35]. This section surveys the current literature and briefly discusses the working principle and capabilities of the different existing classification techniques.

In general, machine learning based diagnosis techniques can be broadly classified into supervised and unsupervised learning approaches. This paper focuses on the supervised learning process in diagnostic classification. The supervised learning is the process of learning a relationship between input values and desired target labels in the form of set of patterns. The error values are evaluated and given as feedback to the learning model to get potential optimal solutions. The learnt relationship/function from training dataset is used as a classifier model to predict unlearnt dataset with unknown patterns. Unlike supervised learning process, unsupervised learning process does not utilize labeled training data. The BNN and the SOM are two representative artificial neural network type diagnosis techniques that are based on supervised learning and unsupervised learning respectively. The BNN possesses a basic neural network structure with three different types of layers: the input layer, the output layer and the hidden layers [27], and is generally trained through optimizing synaptic weights and biases of all neurons till a desired classification rate is obtained. Using BNN, the health diagnosis problem is solved as a health state prediction problem using trained neural networks. Different with the BNN, the SOM is an unsupervised learning technique working based on neurons that determines a closest best-matching unit distance to the input vector [31], and uses it to construct class boundaries graphically on a two dimensional map. The BNN and the SOM have been used as state classification methods for different health diagnosis applications [27]. The main drawback of the BNN and SOM is the over-fitting of the data to the classification model and leading to significant error values in complex scenarios.

The SVM is a state-of-the-art technique for multi-dimensional classification based on supervised learning. The SVM organizes input data D into two sets of vectors in a p-dimensional space as

$$D=\{(\mathrm{x}_i,c_i)\,|\,\mathrm{x}_i\in R^p,\quad c_i\in\{0,1,2\},\ \ i=1,2,\ldots,p\},\tag{1}$$

where c_i is the ith class label (e.g., 0, 1, or 2) indicating the class to which data point $\mathbf{x}_i$ belongs. Each $\mathbf{x}_i$ is a p-dimensional real vector, shows the preprocessed p-dimensional sensory data. With the organized input data, the SVM constructs hyper-planes with maximum margins to divide data points with different c_i values [33]. A hyper-plane can be written as a set of points $\mathbf{x}$ satisfying

$$\mathbf{w}\cdot\mathbf{x}-b=0,\tag{2}$$

where vector $\mathbf{w}$ is a normal vector that is perpendicular to the hyper-plane. The parameter $b/\|\mathbf{w}\|$ determines the offset of the hyper-plane from the origin along the normal vector $\mathbf{w}$. We want to choose the $\mathbf{w}$ and b to maximize the margin, or distance between the parallel hyper-planes of the margin. The optimization problem is defined as

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{s.t.} \qquad c_i(\mathbf{w}\cdot\mathbf{x}_i - b) \ge 1. \end{aligned} \tag{3}$$

Solving the optimization problem above will eventually provide a set of optimized $\mathbf{w}$ and b that define the classification margins [33]. The optimal hyper-plane algorithm for non-linear input space can be performed by employing kernel functions, through which the input vector $\mathbf{x}$ is transformed into a high dimensional feature space and maximum-margin hyper-planes in a transformed feature space classify different classes for the data from the non-linear input space. With the nonlinear kernel functions, the complexity of the classification problem depends only on the dimensionality of the input space instead of the high dimensional feature space.

The above described methods, the BNN, the SOM and the SVM, are different machine learning techniques for health state classification. Unlike these methods, the MD classifier is a classification technique based on statistical inference using the Mahalanobis distance measure. The MD measure shows the degree of deviation of the measured data point $\mathbf{x}_j$ from a reference training set, which can be calculated as

$$D(\mathbf{x}_j) = \sqrt{(\mathbf{x}_j - \boldsymbol{\mu}_j)^T \mathbf{S}_j^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_j)}, \tag{4}$$

where $\mathbf{x}_j = (x_1, x_2, \ldots, x_n)$ is a multi-dimensional data vector, $\boldsymbol{\mu}$ and $\mathbf{S}$ are the mean vector and variance matrix of the reference training data set. Wang et al. [35] classified different health states using this statistical measure and the testing dataset was classified into one health state with a minimum MD measure. The advantages and potentiality of different algorithms can be combined to overcome some of the drawbacks of the existing algorithms. Some researchers combined two or more existing machine learning techniques and formed hybrid models from different individual algorithms. Zhang [38] proposed a bearing fault diagnosis methodology using multi-scale entropy (MSE) and adaptive neuro-fuzzy inference system. Saimurugan et al. [39] presented a multi-component fault diagnosis of rotational mechanical system based on decision tree and support vector machine.

Despite the demonstrated applicability of the above mentioned classification methods, continuous health monitoring through multi-sensor health diagnosis remains as one of the challenge problems to be addressed in the field of state classification and health diagnosis [40]. In general, the complexity of a classification model increases with increase in number of sensors and heterogeneity of sensory signals. There is another set of research in the field of classification techniques apart from health diagnosis, which is pattern classification in the image processing field. The problems faced in the diagnostic classification problem with multi-sensors are similar to the pattern classification problems with the high dimensionality of image data. In the last decade, pattern classification has advanced into a new paradigm with emerging techniques of deep machine learning. The requirement of handling multidimensional heterogeneous sensory signals in the diagnostic classification and advantages of deep machine learning techniques to handle the high dimensionality of data motivated the emergence of a perfect collaboration in their learning techniques. However, the advantages of evolving deep machine learning techniques have not been fully utilized for failure diagnosis. Thus, this paper presents a novel multi-sensor health diagnosis method based on DBN, which utilizes the benefit of deep machine learning process to handle the complexity of sensory signals for failure diagnosis applications. DBN is a recent offspring of supervised machine learning with deep learning capability, fast inference, fast unsupervised learning and ability to encode richer and higher order network structures [41]. Multi-layered deep belief networks enhance the classification capability for complex diagnosis problems. The following section details the proposed health diagnosis approach using DBN based state classification.

3. System health diagnosis using DBN

This section details the proposed multi-sensor health diagnosis approach using DBN based state classification. Section 3.1 overviews the DBN architecture and the methodology involved in DBN machine learning. Section 3.2 presents the learning function and the validation process of DBN for health state classification. Section 3.3 presents the overall procedure of the proposed diagnosis approach.

3.1. Deep belief network architecture

This section introduces the DBN architecture and the general methodology involved in DBN machine learning. DBN employs a multilayered architecture which consists of one visible layer and multiple hidden layers as shown in Fig. 1. The visible layer of a DBN accepts the input data and transfers the data to the hidden layers in order to complete the learning process [41].

The DBN structure is similar to the stacked network of the restricted Boltzmann machine (RBM) [42]. Each RBM consists of two layers, namely visible layer and hidden layer. As noted by the name, the connections between the nodes within each RBM layer (visible layer or hidden layer) are restricted. The process of transformation of data from visible layer to the hidden layer is finished through a sigmoid activation function based on the RBM learning rule [42]. An example of DBN structure is shown in Fig. 1, which consists of three stacked RBMs, as layer 1 (visible layer) and layer 2 (hidden layer 1) forms the first RBM 1, layer 2 (hidden layer 1) and layer 3 (hidden layer 2) forms the second RBM, and layer 3 (hidden layer 2) and layer 4 (hidden layer 3) forms the third RBM.

