Machine Health Surveillance System by Using Deep Learning Sparse Autoencoder

Faizan Ullah
Bacha Khan University Charsadda
Abdu Salam
abdul wali khan university
Muhammad Abrar
Bacha Khan University Charsadda
Masood Ahmad
Abdul Wali Khan University Mardan
Fasee Ullah (faseekhan@gmail.com)
Sarhad University of Science and Information Technology
Atif Khan
Islamia College Peshawar
Abdullah Alharbi
Taif University
Wael Alosaimi
Taif University

Research Article

Keywords: Machine Health Surveillance System, Deep Learning, Autoencoder, rapidly growing research area, art achievement, accuracy

Posted Date: November 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-911385/v1

Version of Record: A version of this preprint was published at Soft Computing on January 18th, 2022. See the published version at https://doi.org/10.1007/s00500-022-06755-z.

Machine Health Surveillance System by Using Deep Learning Sparse Autoencoder

Faizan Ullah[1], Abdu Salam[2], Muhammad Abrar[3], Masood Ahmad[2], Fasee Ullah[4*], Atif Khan[5], Abdullah Alharbi[6], Wael Alosaimi[6]

[1]Department of Computer Science, Bacha Khan University, Charsadda, Pakistan
[2]Department of Computer Science, Abdul Wali Khan University, Mardan, Pakistan
[3]Department of Computer Science Mohi-Ud-Din Islamic University, Nerain Sharif AJ & K, Pakistan
[4]Department of Computer Science, Sarhad University of Science and Information Technology, Pakistan
[5]Department of Computer Science, Islamia College Peshawar, Peshawar 25120, Pakistan
[6]Department of Information Technology, College of Computers and Information Technology, Taif University, P.O.Box 11099, Taif 21944, Saudi Arabia

Corresponding Author:
Fasee Ullah[4]
Email address: fasee.csit@suit.edu.pk

Abstract

Deep learning is a rapidly growing research area having state of art achievement in various applications including but not limited to speech recognition, object recognition, machine translation, and image segmentation. In the current modern industrial manufacturing system, Machine Health Surveillance System (MHSS) is achieving increasing popularity because of the widespread availability of low cost sensors internet connectivity. Deep learning architecture gives useful tools to analyze and process these vast amounts of machinery data. In this paper, we review the latest deep learning techniques and their variant used for MHSS. We used Gearbox Fault Diagnosis dataset in this paper that contains the sets of vibration attributes recorded by SpectraQuest's Gearbox Fault Diagnostics Simulator. In addition, the authors used the variant of auto encoders for feature extraction to achieve higher accuracy in machine health surveillance. The results showed that the bagging ensemble classifier based on voting techniques achieved 99% accuracy.

Introduction

Nowadays, the deep learning is the promising area of research in artificial intelligence. Deep learning is a subcategory of machine learning that uses the neural network to design a highly accurate system. Neural network architecture contains several layers, made of neuron, that apply a non-linear and linear transformation between the layers. There are a number of successful implementations of supervised and unsupervised based deep learning techniques in computer vision and natural language processing (Almiani, AbuGhazleh, Al-Rahayfeh, Atiewi, & Razaque, 2020). The deep learning models learns inn hierarchical fashion where the lower-level features derived from high-level features. For example, in classification of an image-processing task, the deep learning algorithm grabs pixel value in the input layer and allocates label value to the object in the output layer. Among these two layers, there are a number of internal layers, known as hidden layers, that assembles successive higher-order features (LeCun, Haffner, Bottou, & Bengio, 1999). The term “deep” in DL represents several transformation layers of representation that lie between the model’s inputs and output. In deep neural network, there is no standard number of layers, but most research in this area considers at least more than two layers must be present. The reason for the success of deep learning is that it avoids the process of feature engineering (Komar, Yakobchuk, Golovko, Dorosh, & Sachenko, 2018). In conventional machine learning algorithms, feature engineering is the task of choosing relevant features compulsory for the algorithm to work efficiently. This task is complex and time-consuming as the precise selection of features is important to the performance of the algorithm(Bouktif, Fiaz, Ouni, & Serhani, 2018). In this paper, four deep learning algorithms i.e. Auto-Encoders (AE), Restricted Boltzmann Machine (RBM), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) and variants have been discussed and applied in the area of MHSS in detail. The data-driven algorithms and the industrial internet of thing (IoT) are having a revolution in the manufacturing. It is empowering computer network to collect a massive amount of data from the linked machines and transform it into valuable information (O’Donovan, Leahy, Bruton, & O’Sullivan, 2015; F. Tao, Qi, Liu, & Kusiak, 2018). As an important module of the modern manufacturing system, monitoring of health through the machine has fully accepted the big data revolution (M. Chen, Ma, Song, Lai, & Hu, 2016; Luo, Wu, Gopukumar, & Zhao, 2016). New models of bottom-up solutions for fault detection have overtaken traditional physics-based techniques. The data-driven systems for MHSS accomplish this by diagnosing certain faults arising in the system and has the ability to predict the remaining useful life of the machine (Nuhic, Terzimehic, Soczka-Guth, Buchholz, & Dietmayer, 2013). The developmental paradigms for complex and dynamic systems are difficult because of the noisy working conditions which delays constructing physical models. Similarly, the effectiveness and flexibility of physics-based models are hindered as they have no ability of the real time data updates(Mosterman, 1997). Deep learning has proven to perform as a connecting entity between huge machinery data and intelligent MHSS. Deep learning models are developed using multi-layer approach to classify the data pattern. Deep learning goes as back as the 1940s, but recent popularity contributes mostly to the computer vision, speech recognition systems, bioinformatics and audio recognition(W. Liu et al., 2017). The increased use of the deep learning based model can be contributed to the cheaper GPU, exponential growth in data, and the deep learning research strength.

Cheaper GPU: The performance of deep learning models requires high end GPUs, and the recent development of cheaper GPU made it easier to build more efficient models. Resultantly, it has significantly reduced the time required to run algorithms specific to deep learning. As in (Raina, Madhavan, & Ng, 2009) the research has shown, the running time for a four-layered Deep Belief Network (DBN) reduced to a single day, as compared to several weeks, for over 100 million parameters.

Exponential growth in data: Our every operation is digitized nowadays, stored by sensors and computer, connected to the Internet, and stored in the cloud. As shown in (Yin, Li, Gao, & Kaynak, 2014) that in industry associated systems such as electronics and industrial informatics, having 1000 Exabyte produced per annum and a 20-fold rise can be contemplated in the coming ten years. Research in(Al-Sarawi, Anbar, Abdullah, & Al Hawari, 2020) predicts that a minimum of 30 billion systems will be connected by end of 2025.

Deep Learning strengthens research: the ever first success of deep learning is the pre-training practice in the unsupervised manner (Hinton, 2007), Hinton suggested maintaining a single layer at a time deploying RBM and fine-tune by utilizing backpropagation in 2007.

The main contribution of this article is to monitor the health of a by using Gearbox Fault Diagnosis dataset, that contains the sets of vibration attributes recorded by SpectraQuest's Gearbox Fault Diagnostics Simulator. The rest of the paper is organized as follows: Section 2 has the literature about recent work on deep learning models used for MHSS. In Section 3, we present our proposed methodology in detail. In section 4, experimental study and results have been carried out in a tool wear prediction task. In section 5 we conclude this paper and outline for future direction.