Each successive layer in the DBN structure follows the same transformation concept and passes the regularity throughout the DBN architecture [43]. The overall learning process of the DBN classifier model is described by the flowchart shown in Fig. 2. As shown in the figure, the model inputs include preprocessed batch data, total number of layers in the DBN classifier model, total number of hidden neurons in each hidden layer and

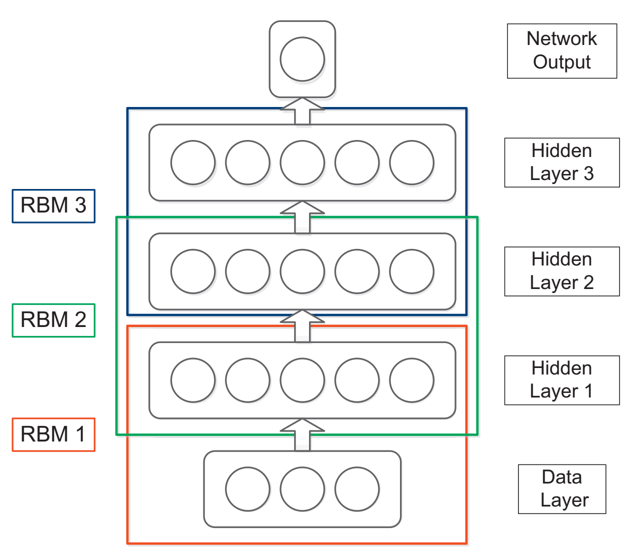

Fig. 1. Deep belief network structure.

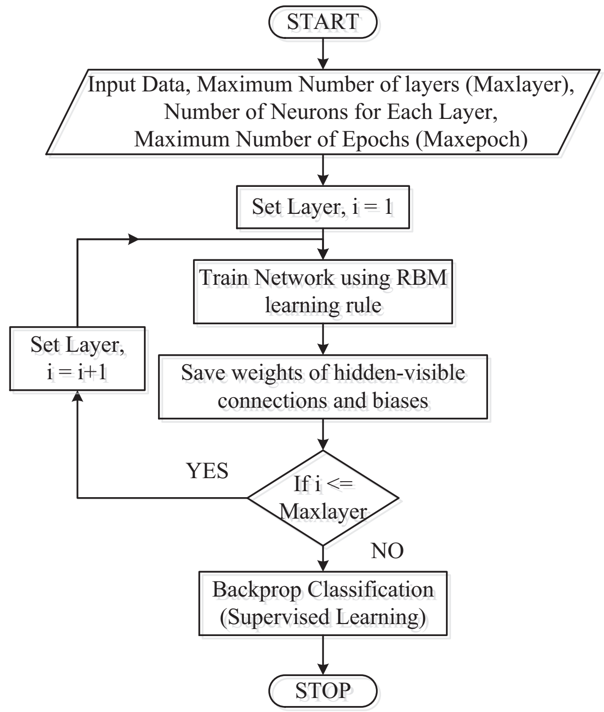

Fig. 2. Flowchart of the DBN training process.

maximum number of epochs for the training process. These input parameters are initialized before the start of the DBN training process.

Each layer of the DBN is trained using the RBM learning rule with two learning steps, namely positive and negative phases. The positive learning phase transfers data from bottom visible layer to hidden layer and determines the probability of generating hidden units as $p(h|v,W)$, whereas the negative phase operates as a reconstruction of the data from previous visible layer and determines the probability of generating visible units as $p(v|h,W)$. The observed data vector is used as an input to the visible units. The repetitive positive and negative phases in the training of RBM layers will result in trained weights and generated visible units after a maximum number of training epochs. The overall system function of DBN learning procedure can be expressed as the probability of generating a visible vector (v) as a function of weights and hidden vectors based on RBM learning rule [42]. The probability of generating a visible vector $p(v)$ by DBN learning process can be formulated using the probability of generating visible units in the reconstruction phase of the previous epoch $p(v|h,W)$ and the probability of generating hidden units in the positive phase of the current epoch $p(h|v,W)$, as

$$p(v) = \sum_h p(h|v,W)\, p(v|h,W). \tag{5}$$

The learning process continues through an iterative process from a lower layer to a higher layer till the maximum number of layers is trained. During the DBN layer by layer training process, each RBM is individually trained and the weights and biases are saved for further analysis. At the end of the training process, the data is transferred from bottom visible layer (data layer) to higher invisible layers throughout the DBN architecture [41]. Weights between visible and hidden layers and biases of the neurons in each DBN layer are optimized until the number of maximum epochs is reached. Notice that the DBN layer by layer training is an unsupervised learning process where class labels of the training data are not provided. The label information of the training data will be used during the succeeding supervised learning process, namely the back-propagation training. Detailed steps involved in the RBM training process are explained in the following subsection.

3.2. Training and validation of DBN classifier model

This section discusses training and validation of DBN classifier models. The DBN training process can be divided into two steps, namely stacked RBM learning and back-propagation learning. The following subsections describe these two steps in detail.

3.2.1. Stacked RBM learning

As discussed in Section 3.1, DBN is structured with stacked RBMs and training of a DBN classifier model is completed through sequential training of each individual RBM structure using the RBM learning rule. Each RBM unit consists of two layers. There are a number of neurons in each layer and there is no synaptic weight connection between neurons within the same layer [44]. Thus, the major training parameters considered in the RBM training process are the synaptic weights between layers and the biases of neurons. The transformation function used in the training process is the sigmoid transformation with log-likelihood function which transforms the data from one layer to another [42]

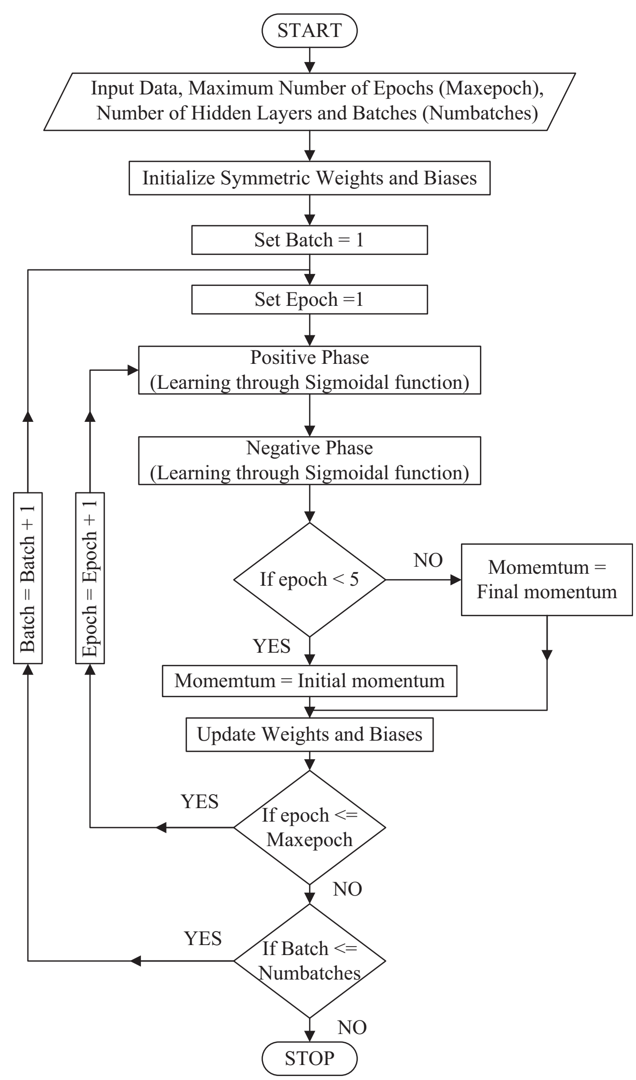

Fig. 3. Iterative learning process of RBM unit.

and it is shown as

$$P(s_i=1)=\frac{1}{1+\exp(-b_i-\sum_j s_j w_{ij})}, \tag{6}$$

where s_i and b_i are the state and bias of the ith neuron in the hidden layer, respectively; s_j is the state of the jth neuron in the visible layer, and w_{ij} represents the synaptic weight between the ith and jth neurons.

Fig. 3 shows the iterative learning process for one RBM unit. At the beginning of the training process, the synaptic weights and biases of all neurons in each RBM layer are initialized. After the initialization, the RBM unit will be trained iteratively with the input training data. The training dataset is often divided into mini-batches with a small number of data vectors and the weights are updated after treating each mini-batch. The states of neurons in the RBM hidden layer are determined through transforming the states of neurons in the visible layer with corresponding synaptic weights and the biases of hidden layer neurons with a transformation function. Each training epoch consists of two phases, the positive phase and the negative phase. The positive phase transforms the input data from the visible layer to the hidden layer whereas the negative phase operates as a reconstruction of the neurons of the previous visible layer. The positive and negative phases of RBM learning can be expressed mathematically by Eqs. (7) and (8) respectively, as

$$P(h_j=1\,|\,v)=\mathrm{sigm}\left(-b_j-\sum_k v_k w_{jk}\right), \tag{7}$$

$$P(v_k=1\,|\,h)=\mathrm{sigm}\left(-b_k-\sum_j h_j w_{jk}\right), \tag{8}$$

where h_j and v_k are the states for the jth neuron in hidden layer and the kth neuron in the visible layer, respectively. Eq. (7) denotes the positive phase learning process in which the states of neurons in the hidden layer are determined through sigmoid transformation, whereas Eq. (8) describes the negative phase learning process while the states of neurons in the visible layer are reconstructed. The visible and hidden layer neurons are binary stochastic neurons with binary states 0 or 1, which representing on and off conditions of the neurons in the learning process.

After the learning process for both positive and negative phases, synaptic weights and biases can be updated based on state vectors of neurons in both hidden and visible layers [42]. The update of synaptic weight, w_{jk}, can be formulated as

$$\Delta w_{jk}=\delta(\langle v_k h_j\rangle_{data}-\langle v_k h_j\rangle_{recon}), \tag{9}$$

where δ is a value between 0 and 1, denoting the learning rate; $\langle v_k h_j\rangle_{data}$ is the pairwise product of the state vectors for the jth neuron in the hidden layer and the kth neuron in the visible layer after the positive phase learning process, whereas $\langle v_k h_j\rangle_{recon}$ denotes the pairwise product after the negative phase learning process for reconstruction of the visible layer. The same learning rule is utilized for bias updating, but individual hidden and visible units are used instead of pairwise products [42]. To stabilize the RBM learning process, a momentum (m) is usually used in updating the synaptic weights and biases. With momentum, the weight update, Δw_{jk}, at the current epoch can be related to the weight update in the previous epoch and formulated as

$$[\Delta w_{jk}]_n=(m[\Delta w_{jk}]_{n-1})+\delta(\langle v_k h_j\rangle_{data}-\langle v_k h_j\rangle_{recon}). \tag{10}$$

In this study, the initial and final momentums utilized in the RBM training process are 0.5 and 0.9 respectively [42]. The learning parameters such as weights and biases of each RBM in a DBN classifier model will be continuously optimized until a maximum number of training epochs are reached. This completes the training of one RBM and the process will be continued until all RBMs in the DBN structure are trained. This iterative learning process that trains one individual RBM at a time successively is referred to as the layer by layer DBN training process.

Fig. 4 demonstrates the unsupervised DBN learning process using a sample DBN structure with four hidden layers. Considering the training process for the first RBM unit in the DBN structure as shown in Fig. 4(a), the input data is first given to the visible layer of this RBM unit. The next step is to transform the input data from the RBM visible layer to the hidden layer using visible layer parameters. While the training epoch reaches its maximum number and the training of the first RBM is accomplished, the hidden layer of this RBM unit becomes the visible layer of the second RBM unit. The training process is continued for the second and the third RBM units as shown in Fig. 4(b) and (c) respectively. The training of the DBN is accomplished through the successive training of each individual RBM unit, as shown in Fig. 4(d).

The trained weights and biases of the RBM layers from the unsupervised RBM learning are utilized in the next step of the DBN training process, referred to as supervised back-propagation learning process.

3.2.2. Back-propagation learning

After the successive layer by layer learning process, the next step of the DBN training is the supervised learning process, which will be accomplished by the back-propagation training algorithm. The supervised learning process further reduces the training error and improves the classification accuracy of the DBN classifier model. The supervised learning uses labeled data for the training of the DBN model. Unlike the unsupervised DBN training process that considers one RBM at a time, the back-propagation DBN training process considers all DBN layers simultaneously. The training error is calculated using model outputs and the target label data. The parameters of the DBN classifier model are updated in order to minimize the training error [45]. The back-propagation learning is continued until the network output reaches the maximum number of epochs. After the supervised back-propagation training process, the trained DBN classifier model can be further fine-tuned to improve the classification accuracy through certain fine-tuning algorithms. In this study, the conjugate gradient approach is employed for the DBN classifier model fine tuning purpose. More information regarding the DBN fine tuning can be found in Ref. [41].

After the DBN training and fine tuning process, another crucial step required to check the validity of the trained DBN classifier is the DBN model validation. Through the model validation process, the trained DBN model will be assessed for its classification validity. The synaptic weights and biases saved during the DBN training process will be used to determine misclassification errors, which will be employed as the validation metric to validate the performance of trained DBN classifier model. The misclassification error is calculated as the ratio of number of misclassified data points to the total number of data points in the input dataset. The validation process of the DBN classifier model is carried out for both the training and testing datasets. Based on misclassification errors, DBN model can be validated and then applied for the health diagnosis based on the acquired sensor data [28].

3.3. General procedure for DBN classification

Table 1 summarizes the procedure of applying the DBN based state classification for health diagnosis. As shown in the table, the first step is to define the diagnosis problem and identify possible system health states. After different system health states being identified, collecting sensory data from each health state is the