Related work

In the area of MHSS, the multi-layer conventional neural network has been used for many years (Su & Chong, 2007). In MHSS, recently a huge number of deep learning algorithms has significantly increased. Deep neural network based on Restricted Boltzmann machine or autoencoder smooths the training phase and improve the classification power to classify data. Recursion neural network and convolution neural network come up with highly complex and advance composition to extract a pattern from machine data. For MHSS based on deep learning, the first layer is used for input. To diagnose a task in the top layer, discrete values are used which is called the softmax layer. To diagnose a task having continuous values, linear regression is used. A deep learning survey based on MHSS is carried out in three different architectures and summarized in Table 1.

MHSS using Auto encoder

Auto encoder (AE) algorithms can extract high-level information automatically from machine data. Honga et al. (Hoang & Kang, 2019) suggested a single layer AE solution to classify induction motor fault based on the neural network as shown in Figure 1 and Figure 2. The author focuses on how to overcome overfitting due to training data because the available dataset is having a limited dimension. Their model uses dropout architecture to hide the random output of the hidden layer. Lu et al. (Lu, Wang, Qin, & Ma, 2017) proposed a compressive study on stacked autoencoder denoising techniques having three hidden layers to diagnosis a fault of the rotary machine. In the suggested study, experimental work was carried out in cross working and single working environment which effects deep architecture, input size denoising operation, and sparsity control check. Tao et al.(S. Tao, Zhang, Yang, Wang, & Lu, 2015) presented many structures of deep learning based on two-layer SAE. To perform a classification task of fault diagnosis proposed system consists of the various hidden layer having masking probability. The input dataset is huge in size, which causes overfitting and high computation. Therefore, some authors suggested normalizing input data using AE models. Jia et al. (Jia, Lei, Lin, Zhou, & Lu, 2016) use spectra frequency of time series for features extraction from raw input data. The extracted features are utilized for the classification of fault diagnosis applications. Sun et al. (Sun, Yan, & Wen, 2017) proposed a soft threshold non-linear and digital wavelet model to operate the vibration signal. SAE is used to classify fault diagnosis over the preprocessed input signal. Liu et al. (H. Liu, Li, & Ma, 2016) proposed Fourier wavelet transform used to extract features. SAE based on a deep neural network is used to classify roller bearing fault diagnosis. SAE utilized a dropout and ReLU activation function, which cause to prevent overfitting issue. In Input, the data-set is normalized by short time Fourier transform to generate a normalized spectrogram. The two-layer SAE based DNN is then used to classify bearing rolling fault diagnosis. Galloway et al. (Galloway, Catterson, Fay, Robb, & Love, 2016) extracts spectrograms from tidal turbine data using two unit layer SAE based on DNN to diagnose the fault. In (K. Li & Wang, 2015) SAE based on DNN is used to classify fault diagnosis and principal component analysis has been utilized to extract features from input.

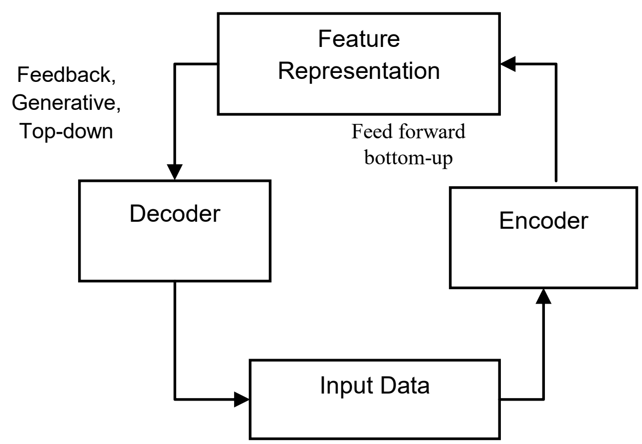

Figure 1: Encoding of Recovering the Original Data and the Input Data

Auto encoders have been also utilized to extract a useful pattern from multiple sensors and time series data by Chen and Li (Zhuyun Chen & Li, 2017) and Kishore et al. (Reddy, Sarkar, Venugopalan, & Giering, 2016). Informative statistical input features with high frequency and specific time domain are derived from signal vibration and then utilized for pattern classification. Some researchers have taken a step further to investigate the application of autoencoders with other machine learning approaches.

Jiang et al. (Jiang, Xie, Wang, Chen, & He, 2017) proposed a scheme to extract frequency domain features by using auto-encoders and use conventional classifier algorithms such as random forest and support vector machine to perform the classification task. Wang et al. (Wang, Zhao, Pei, & Tang, 2016) suggested a novel auto-encoder unsupervised features learning technique, which prevents overfitting and changes gradient direction for a fault classification task. Mao et al. (Mao, He, Li, & Yan, 2017) presented a novel version of autoencoder for a fault recognition named extreme learning machine. Compare with conventional auto encoder extreme machine is quick and efficient.

MHSS using RBM

A significant type of autoencoder is called Restricted Boltzmann Machine (RBM). He et al. (He, He, & Bechhoefer, 2016) presents a RBM based technique for the prediction of the remaining useful life (RUL) of machine. After training the input data, an additional layer is appended at the top of RBM. Zhu et al. (Zhu, Chen, & Peng, 2018) suggested a modified RBM for RUL prediction of machine. In their proposed study, a new hidden unit is appended to achieve training function objective of RBM. After that, a Self-Organizing Map (SOM) unsupervised algorithm has been modelled to transform a pattern recognized by a novel RBM called health value. The health value has been, then, utilized for the prediction of RUL using harmony methodology. In (C. Li et al., 2015) support vector multi-model classification technique was suggested for fault recognition of gearbox. The proposed algorithm extracts three types of feature from vibration signal: time-frequency, frequency and time. After that, a three layer Gaussian Bernoulli Boltzmann Machines (GDBM) has been applied to extracted features. In every GDBM softmax, the unit layer is attached at the top. After fine-tuning, the statistical output of the softmax layer generated from three GDBM has been assessed by support vector classification structure to predict the final stage of classification. Li et al. (C. Li et al., 2016) developed a single unit GDBM for feature containing three different tones i.e. time-frequency, frequency and time. To predict a fault, the combination of softmax and the stacked layer is applied on top of GDBM. Li et al. (Tseng, Chen, Kao, & Lin, 2016) applied a two layers DBM to extract deep information having statistical parameters based on wavelet packet transform (WPT) of input sensory signal to recognized fault diagnosis for the gearbox. In the proposed study author focus on the fusion of data, two different DBMs were trained over vibratory and acoustic signals. Random forest was applied to fuse information recognized by two layers of DBM. To make a useful application of DBN having DNN, Ma et al. proposed a novel model to minimize the evaluation under bearing accelerated life test (Ma, Chen, Wang, Liu, & Li, 2016).