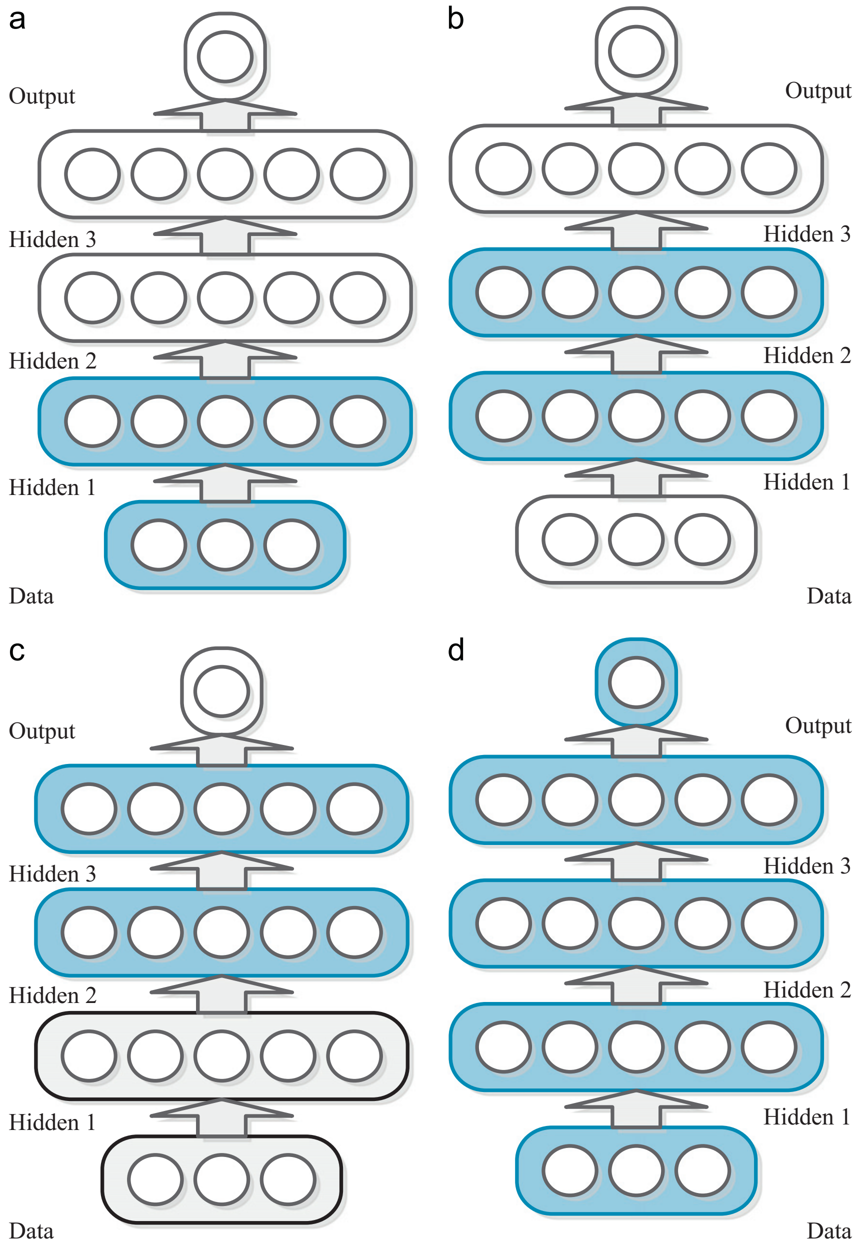

Fig. 4. Layer by layer training of DBN.

Table 1
Procedure for DBN classification.

Step 1:	Define the diagnosis problem and system health states
Step 2:	Collect and preprocess the sensory data for each predefined health state
Step 3:	Divide the dataset into training and testing datasets separately
Step 4:	Divide the training and testing datasets input into set of mini-batches with its target output
Step 5:	Develop the DBN classifier model
Step 6:	Train the DBN classifier model using the training dataset
Step 7:	Determine misclassification errors and validate the developed DBN classifier model
Step 8:	Perform diagnosis using the trained DBN classifier model

second step in the diagnosis process. To ensure a good diagnosis result, preprocessing of the sensory data becomes an essential step to convey the health relevant information from the raw sensory signals for the classification process. Preprocessed sensory data with known classes will be divided into training and testing datasets, which will be further divided into a set of mini-batches for the training of the DBN classifier model. The DBN classifier model is initialized and trained using the mini-batches of the training dataset. The detail architecture of the DBN classifier model is altered based on the input patterns with different known health states. The DBN classifier model is validated using the misclassification error determination process as discussed in the previous subsection. Steps 6–7 in Table 1 will be iteratively repeated until the model reaches the maximum number of epochs. Completely trained and validated DBN classifier model is then set to classify real time sensor signals to the corresponding health states.

4. Health diagnosis applications

The proposed health diagnosis using the DBN based state classification approach is demonstrated with benchmark diagnosis datasets and two structural health diagnosis case studies. The first structural health diagnostic case study employs 2008 IEEE PHM challenge data for aircraft engine health diagnosis, while the second case study used in this paper is power transformer mechanical fault diagnosis. The performance of health diagnosis using the proposed DBN based state classification approach for benchmark datasets and two case studies is compared with four existing diagnosis algorithms and the results are presented in the follow.

4.1. Benchmark datasets

Benchmark diagnosis datasets employed in this study are iris dataset [46], wine dataset [47], Wisconsin breast cancer diagnosis dataset [48] and *Escherichia coli* dataset [49]. The number of real attributes involved in the classification problem, number of classes and total number of instances of each dataset are listed in Table 2. The total instances for each benchmark dataset are divided into training dataset and testing dataset equally.

The architectures of each classifier model used in the comparison study are as follows. The trained DBN model architecture has one bottom data layer, one top output layer and three hidden layers in between. The data layer and the output layer are constructed with neurons denoting the input parameters and the target classes respectively, whereas 100 neurons are used for each hidden layer of the DBN models for all benchmark datasets. The BNN architecture has three processing layers: input, hidden, and output. Similar to the DBN architecture, the input layer and the output layer are constructed with neurons denoting the input parameters and the target classes respectively. A single hidden layer with 10 neurons and the sigmoidal transfer function are used to model the BNN model for each dataset. The 10×10 architecture of neurons is used to develop SOM models for all datasets. A Gaussian kernel function in one against all SVM model is used to train SVM classifiers for all benchmark datasets. To account for the stochastic nature of machine learning based diagnosis algorithms, the classification process for each dataset is repeated for 100 times and the average classification rates for training and testing datasets are summarized in Tables 3 and 4 respectively.

As shown in Tables 3 and 4, due to a higher complexity of the breast cancer dataset compared with other benchmark datasets used in the case study due to a relatively larger number of attributes involved in the classification, the correct classification rate of the breast cancer dataset is relatively lower for all algorithms. However, because of the capability to handle high complexity of sensory data through encoding richer and higher order network structures, the proposed DBN based health diagnosis approach provides a high diagnosis accuracy that outperforms other diagnosis algorithms used in the comparison, especially when dealing with high dimensional heterogeneous diagnosis inputs. As also revealed by other benchmark datasets, the overall performance of the proposed DBN based health diagnosis approach is robust and provides good diagnosis results in all the cases in comparison with four existing diagnosis algorithms.

Table 2
Problem complexity of the benchmark datasets.

Data	No. of real attributes	No. of classes	No. of instances
Iris	4	3	150
Wine	13	3	178
Breast cancer	30	2	569
E. coli	7	8	336

Table 3
Classification results of benchmark datasets (training).

Benchmark datasets	Classification rate training (%)				
	DBN	SVM	BNN	SOM	MD
Iris	100.00	100.00	100.00	96.67	98.67
Wine	100.00	100.00	97.73	95.46	94.31
Breast cancer	97.18	96.52	88.41	92.37	88.38
E. coli	98.39	95.74	94.59	93.21	96.77

Table 4
Classification results of benchmark datasets (testing).

Benchmark datasets	Classification rate testing (%)				
	DBN	SVM	BNN	SOM	MD
Iris	98.44	97.31	94.73	96.01	93.33
Wine	98.39	96.65	95.32	91.63	92.22
Breast cancer	96.78	93.58	84.88	87.25	86.31
E. coli	97.13	95.07	93.18	91.22	94.19

4.2. 2008 PHM challenge problem

The 2008 PHM challenge problem [50] is employed in this study to demonstrate the proposed DBN based diagnosis approach.

4.2.1. Case study description

The challenge dataset comprises of multivariate time series signals that were collected from an aircraft engine dynamic simulation process. Each time series signal is derived from a different degradation instance of the stochastic simulation for the same aircraft engine system. The complete dataset for each cycle of each engine unit consists of unit ID, operating cycle index, 3 values denoting different operational settings and 21 values for measurements from 21 sensors. The initial health condition for different engine unit is different and simulated sensory signals are contaminated with measurement noise. The operational setting values have significant effect on the engine performance and degradation behaviors, resulting in six different operation regimes [51,52]. Sensory signals were collected from six different operation regimes for all 218 engine units. The objective of this case study is to classify the different health conditions based on the sensory signals.