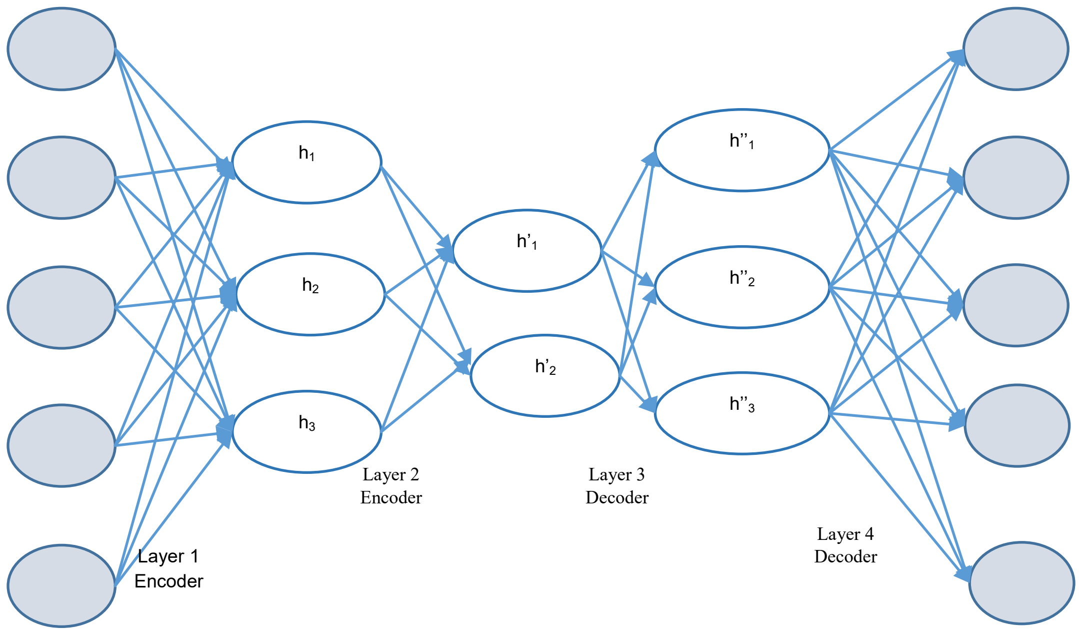

Figure 2. **Rearrange the Value for Single Layer of an Autoencoder**

Root mean square (RMS), a variant of statistical feature, used by Weibull statistics that can keep away fluctuation areas. Then the feature of the frequency domain has been extracted as input raw data. To conclude, the proposed methodology is given in Figure 3. Shao et al. (Shao, Sun, Yan, Wang, & Gao, 2017) proposed a DBN framework to diagnosis a fault of induction motor having the uninterrupted utilization of signal vibration data. Tao et al. (J. Tao, Liu, & Yang, 2016) proposed a novel framework multi-sensor DBN base information fusion model to predict bearing fault diagnosis. In the first phase, three signals of vibration extracted from three sensors and has been combined to extract 14 time-domain statistical features using DBN as input model. In the training phase, to determine the iteration number, a predefined threshold value was utilized. In (Zhiqiang Chen, Li, & Sánchez, 2015) the researchers presented a model to extract a feature vector speed, load measure, domain frequency features, and domain time features was given as input to DBN having DNN for classification of gearbox fault. In (Gan & Wang, 2016), Gan et al. presented the hierarchical framework to diagnosis network fault pattern and classification. The spinning element of bearings containing two adjacent stages having four variant faults are determined.

MHSS using CNN

A new approach, which gains high attention of researcher for the high dimensional data such as time series, images, and information is CNN. CNN is made of neural networks, which contain extracted feature during the training phase and the assigned weights are adjusted during the training phase. The fundamental truth that CNN can resolves is the issue of manual extraction of features The main feature and advantage of CNN is the process of feature extraction automation

(Hussain et al., 2019). CNN is a simplified form of neural networks, which utilize the convolution process instead of a conventional multiplication matrix.

In some frameworks, the raw machinery information can be gathered in a single 2D format like spectrum frequency time, where in some cases, the presented data is in a 1D format such as time series. CNN algorithms are capable to learn robust and complex pattern with the convolutional layer from both format. Traditionally, convolutional layers, with the help of filters, are able to extract local patterns from raw data while convolutional layers can further be stacked together to create meaningful patterns. Liu et al. (R. Liu, Meng, Yang, Sun, & Chen, 2016) presented a 2D-CNN framework for four variety of spinning machinery conditions classification. Two sensors, that are set perpendicular to each other, generate two accelerometer signals processed by DFT. The adopted CNN model consists of a directly connected layer and a single convolutional layer. After that, the top softmax-layer is applied for classification. Babu et al. (Babu, Zhao, & Li, 2016) created 2D deep convolution neural network for the prediction of RUL of a machine over normalized variable time sequence collected from sensor signals. In the proposed study, the mean pooling approach is applied rather than max pooling. As the value RUL is a continuous number, so the top layer is set off as a linear-regression layer.

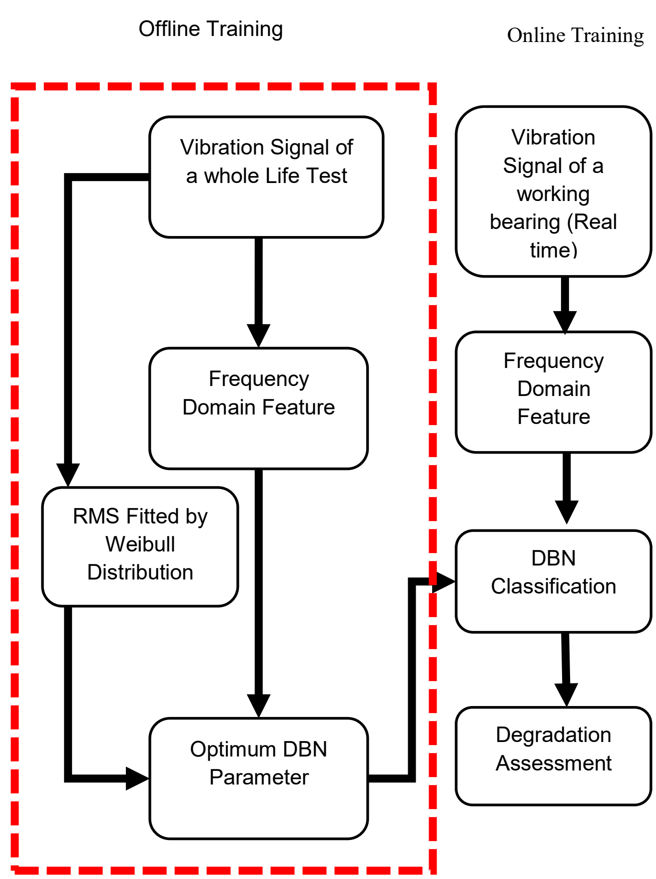

Figure 3. Proposed Framework for DBN-DNN for Assessment of Bearing

Ding et al. (Ding & He, 2017) suggested a novel approach based on deep Convolutional Network (ConvNet), having wavelet packet energy (WPE) picture that has been utilized as an input to predict spindle bearing fault. To completely recognize the hierarchical structure, the multi-scale layer is appended succeeding to the end of the convolutional layer. The output of the last convolution layer is concatenated with the first pooling layer. Guo et al. (Guo, Chen, & Shen, 2016) suggested an algorithm based on adaptive deep convolution neural network (ADCNN). The hierarchical module is designed to predict fault size and pattern of fault. The process of fault pattern decision phase ADCNN was first applied to recognize the type of fault. ADCNN having an identical structure is utilized to recognize fault dimension. A function f is used to classify the fault type. The function f is determined as summation of probability given and shown in (1):

$$f = \sum_{j=0}^{c} a_i p_j \tag{1}$$

Table 1. Summary Of Deep Learning Used In Health Monitoring System Applications

Technique	Ref.	Application	Strengths	Limitations
Auto-Encoders	(Zhuyun Chen & Li, 2017)	Bearing fault diagnosis	Capable to update the values to learn useful	For training process a lot of fine tuning
	(Hoang & Kang, 2019)	Rolling element bearing fault diagnosis	Overcome over fitting due to training	A lot of processing time required
	(Lu et al., 2017)	Fault diagnosis of rotary machinery	Easy to implement	
	(S. Tao et al., 2015)	Bearing fault diagnosis	Masking probability	Over fitting and low computation
	(Jia et al., 2016)	Diagnosis of rotating machinery	spectra frequency of time series	Threshold occur
	(Sun et al., 2017)	Bearing fault diagnosis	Dimensionality reduction	
	(H. Liu et al., 2016)	Rolling bearing fault diagnosis	prevent over fitting issue	Low computation
	(H. Liu et al., 2016)		generate normalized spectrogram	Cant extract useful information
	(Galloway et al., 2016)	Diagnosis of tidal turbine vibration		Low computation
	(K. Li & Wang, 2015)	Recognition of Signal and diagnosis for spacecraft	prevent over fitting	Cant extract useful information
	(Wang et al., 2016)	Transformer fault diagnosis	Easy to track the cost/ loss	
	(Reddy et al., 2016)	Fault classification in large flight data	statistical features are extracted in frequency and time domain	A lot of processing time required
	(Jiang et al., 2017)	Intelligent fault diagnosis of rotary machinery	Handcraft based extracted features	Handcraft based extracted features
	(Mao et al., 2017)	Bearing fault diagnosis		A comparative study
Denoising auto encoder	(Thirukovalluru, Dixit, Sevakula, Verma, & Salour, 2016)	To diagnosis fault in health prognosis applications	Higher accuracy for feature extraction	Noise at input level