4.2.2. Data preprocessing and health state definition

The stepwise procedure for data preprocessing of the 2008 PHM challenge problem is detailed in Table 5. Following the literature, out of 21 sensors, sensory signals from 7 sensors are selected for all 218 engine units [51]. Each operating cycle is assigned to one corresponding operational regime based on its operational setting values. Each simulated engine unit has unique failure time and these failure times measured by the number of total operating cycles. The time series data of each engine unit is divided into four health conditions based on its proximity to the

Table 5
Procedure for preprocessing of PHM challenge data.

Step 1:	Selecting 7 sensory signals out of 21 sensory signals from complete engine simulation data
Step 2:	Collecting 7 sensor data for each cycle of 218 engine units
Step 3:	Assigning different operational regime class based on its operational setting values
Step 4:	Dividing the complete data into four health conditions based on its proximity to the failure time
Step 5:	Dividing each regime data into training and testing set

Table 6
DBN based diagnosis procedure of the 2008 PHM challenge problem.

Step 1:	Randomizing training and testing dataset
Step 2:	Dividing training and testing datasets input into 100 set of mini-batches
Step 3:	Developing target dataset for its corresponding 100 mini-batches of training and testing input datasets
Step 4:	Training DBN classifier model using RBM and back-propagation learning with training dataset
Step 5:	Testing the trained DBN classifier model
Step 6:	Determining classification rate for both training and testing data set of each regime

Table 7
2008 PHM challenge problem—different cases.

Case	Training dataset %	Testing dataset %
C1	70	30
C2	50	50
C3	30	70

failure time as follows. The sensory signals of each engine unit is first arranged in descending order based on the operation cycle index and the first 50 data points are termed as HS 4 (failure HS); the region between 75 and 125 data points is termed as HS 3; the region between 150 and 200 data points is termed as HS 2; and the region greater than 220 data points is termed as HS 1 (healthy HS). For each operation regime, the dataset for four health conditions is divided into training and testing datasets. The classifier models for each regime is then trained separately and tested with the testing dataset.

4.2.3. *DBN based health state diagnosis*

The procedure of developing a DBN classifier model for the 2008 PHM challenge problem is outlined in Table 6. The training and testing datasets of each regime are divided into 100 mini-batches. The training dataset is given as an input to the DBN classifier model for training using RBM and back-propagation learning. The trained DBN classifier model has one data layer in the bottom, one output layer at the top and three hidden layers with 100 neurons each. The maximum number of training epochs used in the RBM learning process and back-propagation learning process are set to 50 each in this study. The data layer and the output layer are constructed with 7 and 4 neurons denoting input parameters and target health states respectively. Similarly, the architectures of each classifier model used for the comparison purpose are as follows. The BNN architecture has three processing layers: input, hidden, and output, with 7, 5, and 4 neurons in each layer, respectively, and the transfer function used is TanH. A Gaussian kernel function is used to train the one-against-all SVM classification model. The SOM is trained using the architecture of 10×10 neurons. The health diagnostic model of each classification technique is trained using the training dataset and the accuracy and efficiency of the trained diagnostic model is validated with the testing dataset.

The case study is conducted in three different test cases denoting as C1 to C3 as shown in Table 7, which represent different sample sizes of training and testing datasets used in training and testing of the classification models. Similarly, the diagnosis process is repeated 100 times for all diagnosis algorithms and average classification rates from 100 runs are listed in Tables 8–13.

From the diagnosis results, clearly the classification rates of all algorithms are decreasing from case 1 to case 3, owing to a decrease of the number of training data being used. Compared with four existing diagnosis algorithms, the proposed DBN based diagnosis approach generally produces higher correct classification rates in most of the cases for regimes 1–6, as shown in Tables 8–13. This is mainly because of its capability of learning high

Table 8
Regime 1 classification results.

Methods	Classification rate training (%)			Classification rate testing (%)		
	C1	C2	C3	C1	C2	C3
DBN	92.38	91.29	85.21	90.72	90.20	85.64
SVM	90.39	89.26	84.32	92.11	88.53	84.16
BNN	88.23	84.88	81.11	86.82	84.27	80.93
SOM	84.20	83.55	80.60	83.56	81.27	79.47
MD	91.31	89.67	88.61	90.55	89.17	86.42

Table 9
Regime 2 classification results.

Methods	Classification rate training (%)			Classification rate testing (%)		
	C1	C2	C3	C1	C2	C3
DBN	96.44	93.45	92.13	95.80	92.56	89.89
SVM	93.32	92.34	92.63	91.71	90.83	89.78
BNN	90.40	89.31	88.11	90.36	87.04	83.63
SOM	87.94	85.64	82.63	85.32	85.18	81.71
MD	92.38	90.17	89.17	91.94	89.67	87.73

Table 10
Regime 3 classification results.

Methods	Classification rate training (%)			Classification rate testing (%)		
	C1	C2	C3	C1	C2	C3
DBN	95.45	93.43	92.42	94.86	91.54	91.35
SVM	94.50	92.42	91.33	93.19	90.09	90.54
BNN	89.27	87.29	83.50	88.45	86.50	80.75
SOM	84.39	81.49	81.01	82.36	79.31	78.00
MD	94.28	93.17	92.78	93.61	90.17	88.92

Table 11
Regime 4 classification results.

Methods	Classification rate training (%)			Classification rate testing (%)		
	C1	C2	C3	C1	C2	C3
DBN	94.37	93.44	92.18	92.46	92.71	92.64
SVM	94.63	92.65	90.74	93.88	90.73	90.46
BNN	90.43	88.79	87.86	89.34	88.50	86.55
SOM	87.77	87.10	84.67	86.68	84.28	81.84
MD	93.45	92.17	90.28	92.78	91.67	89.76

Table 12
Regime 5 classification results.

Methods	Classification rate training (%)			Classification rate testing (%)		
	C1	C2	C3	C1	C2	C3
DBN	92.42	91.57	90.60	91.67	91.10	90.80
SVM	91.54	93.82	89.23	90.41	91.97	89.83
BNN	86.70	84.96	81.83	85.64	83.69	79.74
SOM	83.44	81.93	80.38	81.02	79.54	76.75
MD	90.83	90.17	88.33	89.44	88.33	87.51

Table 13
Regime 6 classification results.

Methods	Classification rate training (%)			Classification rate testing (%)		
	C1	C2	C3	C1	C2	C3
DBN	96.70	95.31	93.73	95.70	93.39	92.67
SVM	95.49	94.51	92.54	94.08	93.68	91.31
BNN	91.49	89.08	88.02	91.05	86.16	84.45
SOM	88.55	86.81	84.69	88.38	84.28	82.92
MD	93.33	92.83	90.55	91.67	90.33	88.92

complexity non-linear relationships between the input sensory signals and the different health states through encoding richer and higher order network structures in the deep learning process with both supervised and unsupervised training. Taking the diagnosis results for the training dataset in C1 and regime 1 as an example, the classification rates are 92.38%, 90.39%, 88.23%, 84.20% and 91.31% respectively for DBN, SVM, BNN, SOM, and MD. Although DBN provided generally better results compared to for existing diagnosis algorithms in most of the cases, notice that the SVM classifier performs better for the testing dataset in C1 and regime 1 as shown in Table 8. As also indicated by the diagnosis results in this case study, the SOM algorithm generally performs less accurately compared to all other algorithms, mainly due to its inefficiency in learning the non-linearly separable health conditions based on the complex sensory signals [53].

4.3. Power transformer health diagnosis

In high voltage power systems, power transformers are the most expensive elements. Continuous monitoring of the transformer health conditions facilitates the evolution from traditional scheduled maintenance to condition-based maintenance, resulting in significant savings in lifecycle maintenance costs [54]. Due to the difficulty in acquiring direct measurements inside the transformer, indirect data acquisition techniques are often used in data collection for both diagnosis and prognosis of power transformers [55]. For example, electric parameters and analysis of moisture content of the cooling oil are often performed for the diagnosis and condition-based maintenance purposes [56]. The vibrations of the magnetic core and of the windings could characterize transitory overloads and permanent failures before any irreparable damage occurs. The indirect measurement of transformer vibration responses induced by the magnetic field loading enables the detection of mechanical failures of winding support joints inside the transformer. This case study aims at diagnosis of the mechanical faults of the power transformer winding support joints based on vibration signals acquired from the sensor network placed on the transformer external wall surface.

4.3.1. Description of the case study

In this case study, the winding support joint loosening is considered as the failure of interest. Detection of the failure modes is enabled by collecting vibration signals, induced by the magnetic field loading with a fixed frequency on the power transformer core. A validated finite element (FE) model of a power transformer was created in ANSYS 10 as shown in Fig. 5, where one exterior wall is uncovered to make the interior structure visible [35]. Fig. 6 shows 12 simplified winding support joints with 4 for each winding. The transformer is fixed at the bottom surface and a vibration load with the frequency of 120 Hz is applied to the transformer core. The joint loosening was realized by reducing the stiffness of the winding joints. Different combinations of the loosening joints will be treated as different health states of the power transformer which will be detailed in the next subsection.

4.3.2. Health states and simulations

For the purpose of demonstrating the proposed DBN based health diagnosis technique, 9 representative health states, as detailed in Table 14, were selected from all possible combinations of 12 winding support joint failures. Among these 9 selected health states, HS1 denotes the healthy condition without any loosening joint, whereas HS2 to HS9 are health states with either one or two loosening joints. 200 sets of random samples were

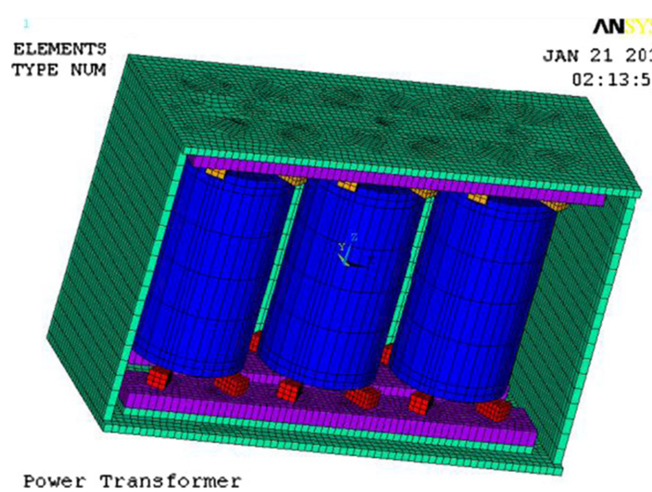

Fig. 5. A power transformer FE model.

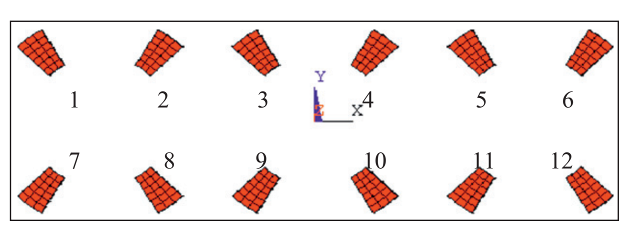

Fig. 6. Winding support joints.

Table 14
Definition of system health states.

Health state	1	2	3	4	5	6	7	8	9
Loosening joints	–	1	2	3	1,2	1,3	1,5	1,9	1,11

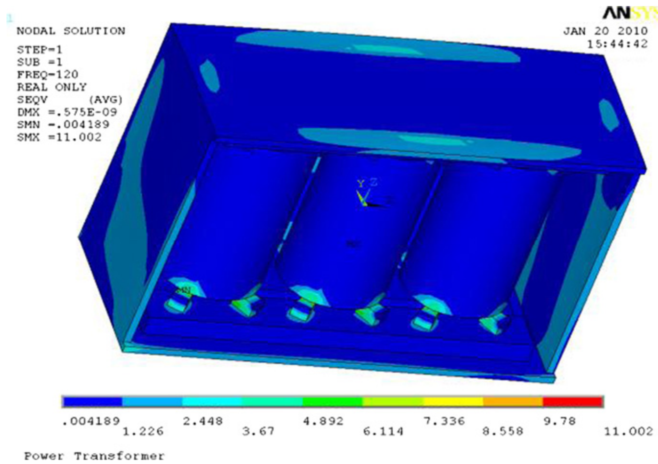

Fig. 7. Stress contour of the winding supports for the healthy state power transformer.

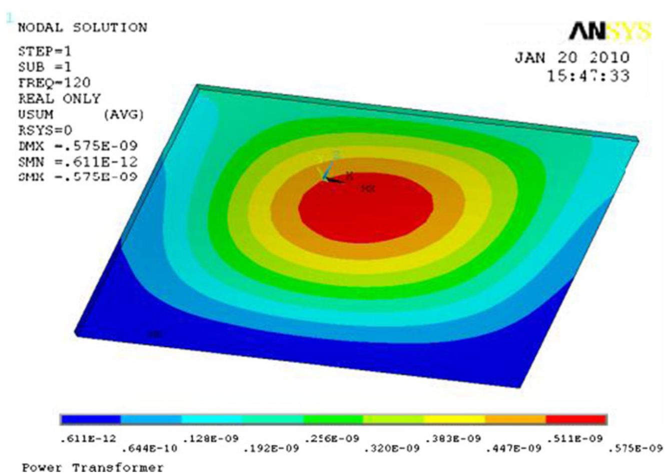

Fig. 8. Vibration displacement contour of the covering wall for the healthy state power transformer.

generated and the simulations for each of 9 health states were carried out and the vibration response of the displacement amplitudes for all the finite element nodes on the outer wall surfaces were saved as the simulation results. The stress contour of the healthy state power transformer from the structural simulation is shown in Fig. 7, whereas the vibration response of the covering wall is shown in Fig. 8. Considering uncertainties of the transformer and measurement noise, a sensor network has been designed for transformer faulty diagnostics in which 9 sensors has been optimally placed at the external wall surface of the transformer as shown in Fig. 9 [35]. With these 9 sensors placed on the external wall, the vibration amplitudes of the nodes at these 9 locations on the surface of the covering wall will be used as the simulated sensor (accelerometer) output. For more information regarding the sensor placement and sensory signal collection, readers can refer to the authors' previous studies [35].