Technique	Ref.	Application	Strengths	Limitations
	(Lu et al., 2017)	Health state identification	supervised fine-tuning	High computation
	(Yan & Yu, 2019)	Anomaly detection for gas turbine	DE noising or Compression.	
Vibrational auto encoder RBM	(Yoon et al., 2017)	Predict remaining useful life (RUL) estimation	Learns which noise distribution to Insert at code.	Can be difficult to optimize
	(He et al., 2016)	RUL	Predict Future root mean square (RMS) value	
	(Zhu et al., 2018)	Prediction of bearing RUL	Transform the learned pattern	Difficult to train
	(C. Li et al., 2015)	Fault recognition of gearbox	extract three types of feature from vibration signal	High computation
	(C. Li et al., 2016)	Gearbox fault diagnosis	extract deep information having statistical parameters	Fusion of data
	(Ma et al., 2016)	Bearing degradation assessment	capable to generate patterns if missing data	Difficult to train model
	(Shao et al., 2017)	Fault diagnosis of induction motors	away areas of fluctuation having statistical parameter	Threshold occur
	(J. Tao et al., 2016)	Bearing fault diagnosis	multisensory information fusion	
DBM	(C. Li et al., 2015)	Gearbox fault diagnosis	Able to learn good generative models	Training may be slower than DBN.
	(Jia et al., 2016)	Fault diagnosis for rotating machinery	Retains much of desired data present in DBN	Cost the joint optimization of Parameters impractical for large datasets.
DBN	(AlThobiani & Ball, 2014)	Diagnosis fault of reciprocating compressor valves	Work well for single dimensional data.	Training may be very low and inefficient
	(Ma et al., 2016)	Assessment of bearing degradation	steadily achieve high performance over raw signal without too much data preparation	
CNN	(Lee, Yoo, Kim, Lee, & Hong, 2019)	Bearing fault diagnosis	Good for multi-dimensional data	High computation
	(Kaplan, Kaya, Kuncan, Mi̇naz, & Ertunç, 2020)	Bearing fault diagnosis	improved feature extraction	Requires more training time

Technique	Ref.	Application	Strengths	Limitations
	(R. Liu et al., 2016)	Diagnosis of fault approach for electric machine	High performance on local feature extraction	
	(Babu et al., 2016)	Estimation of remaining useful life	Easy to implement	
	(Ding & He, 2017)	Spindle bearing fault diagnosis	multiscale feature learning	High computation
	(Guo et al., 2016)	Bearing fault diagnosis		Requires more training time

METHODOLOGY

In this section, the proposed methodology is discussed in detail to prove the effectiveness of the proposed MHSS algorithm for fault diagnose. A data set of six gearboxes was recoded under a variant environment having different rotation speed. The framework of the proposed methodology is shown in Figure 4.

Data Set

Dataset used in this paper contains the vibration dataset recorded using SpectraQuest's Gearbox Fault Diagnostics Simulator(ZhiQiang Chen, Li, & Sanchez, 2015). Dataset is collected with the aid of four variant sensors. These sensors were placed at four different locations. Data set has been gathered under different load percentage and the range was from zero to 90 percent load. Data set has been gathered in the following two scenarios. Broken Tooth Condition and Healthy condition.

Preprocessing

The Running time span of every signal was 0.5 s, a total of 120 fragments were extracted from the original signal for every condition. Every sample was then preprocessed using CAE. Therefore, each type of gearbox fault had 120 records to get a balanced dataset. There are a total number of 720 samples for six variant health conditions. The dimension size of the total sample was 6140. Hence the dimension of fault gearbox was 6140×720. In the output layer, the numbers of the neuron are six, showing six different health states. Fine-tuning process and Iteration level of each hidden layer is set to 100 for both. Fine-tuning process, the value 100 is set for the first layer. The corresponding learning rates were 0.01 and 0.1. A subset 30% of the total samples was used for testing, whereas all of the reaming datasets were used for training.

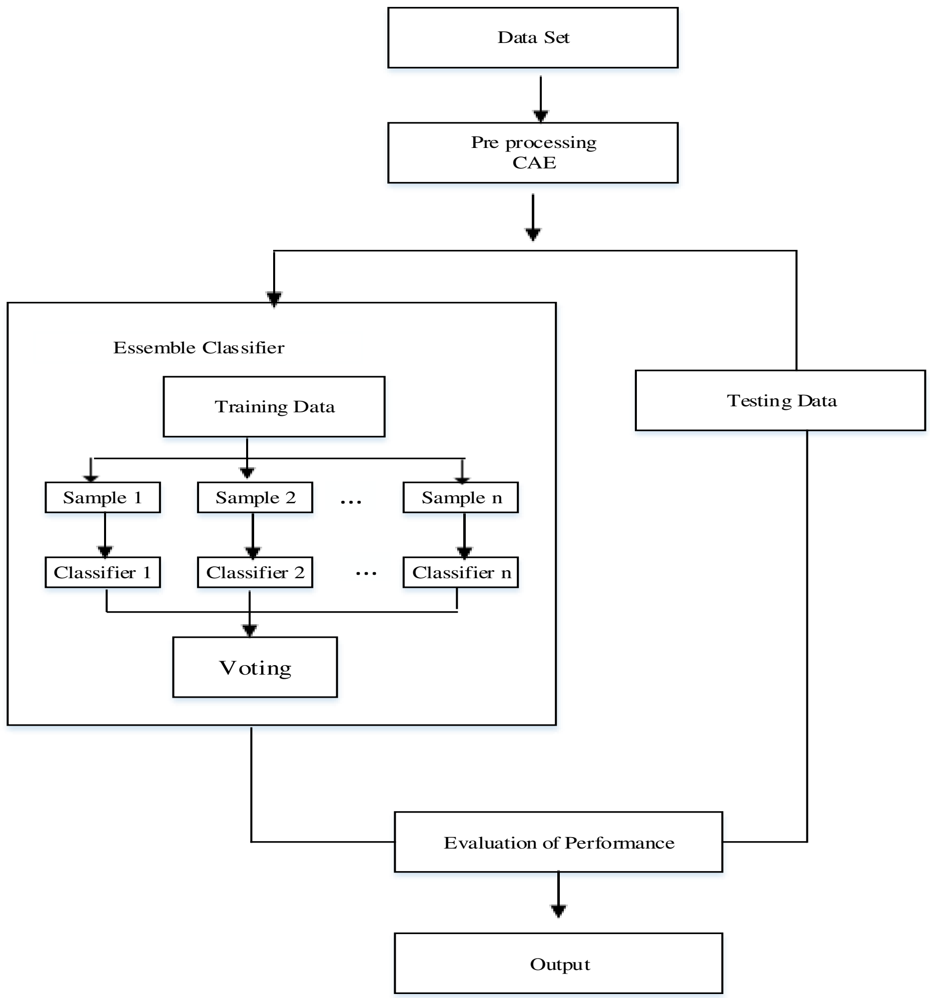

Figure 4. Framework of Proposed Methodology

Unsupervised Bagging Ensemble Classifier

In proposed MHSS model, an ensemble classifier was utilized using bagging model. The extracted features from the dataset are input to ensemble technique to form a set of classifiers. Using the newly constructed set of classifiers, up-to-date data points are classified by measuring vote of their predictions, resulting in the improvement of predictive performance for fault diagnosis. Figure 5 demonstrate the block diagram of bagging ensemble classifier model.

Table 2. Classification of fault process

Faulty Pattern	Location of Fault	Route Transmission
Normal	Normal	5 speed
G22	Gear tooth broken	2 speed
G20	Gear tooth peeling off	5 speed
B	Ring Fault	5 speed
G22 and B	Gear tooth broken + Ring	5 speed

	fault	
G20 and B	Gear tooth peeling off + Ring fault	2 speed

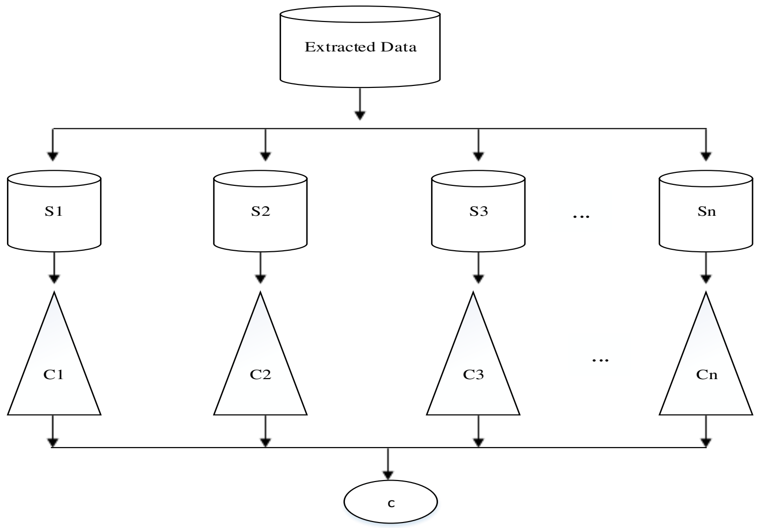

Figure 5: Unsupervised bagging ensemble classifier

As demonstrated in Figure 5, the features extracted using machine learning technique forms the input layer to the proposed model. Several extracted samples ‘s1, s2, s3…, sn’ in the datasets are preprocessed into feature ‘SFE1, SFE2, SFE3 …, SFEn’. Several classifiers ‘C1, C2, Cn’ are created with many samples. Lastly, a blend classifier named ‘C′ is acquired. With the motivation of refining the accuracy of fault diagnosis. The proposed Bagging Ensemble Classifier processes obtained samples (i.e., features) concurrently. Every sample gets an equivalent weight. After that, the classifier is trained for each sample in the bagging ensemble model. Every sample feature is exploited to train the classifier and the final decision is made by using the vote of each component.

Autoencoders (AE)

The idea of AE was presented by Lecun (Rolfe & LeCun, 2013). AE comprise of two parts;

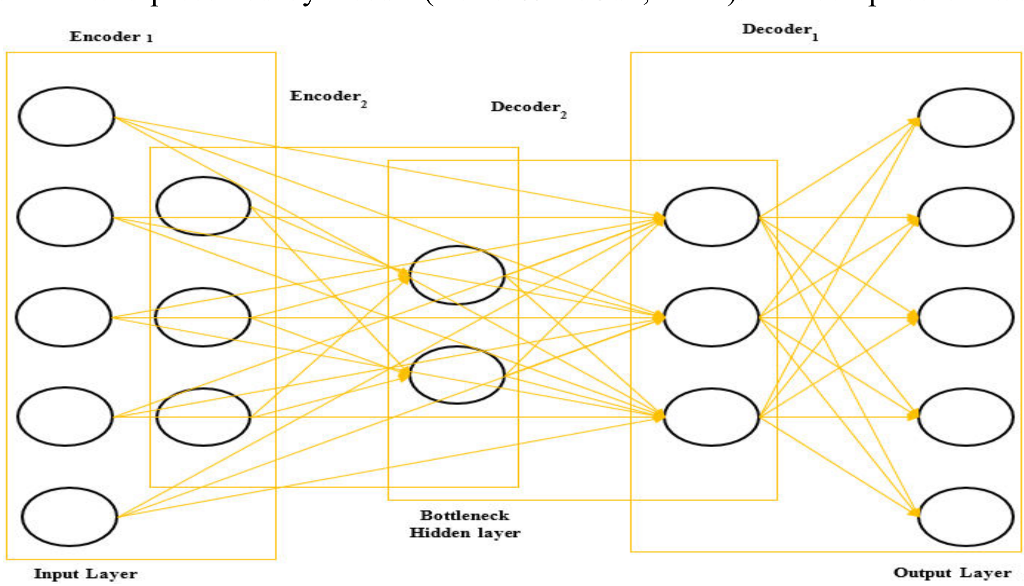

Figure 6: Structure of Stacked Autoencoder

decoder and encoder. Which is designed to learn from input data by reconstructing a new representation of data. The decoder and encoder can be observed as two different functions (Hinton & Zemel, 1994).

$$w = f(z) \,\&\&\, y = g(w) \qquad (2)$$

The function f(z) map data point z to feature space from data space, where g(w) creates re-establishment of data point z by mapping w data space from feature space. In recent AE the two parameters i.e. $w = f(z)$ and $y = g(w)$ are normally a variable function. Where the encoder in (w/z) decoder is (y/w). However y is the reconstruction of z. it is foremost to record that AE does not learn to copy from input z (Bourlard & Kamp, 1988).

Sparse Auto Encoder (SpAE)

To view the internal shape of data, in our model, an extra supervision checks on the hidden layer is applied. A neuron is called active whose output is near to one, while a neuron is called inactive whose output gain is near to zero. The main motivation of SpAE is to minimize the number of inactive neurons.

Given a sample set of training data $y_1, y_2, y_3 ...... y_m$. The mean activation ith hidden layer is $p_i = (1/m)\sum_{j=1}^{m}[h_i(x_j)]$. SpAE impose the restriction $p_i = p$, while p is the activation value of the deserved average. In the hidden layer most of activation gain is near to zero. Thus it was restriction using the following equation.

$$f(x) = a0 + \sum_{j=1}^{s_2}\left(p\log\frac{p}{p_j} + (1-p)\log\frac{1-p}{1-p_i}\right) \qquad (3)$$

Where pj means activation while p is the predetermined average value of activation target of ith hidden neuron in the complete dataset. The sparsity addition checks can be learned from hidden