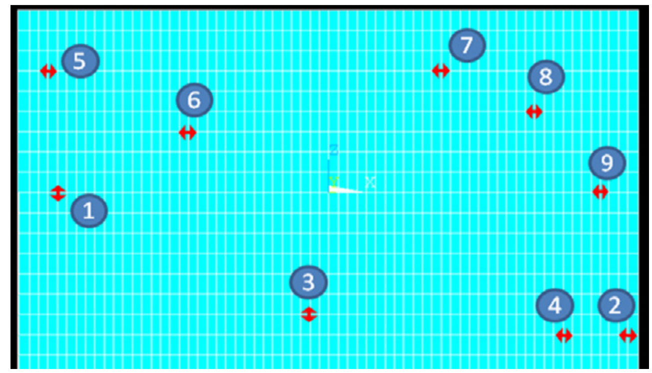

Fig. 9. Sensor placement on the covering wall of the power transformer.

Table 15
Case study 2—different diagnosis scenarios.

Case	HS1	HS2	HS3	HS4	HS5	HS6	HS7	HS8	HS9
Training set (sample size)									
C1	100	100	100	100	100	100	100	100	100
C2	50	50	50	50	50	50	50	50	50
C3	100	100	100	100	50	50	50	50	50
Testing set (sample size)									
C1	100	100	100	100	100	100	100	100	100
C2	150	150	150	150	150	150	150	150	150
C3	100	100	100	100	150	150	150	150	150

4.3.3. Health diagnosis using DBN approach

Three test cases are developed with different types of training and testing dataset combinations as shown in Table 15. Case 1 has equal number of data points in both training and testing datasets with 100 for each health state. Case 2 has less number of training data points but more testing data points for each health state, whereas case 3 assumes that available data for loosening of two joints is limited. The DBN model is then trained using the RBM learning function and the back-propagation algorithm as discussed in Section 3. The trained DBN model architecture has one data layer in the bottom, one output layer at the top and three hidden layers with 50, 50, and 100 neurons respectively. The datasets of each case are further divided into 50 mini-batches for training and testing of DBN classifier models respectively. The number of training epochs used in the RBM learning process and back-propagation learning process is set to 50. The data layer and the output layer are both constructed with 9 neurons based upon the dimensions of input and output variables. Architectures of existing diagnosis algorithms used for comparison are as follows. The BNN is composed of four processing layers: one input, two hidden, and one output, with 5, 6, 10 and 9 neurons in each layer, respectively, and the sigmoid transfer function is used. Similarly, a Gaussian kernel function is used in training of the one against all SVM model. The SOM is trained using a 15×15 architecture of neurons. With the trained health diagnosis models, the accuracy and efficiency of the diagnosis techniques are validated with the testing dataset. The average classification rates out of 100 repeated runs for the proposed DBN diagnosis approach as well as the SVM, BNN, SOM and MD classifier methods are compared and the results are listed in Table 16.

The classification rate of the trained DBN classifier model for the case 1 is found to be 90.99% for the training data set and 89.16% for the testing data set respectively. Although the misclassification of the training and testing data is relatively high due to large variability of the sensory data, the health diagnosis results by the proposed DBN diagnosis approach outperform the

Table 16
Case study 2—classification results.

Methods	Classification rate training (%)			Classification rate testing (%)		
	C1	C2	C3	C1	C2	C3
DBN	90.99	87.07	88.98	89.16	86.71	87.31
SVM	88.98	85.37	87.53	88.27	87.45	86.18
BNN	84.99	82.24	82.73	81.82	81.02	81.99
SOM	81.96	80.49	80.88	79.42	79.82	80.35
MD	87.56	82.22	88.15	86.44	81.18	87.39

other four existing diagnosis methods used in the comparison. This is mainly because of the capability of the DBN to learn the highly nonlinear relationships between the input sensory signals and the different health states of the power transformers. Notice that the classification rates of the trained SVM model for the training and testing data are better compared the other three existing diagnosis techniques. Also it is obvious from the case study results shown in Table 16, the SOM method generally performs less accurately, as it is an unsupervised learning technique which does not incorporate any prior knowledge about power transformer health conditions in the diagnosis thus performs less effective in learning the nonlinear relationships between the input sensory signals and corresponding system health states.

As observed from the case study results that SVM has comparable results with DBN, mainly because that the diagnostics case studies employed in the paper do not belong to the highly varying class [57]. As pointed out in the literature [45,57], SVM cannot handle the highly varying classes due to the curse of dimensionality. Moreover, Larochelle et al. presented that deep learning techniques can perform better classification for highly varying class functions compared to kernel based techniques, such as SVM [58]. It will very interesting in the future work to compare these two diagnosis techniques with highly varying class functions. As a summary, the proposed DBN diagnosis approach is able to provide accurate yet robust results for complex system health diagnosis applications, as demonstrated by the case studies that may involve multidimensional heterogeneous sensory signals and highly nonlinear relationships between diagnosis inputs and corresponding system health states.

5. Conclusion

This paper presented a novel multi-sensor health diagnosis approach using DBN based state classification. The DBN classifier model employs a hierarchical structure with multiple stacked Restricted Boltzmann Machines and works through a layer by layer successive learning process. The developed DBN based health diagnosis approach can be structured in three consecutive stages: (1) defining health states and collecting sensory data for DBN training and testing, (2) developing DBN based classification models for the diagnosis of predefined health states, and (3) validating DBN classification models with testing sensory dataset. The feasibility of the health diagnosis using DBN based health state classification was demonstrated with the classification benchmark datasets and two structural health diagnosis case studies. The diagnosis performances of the DBN classifier models were compared with four existing classification algorithms. Case study results indicated that DBN classifier model generally results in a better diagnosis performance in most of the cases for health diagnosis of complex systems, compared to other classification methods. In this paper, we assumed that the training data are labeled to different health states. In future study, it will be very interesting to look at the diagnosis performances of the DBN approach with a mix of labeled and unlabeled training data. The DBN will be further investigated for the health prognosis applications.

Acknowledgment

This research is partially supported by National Science Foundation through award CMMI#1200597 and Wichita State University through the University Research Creative Project Awards (UCRA).