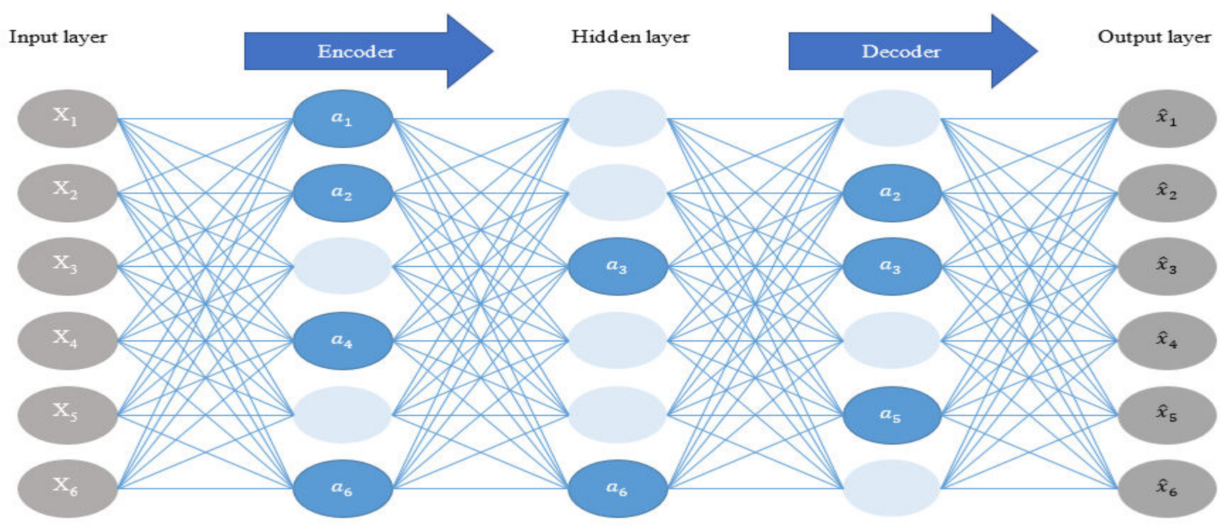

Figure 7. **Structure of Sparse Autoencoder**

layer to be work as sparse representation. Hence, this type of AE is called SPE as shown in Figure 7.

Contractive Autoencoders (CAE)

Some of the real-life applications of neurons involve data labelling for segmentation, denoising input data, detecting outliers, preprocessing, and replacing missing values in data. Many of these applications additionally work with SAEs. In CAE, the zero is not completely replaced but there are little bit changes. In training dataset, the alternative regularization help mapping. CAE is obtaining with the help of the equation 4.

$$CAE = \sum_{y \in h}(R(y, g(f(y))) + \lambda \left\| Xf(y) \right\| F^2 \quad \ldots\ldots (4)$$

It is easy to observe that square of Jacobian corresponds to Frobenius norm to Y weight decay. gf(y) is an identity function in case of the linear function.

Stacked Autoencoders (SAE)

SAE is the combination of many encoders put jointly having several layers of decoding and encoding. This function permits algorithm to adjust more weights, many layers, and the most important more robust. Stacked Denoising Autoencoder (SDA) may be yielding a productive pre-training solution, to train the model via setting the weight deep neural network (DNN). A type of supervised fine-tuning is applied to reduce classification error over the training label dataset as shown in Figure 6.

Training and Testing

After applying CNN variants to extract all important features, for training the assemble model, the feature map of train CNN variants, actual data and its label are used. We used the extracted feature vectors having class labels and actual data to train the assemble model. Assemble trained classifiers calculate the labels of new records in the form of feature vectors. Later, the performance of the proposed model is calculated. In this research, we utilized an ensemble model based on three different deep learning algorithms, namely, RBM, DBN and DBM. The dataset were divided into test and training set with 70-30 ratio.

Classification

The classification phase has a key role in the field MHSS. The training dataset is used to train the model and the test dataset is used to validate the model's results. For the classification purpose a bagging approach is used that combine the three classifiers i.e. RBM, DBN, and DBM

i) Restricted Boltzmann Machine (RBM)

RBM is a useful variant of recurrent neural networks proposed by Geoff Hinton in (C. Li et al., 2015). RBM is very robust for feature learning, classification, filtering, and dimension reduction. RBM is a type of generative stochastic technique called recurrent neural networks. It has a probabilistic element i.e. neuron, which is used to make up the whole network. RBM is two-layer NN shaping a multidirectional graph, which consists of two classes. The hidden unit d and visible unit b with a restriction that there is no communication between the hidden layer and visible layer and there is no communication between the group with node.

ii) Deep Belief Network

Deep belief network (DBN) is a type of deep learning in which many RBM are combined. Figure 8 illustrate a generic deep belief network (Jia et al., 2016). The process of training is performed in a greedy approach having weight fine-tuning to extract hierarchical feature from given data. The purpose of DBN creation was to delineate a model, which distribute data among hidden, and

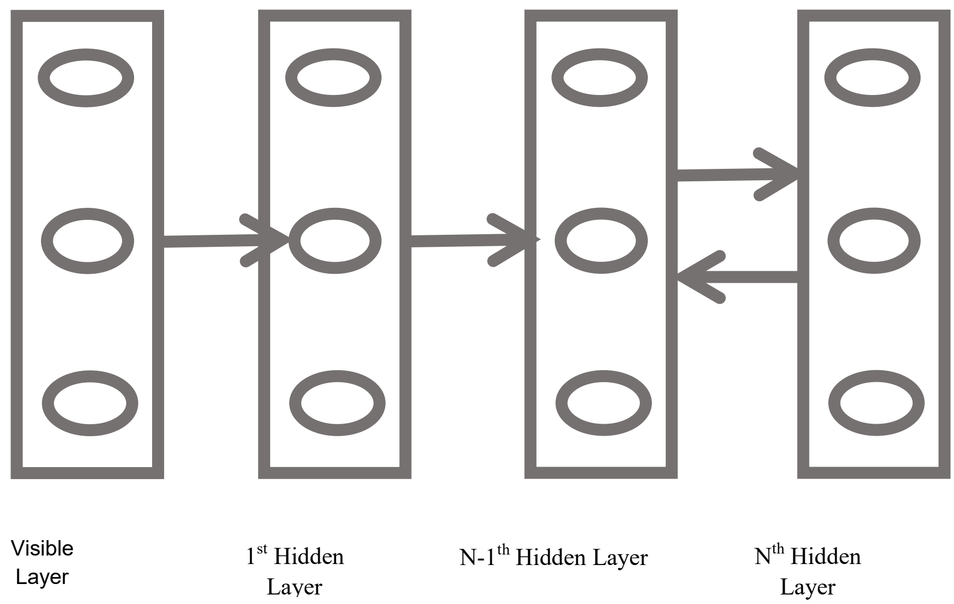

Figure 8: **Deep Belief Network**

input layer such that there is an uninterrupted relationship between upper layer nodes and lower layer nodes. The training process is performed layer-wise in parallel by balancing weight parameter and applying contrastive convergence. Apart from these restriction, probabilities of dispersal DBN is invariable which is also robust to noise transformation(Ma et al., 2016).

iii) Deep Boltzmann Machine

A deep stricture of RBM is combined to form Deep Boltzmann machine (DBM) in which hidden layer are combined in hierarchy shape (C. Li et al., 2015). Allowing RBMs networking restriction no connection among non-neighbour layer and single complete connectivity has been established between subsequent layers. DBM is also called Network of symmetrically grouped stochastic binary layers. The main difference between DBM and DBN is that DBM is completely un-directed graphical algorithm, where DBN is varied undirected/ directed type(Jia et al., 2016).

EXPERIMENTAL RESULTS

This section presents the experimental evaluation of our proposed ensemble model. The mean eigenvalue of variant hidden neuron considering different layers are shown in Figure 9. The output results of the first layer provide the high mean value i.e., 0.745, where the size of hidden neurons in the hidden layer is 451. The mean eigenvalue of the second layer is between 0.805 and 0.76, where the highest eigenvalues is set to 0.381 hidden neurons as shown in Figure 10.

In the third layer, the highest Eigen mean value is 0.966, where the number of neurons in the hidden layer are 201 as shown in Figure 11. As for the last fourth layer, the highest Eigen mean value is larger than 0.93 for the reaming hidden neuron. The eigenvalue of hidden neuron 40, 30 and 20 are 0.980 where the number of neurons in the hidden layers are 191 as shown in Figure 12.

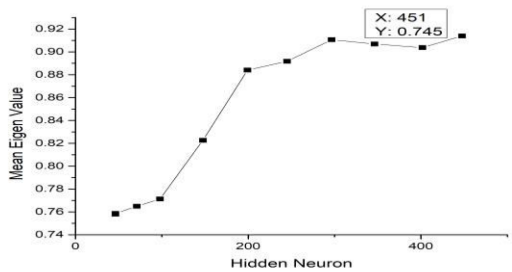

Figure 9. Mean Eigenvalues of Variant Hidden Neurons of First Layer

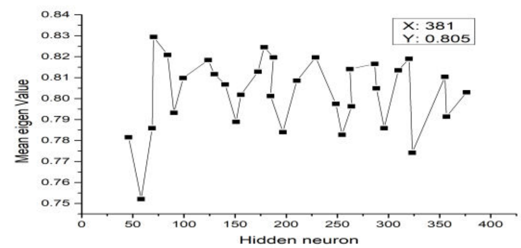

Figure 10. Mean Eigenvalues of Variant Hidden Neurons of Second Layer

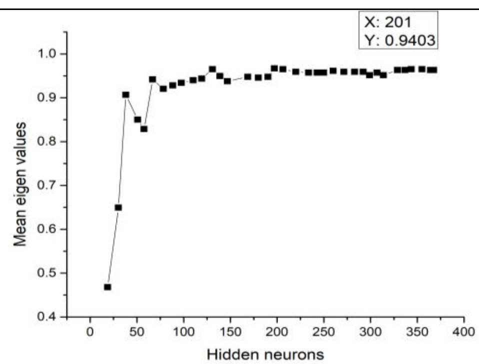

Figure 11. Mean eigenvalues of variant hidden neurons of third layer

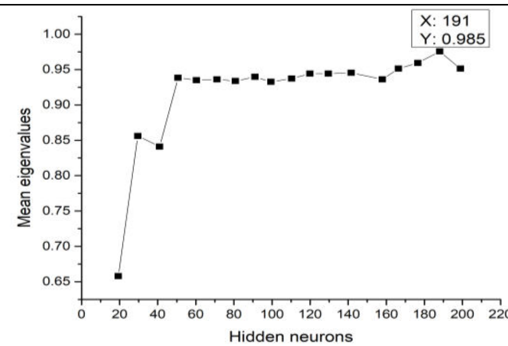

Figure 12. Mean eigenvalues of variant hidden neurons of third layer

The results are shown in Table 3. An increase is observed in Kc value with the increasing hidden layer. When the Kc value is greater than 0.98, hidden neuron at layer fourth gives Kc value 0.985 as the result show the proposed method consists of four hidden layers. Deep learning based on Bozaltman architecture consist of 6141 at input layer correspond to input data. First layer has 451 neurons, 381 neurons at the second layer, 201 neurons at the third layer, 191 at the fourth layer and 7 neurons at the output layer.

Table 3. Best neurons and value of kc at each layer

Figure 13 demonstrate the training performance of the proposed MHSS method. A decrease in the result is shown for MHSS in mean squared with the increasing hidden layers. The mean squared errors of MHSS algorithm is higher than 50 except for the 1st layer. The results also show that the MHSS has a higher feature learning ability than the traditional Algorithms.

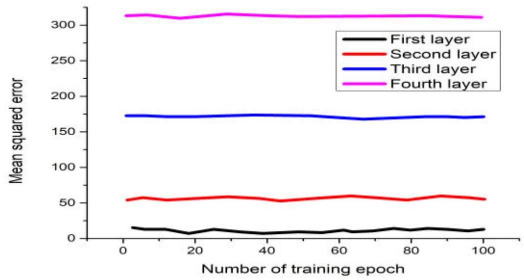

Figure 13. Training performance of SAE and MHSS with four hidden layers

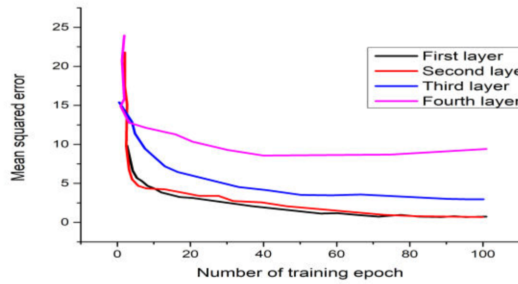

Figure 14. Training performance of proposed MHSS with four hidden layers

To verify the effectiveness of the proposed MHSS, the most commonly utilized techniques are employed in the testing, and testing results are listed in Table 4. The classification accuracy of all algorithms are above 90 percent except SAE. From the above results, it is clear that the performance of SAE and proposed MHSS algorithms are promising than AE, CAE, SAE and SpAE. The Lc value of the MHSS is 0.99, which is better than SAE 0.92. The SpAE method resulted in the lowest Lc value 0.64. It means that the features learnt by MHSS are more distinguishable than those learnt by others techniques. The main features are obtained from the peer techniques are also shown in Figure 14. It should be noted that the two main value features were normalized into [1, 0] along the projecting. direction. The AE, SpAE, and CAE could distinguish “G20,” “Normal,” and “G22 and B” samples from other samples. However, it is clear that the peer algorithms, except for SAE method, especially for the “B” and “G20 and B” samples.

Table 4. Accuracy of testing and clustering performance of the comparative algorithms

Algorithm	Accuracy	Lc
AE	0.9705	0.76
SpAE	0.9898	0.64
CAE	0.9891	0.91
SAE	0.8734	0.92
MHSS	1	0.99

CONCLUSION AND FUTURE DIRECTIONS

Deep learning is a rapidly growing research area having state of art achievement in various application such as speech recognition, object recognition, machine translation and image segmentation. In current modern industrial manufacturing system, MHSS is achieving high popularity because of the widespread availability of sensors having low cost and access to the internet connection. Deep learning architecture gives useful tools to analyze and process these huge amounts of machinery data. In contrast to the traditional machine learning model, deep learning algorithms having superior achievements in the area of machine health monitoring.to perform pre-processing using sparse auto encoder can improve the result accuracy of MHSS. CAE applications are crucially important for MHSS. CNN and their types can handle MHSS feature extraction. Due to the complexity model, hyper-parameter selection is needed to acquire state-of-the-art performances.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Funding: This research was supported by Taif University Researchers Supporting Project number (TURSP-2020/254), Taif **University**, Taif, Saudi Arabia.

Conflicts of interest/Competing interests: The authors declare that we have no conflicts of interests regarding the publication of this article.

Informed consent I consent the journal to review the paper. I inform that the manuscript has not been submitted to other journal for simultaneous consideration. The manuscript has not been published previously. The study is not split up into several parts to increase the quantity of submissions and submitted to various journals or to one journal over time. No data have been fabricated or manipulated (including images) to support my conclusions. No data, text or theories by others are presented as if they were of my own. Proper acknowledgements to other works are provided, and I use no material that is copyrighted. I consent to submit the paper, and I have contributed sufficiently to the scientific work and I am responsible and accountable for the results.

Contributions: Faizan Ullah has produced the idea and converted into writing format. Abdu Salam and Masood Ahmad have performed various steps in simulations. Fasee Ullah has helped in writing the draft of this paper and explained the results. Atif Khan has provided technical support in various sections of this paper. Abdullah Alharbi and Wael Alosaimi have supported the funding for this paper.

References

Al-Sarawi, S., Anbar, M., Abdullah, R., & Al Hawari, A. B. (2020). *Internet of Things market analysis forecasts, 2020–2030.* Paper presented at the 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4).

Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., & Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory, 101*, 102031.

AlThobiani, F., & Ball, A. (2014). An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks. *Expert Systems with Applications, 41*(9), 4113-4122.