References

[1] Pecht M. Prognostics and health management of electronics. New York, NY: John Wiley Publishing Co.; 2008.
[2] Licht T, Deshmukh A. Hierarchically organized Bayesian networks for distributed sensor networks. Ameican Society of Mechanical Engineering, Dynamic System Control Division 2003;71:1059–66.
[3] Marseguerra M, Zio E, Podofillini L. Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation. Reliability Engineering and System Safety 2002;77(2):151–65.
[4] Zio E. Review reliability engineering: old problems and new challenges. Reliability Engineering and System Safety 2009;94(2):125–41.
[5] Dekker R. Applications of maintenance optimization models: a review and analysis. Reliability Engineering and System Safety 1996;51:229–40.
[6] Barata J, Guedes Soares C, Marseguerra M, Zio E. Simulation modelling of repairable multi-component dteriorating systems for 'on condition' maintenance optimization. Reliability Engineering and System Safety 2002;76(3): 255–64.
[7] Grall A, Berenguer C, Dieulle L. A condition-based maintenance policy for stochastically deteriorating systems. Reliability Engineering and System Safety 2002;76(3):167–80.
[8] JAM Weide, van der, Pandey MD, Noortwijk JM, van. Discounted cost model for condition-based maintenance optimization. Reliability Engineering and System Safety 2010;95(3):236–46.
[9] Smidt-Destombes KS de, Heijden MC van der, Harten A. On the availability of a k-out-of-N system given limited spares and repair capacity under a condition based maintenance strategy. Reliability Engineering and System Safety 2004;83(3):287–300.
[10] Ebeling CE. An introduction to reliability and maintainability engineering. Long Grove, IL: Waveland; 1997.
[11] Coit DW, Jin T. Gamma distribution parameter estimation for field reliability data with missing failure times. IIE Transactions 2000;32(12):1161–6.
[12] Elsayed EA. Perspectives and challenges for research in quality and reliability engineering. International Journal of Production Research 2000;38(9): 1953–76.
[13] Myotyri E, Pulkkinen U, Simola K. Application of stochastic filtering for lifetime prediction. Reliability Engineering and System Safety 2006;91(2): 200–8.
[14] Tinga T. Application of physical failure models to enable usage and load based maintenance. Reliability Engineering and System Safety 2010;95(10): 1061–75.
[15] Cadini F, Zio E, Avram D. Model-based Monte Carlo state estimation for condition-based component replacement. Reliability Engineering and System Safety 2009;94(3):752–8.
[16] Zio E, Di Maio F. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear power plant. Reliability Engineering and System Safety 2010;95(1):49–57.
[17] Baraldi P, Razavi-Far R, Zio E. Classifier-ensemble incremental-learning procedure for nuclear transient identification at different operational conditions. Reliability Engineering and System Safety 2011;98(4):480–8.
[18] Niu G, Yang B, Pecht M. Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance. Reliability Engineering and System Safety 2010;95(7):786–96.
[19] Alguindigue IE, Loskiewicz-Buczak A, Uhrig RE. Monitoring and diagnosis of rolling element bearings using artificial neural networks. IEEE Transactions on Industrial Electronics 1993;40(2):209–17.
[20] Gebraeel N, Lawley M, Liu M. Vibration-based condition monitoring of thrust bearings for maintenance management. Intelligent Engineering System Through Artificial Neural Network 2002;12:543–51.
[21] Li Y, Billington S, Zhang C. Dynamic prognostic prediction of defect propagation on rolling element bearings. Lubrication Engineering 1999;42(2): 385–92.
[22] Martin KF. Review by discussion of condition monitoring and fault diagnosis in machine tools. International Journal of Machine Tools and Manufacture 1994;34(4):527–51.
[23] Booth C, McDonald JR. The use of artificial neural networks for condition monitoring of electrical power transformers. Neurocomputing 1998;23: 97–109.

[24] Macian V, Tormos B, Olmeda P, Montoro L. Analytical approach to wear rate determination for internal combustion engine condition monitoring based on oil analysis. Tribology International 2003;36(10):771–6.
[25] Breikin T, Kulikov G, Arkov V, Fleming P. Dynamic modelling for condition monitoring of gas turbines: genetic algorithms approach. In: Proceedings of the 16th IFAC World Congress; 2005.
[26] Arel I, Rose DC, Karnowski TP. Deep machine learning—a new frontier in artificial intelligence research. Computational Intelligence Magazine 2010;5(4): 13–8.
[27] Huang R, Xi L, Li X, Richard Liu C, Qiu H, Lee J. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mechanical Systems and Signal Processing 2007;21:193–207.
[28] Khomfoi S, Tolbert LM. Fault diagnostic system for a multilevel inverter using a neural network. IEEE Transactions on Power Electronics 2007;22(3): 1062–9.
[29] Samanta B. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. Mechanical Systems and Signal Processing 2004;18:625–44.
[30] Yang BS, Hwang WW, Kim DJ, Chit Tan A. Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines. Mechanical Systems and Signal Processing 2005;19:371–90.
[31] Wong M, Jack LB, Nandi AK. Modified self-organising map for automated novelty detection applied to vibration signal monitoring. Mechanical Systems and Signal Processing 2006;20:593–610.
[32] Abbasion S, Rafsanjani A, Farshidianfar A, Irani N. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. Mechanical Systems and Signal Processing 2007;21:2933–45.
[33] Weston J, Watkins C. Multi-class support vector machines. Citeseer; 1998.
[34] Widodo A, Yang BS. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing 2007;21(6):2560–74.
[35] Wang P, Youn BD, Hu CA. Probabilistic detectability-based structural sensor network design methodology for prognostics and health management. In: Annual conference of the prognostics and health management society; 2010.
[36] Cheng S, Tom K, Pecht M. Anomaly detection of polymer resettable circuit protection devices. IEEE Transactions on Device and Materials Reliability 2012;12(2):420–7.
[37] Cheng S, Pecht M. Using cross-validation for model parameter selection of sequential probability ratio test. Expert Systems with Applications 2012;39: 8467–73.
[38] Zhang L. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. Expert Systems with Applications 2010;37(8): 6077–85.
[39] Saimurugan M, Ramachandran KI, Sugumaran V, Sakthivel NR. Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine. Expert Systems with Applications 2011;38(4):3819–26.
[40] Geramifard O, Xu JX, Pang C, Zhou J, Li X. Data-driven approaches in health condition monitoring—a comparative study. In: Proceedings of the 8th IEEE international conference on control and automation (ICCA); 2010. p. 1618–22.
[41] Hinton GE, Osindero S, Teh YW. A Fast learning algorithm for deep belief nets. Neural Computation 2006;18:1527–54.
[42] Hinton GE. A practical guide to training restricted Boltzmann machines. Momentum 2010;9:1.
[43] Ackley DH, Hinton GE, Sejnowski TJ. A learning algorithm for Boltzmann machines. Cognitive Science 1985;9:147–69.
[44] Lee H, Grosse R, Ranganath R, Ng AY. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: International conference on machine learning. Montreal, Canada; 2009. p. 609–16.
[45] Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. Advances in Neural Information Processing Systems 2007:153.
[46] Fisher RA. The use of multiple measurements in taxonomic problems. Annual Eugenics 1936;7(2):179–88.
[47] Forina M, Leardi R, Armanino C, Lanteri S. PARVUS an extendable package of programs for data exploration. Classification and correlation. Amsterdam: Elsevier Scientific Software; 1988.
[48] Mangasarian OL, Street WN, Wolberg WH. Breast cancer diagnosis and prognosis via linear programming. Operations Research 1995;43(4):570–7.
[49] Nakai K, Kanehisa MA. Knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 1992;14:897–911.
[50] Saxena A, Goebel K, Simon D, Eklund N. Damage propagation modeling for aircraft engine run-to-failure simulation. In: Proceedings of the international conference on prognostics and health management; 2008.
[51] Hu C, Youn BD, Wang P. Ensemble of data-driven prognostic algorithms with weight optimization and *K*-fold cross validation. In: Annual conference of the prognostics and health management society; 2010.
[52] Wang T, Yu J, Diegel S, Lee J. A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In: Proceedings of the international conference on prognostics and health management; 2008.
[53] Gao P, Woo W, Dlay S. Neural network approaches to nonlinear blind source separation. In: Proceedings of the 8th international symposium on signal processing and its applications; 2005.
[54] Leibfield T. Online monitors keep transformers in service. IEEE Computer Applications in Power 1998;11(3):36–42.
[55] Rivera HL, Garcia-Souto JA, Sanz J. Measurements of mechanical vibrations at magnetic cores of power transformers with fiber-optic interferometric intrinsic sensor. IEEE Journal on Selected Topics in Quantum Electronics 2000;6(5):788–97.
[56] Allan D, Brundell M, Boyd K, Hinde D. New techniques for monitoring the insulation quality of in-service HV apparatus. IEEE Transactions on Electrical Insulation 1992;27(3):578–85.
[57] Bengio Y, Delalleau O, Le Roux N. The curse of highly variable functions for local kernel machines. In: Proceedings of the advances in neural information processing systems (NIPS'05), p. 18.
[58] Larochelle H, Bengio Y, Louradour J, Lamblin P. Exploring strategies for training deep neural networks. Journal of Machine Learning Research 2009;10:1–40.