Babu, G. S., Zhao, P., & Li, X.-L. (2016). *Deep convolutional neural network based regression approach for estimation of remaining useful life.* Paper presented at the International conference on database systems for advanced applications.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies, 11*(7), 1636.

Bourlard, H., & Kamp, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics, 59*(4-5), 291-294.

Chen, M., Ma, Y., Song, J., Lai, C.-F., & Hu, B. (2016). Smart clothing: Connecting human with clouds and big data for sustainable health monitoring. *Mobile Networks and Applications, 21*(5), 825-845.

Chen, Z., Li, C., & Sanchez, R.-V. (2015). Gearbox fault identification and classification with convolutional neural networks. *Shock and Vibration, 2015.*

Chen, Z., Li, C., & Sánchez, R.-V. (2015). Multi-layer neural network with deep belief network for gearbox fault diagnosis. *Journal of Vibroengineering, 17*(5), 2379-2392.

Chen, Z., & Li, W. (2017). Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Transactions on Instrumentation and Measurement, 66*(7), 1693-1702.

Ding, X., & He, Q. (2017). Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement, 66*(8), 1926-1935.

Galloway, G. S., Catterson, V. M., Fay, T., Robb, A., & Love, C. (2016). Diagnosis of tidal turbine vibration data through deep neural networks.

Gan, M., & Wang, C. (2016). Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems and Signal Processing, 72*, 92-104.

Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement, 93*, 490-502.

He, M., He, D., & Bechhoefer, E. (2016). *Using deep learning based approaches for bearing fault diagnosis with AE sensors.* Paper presented at the Annual conference of the prognostics and health management society.

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in cognitive sciences, 11*(10), 428-434.

Hinton, G. E., & Zemel, R. S. (1994). *Autoencoders, minimum description length and Helmholtz free energy.* Paper presented at the Advances in neural information processing systems.

Hoang, D.-T., & Kang, H.-J. (2019). Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research, 53*, 42-50.

Hussain, T., Muhammad, K., Ullah, A., Cao, Z., Baik, S. W., & de Albuquerque, V. H. C. (2019). Cloud-assisted multiview video summarization using CNN and bidirectional LSTM. *IEEE Transactions on Industrial Informatics, 16*(1), 77-86.

Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing, 72*, 303-315.

Jiang, G.-Q., Xie, P., Wang, X., Chen, M., & He, Q. (2017). Intelligent fault diagnosis of rotary machinery based on unsupervised multiscale representation learning. *Chinese Journal of Mechanical Engineering, 30*(6), 1314-1324.

Kaplan, K., Kaya, Y., Kuncan, M., Minaz, M. R., & Ertunç, H. M. (2020). An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis. *Applied Soft Computing, 87*, 106019.

Komar, M., Yakobchuk, P., Golovko, V., Dorosh, V., & Sachenko, A. (2018). *Deep neural network for image recognition based on the Caffe framework.* Paper presented at the

2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP).

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object recognition with gradient-based learning *Shape, contour and grouping in computer vision* (pp. 319-345): Springer.

Lee, K. M., Yoo, J., Kim, S.-W., Lee, J.-H., & Hong, J. (2019). Autonomic machine learning platform. *International Journal of Information Management, 49*, 491-501.

Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. E. (2015). Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing, 168*, 119-127.

Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. E. (2016). Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mechanical Systems and Signal Processing, 76*, 283-293.

Li, K., & Wang, Q. (2015). *Study on signal recognition and diagnosis for spacecraft based on deep learning method.* Paper presented at the 2015 Prognostics and System Health Management Conference (PHM).

Liu, H., Li, L., & Ma, J. (2016). Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock and Vibration, 2016*.

Liu, R., Meng, G., Yang, B., Sun, C., & Chen, X. (2016). Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. *IEEE Transactions on Industrial Informatics, 13*(3), 1310-1320.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing, 234*, 11-26.

Lu, C., Wang, Z.-Y., Qin, W.-L., & Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing, 130*, 377-388.

Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights, 8*, BII. S31559.

Ma, M., Chen, X., Wang, S., Liu, Y., & Li, W. (2016). *Bearing degradation assessment based on weibull distribution and deep belief network.* Paper presented at the 2016 International Symposium on Flexible Automation (ISFA).

Mao, W., He, J., Li, Y., & Yan, Y. (2017). Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 231*(8), 1560-1578.

Mosterman, P. J. (1997). *Hybrid Dynamic Systems: A hybrid bond graph modeling paradigm and its application in diagnosis*: Vanderbilt University.

Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M., & Dietmayer, K. (2013). Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *Journal of power sources, 239*, 680-688.

O’Donovan, P., Leahy, K., Bruton, K., & O’Sullivan, D. T. (2015). An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *Journal of Big Data, 2*(1), 1-26.

Raina, R., Madhavan, A., & Ng, A. Y. (2009). *Large-scale deep unsupervised learning using graphics processors.* Paper presented at the Proceedings of the 26th annual international conference on machine learning.

Reddy, K. K., Sarkar, S., Venugopalan, V., & Giering, M. (2016). *Anomaly detection and fault disambiguation in large flight data: A multi-modal deep auto-encoder approach.* Paper presented at the Annual Conference of the Prognostics and Health Management Society.

Rolfe, J. T., & LeCun, Y. (2013). Discriminative recurrent sparse auto-encoders. *arXiv preprint arXiv:1301.3775*.

Shao, S.-Y., Sun, W.-J., Yan, R.-Q., Wang, P., & Gao, R. X. (2017). A deep learning approach for fault diagnosis of induction motors in manufacturing. *Chinese Journal of Mechanical Engineering, 30*(6), 1347-1356.

Su, H., & Chong, K. T. (2007). Induction machine condition monitoring using neural network modeling. *IEEE Transactions on Industrial Electronics, 54*(1), 241-249.

Sun, J., Yan, C., & Wen, J. (2017). Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Transactions on Instrumentation and Measurement, 67*(1), 185-195.

Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems, 48*, 157-169.

Tao, J., Liu, Y., & Yang, D. (2016). Bearing fault diagnosis based on deep belief network and multisensor information fusion. *Shock and Vibration, 2016*.

Tao, S., Zhang, T., Yang, J., Wang, X., & Lu, W. (2015). *Bearing fault diagnosis method based on stacked autoencoder and softmax regression.* Paper presented at the 2015 34th Chinese Control Conference (CCC).

Thirukovalluru, R., Dixit, S., Sevakula, R. K., Verma, N. K., & Salour, A. (2016). *Generating feature sets for fault diagnosis using denoising stacked auto-encoder.* Paper presented at the 2016 IEEE International Conference on Prognostics and Health Management (ICPHM).

Tseng, A., Chen, Y., Kao, Y., & Lin, T. (2016). Deep learning for ransomware detection. *IEICE Tech. Rep., 116*(282), 87-92.

Wang, L., Zhao, X., Pei, J., & Tang, G. (2016). Transformer fault diagnosis using continuous sparse autoencoder. *SpringerPlus, 5*(1), 1-13.

Yan, W., & Yu, L. (2019). On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach. *arXiv preprint arXiv:1908.09238*.

Yin, S., Li, X., Gao, H., & Kaynak, O. (2014). Data-based techniques focused on modern industry: An overview. *IEEE Transactions on Industrial Electronics, 62*(1), 657-667.

Yoon, A. S., Lee, T., Lim, Y., Jung, D., Kang, P., Kim, D., . . . Choi, Y. (2017). Semi-supervised learning with deep generative models for asset failure prediction. *arXiv preprint arXiv:1709.00845*.

Zhu, J., Chen, N., & Peng, W. (2018). Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Transactions on Industrial Electronics, 66*(4), 3208-3216.