Journal of Intelligent Manufacturing (2023) 34:1427–1444
https://doi.org/10.1007/s10845-021-01862-4

Autoencoder-based anomaly detection of industrial robot arm using stethoscope based internal sound sensor

Huitaek Yun[1,2] · Hanjun Kim[1] · Young Hun Jeong[3] · Martin B. G. Jun[1,2]

Received: 1 September 2020 / Accepted: 11 October 2021 / Published online: 2 December 2021

Abstract

Sound and vibration analysis are prominent tools for machine health diagnosis. Especially, neural network (NN) strategies have focused on finding complex and nonlinear relationships between the sensor signal and the machine status to detect machine faults. However, it is difficult to collect enough amount of fault data as much as normal status data for training general NN models. To resolve the issue, this paper proposes the autoencoder-based anomaly detection framework for industrial robot arms using an internal sound sensor. The autoencoder uses signals in the normal state of the robots for training the model. It reconstructs the input signals as output, and anomalous states are found from high reconstruction error. Two stethoscopes were attached to the surface of the robot joint as sensors, and the sounds were recorded by USB microphone attached to the outlet of the stethoscopes. Features were extracted from STFT spectrogram images of the gathered sound, then used to train and test an autoencoder model. The reconstruction errors of the autoencoder were compared to distinguish the abnormal status from normal one. The experimental results suggest that the stethoscopes prevent the interference of noise, and the collected sound signals can be utilized for detecting machine anomalies.

Keywords Sound spectrogram · Autoencoder · Neural network · Stethoscope · Industrial robot arm

Introduction

Unexpected performance degradation, malfunction, and even breakdowns of a machine on shop floors may lead to severe deterioration in productivity, or even worse, dangerous accidents. To avoid such cases, a wide range of studies on timely and precise diagnostics of machines and its components has been conducted (Jardine et al., 2006). In general, there are two different methodologies in identification and prediction of the machine conditions and their dynamic behavior, which are the physics-based and the data-driven methods (Heng et al., 2009). Although their processing strategies are quite different, both methods utilize various sensor signals such as torque, vibration, sound emission, and currents (Safizadeh & Latifi, 2014). The physics-based approach takes a variety of static and dynamic parameters into considerations to develop a precise model. Therefore, it demands comprehensive knowledge about physical attributes of the subjects (Bittencourt & Gunnarsson, 2012; Cong et al., 2013; Khan & Yairi, 2018). On the other hand, the data-driven methods emphasize feature extraction, which accounts for deriving compressed and meaningful information from raw sensor signals (Al-Ghamd & Mba, 2006; Dohnal & Sekhar, 2014; Immovilli et al., 2010). In particular, recent advances in sensors and computing technologies encourage a huge amount of sensor data to participate in establishing accurate models through the data-driven methods.

The feature extraction enables to manage high-dimensional data efficiently. Features for effective machine health diagnosis should extract the interesting machine conditions. Some statistical characteristics of time domain signals such as root-mean-square (RMS) and kurtosis have been used to capture the moment before breakdown from acceptable conditions (Martin & Honarvar, 1995; Safizadeh & Latifi, 2014). For non-stationary system analysis, the feature extraction processes are expanded into frequency

Huitaek Yun and Hanjun Kim have contributed equally to this work.

✉ Martin B. G. Jun
mbgjun@purdue.edu

[1] School of Mechanical Engineering, Purdue University, West Lafayette, IN 47906, USA

[2] Indiana Manufacturing Competitiveness Center (IN-MaC), Purdue University, West Lafayette, IN 47906, USA

[3] School of Mechanical Engineering, Kyungpook National University, Daegu 41566, Korea

domain by several methods; short-term Fourier transform (STFT) (Wang et al., 2013), wavelet transform (Chebil et al., 2009; Jaber & Bicker, 2016), and Hilbert-Huang transform (Peng et al., 2005; Rai & Mohanty, 2007). Recently, these data analysis strategies have been supported by advances in computing technologies. Especially, progresses in artificial intelligence (AI) have taken the level of the data analysis techniques to higher steps. There are several applications of AI techniques such as principal component analysis (PCA) (Safizadeh & Latifi, 2014), support vector machine (SVM) (Pan et al., 2009), and neural network (NN) for machine monitoring by categorical classification. NN strategies are also applicable to fault detection (Eren et al., 2019; Janssens et al., 2016; Zhang et al., 2018) and prediction of remaining life (Guo et al., 2017).

However, the classification problem requires a large amount of information from anomalous behaviors of a machine. To construct the prediction model, it is necessary to predefine the categories of faults according to the machine status. It is not always possible to denote every sort of anomaly, which may lead to diagnostic failure (or false alarm). Furthermore, it is not easy to collect the data from machines under the anomalous conditions. Imposing artificial defects on machine components allows to collect useful data, however it may induce severe failure in the machine. To resolve the issue, extensive studies on the autoencoder using reconstruction error (RE)-based anomaly detection, which is applicable for outlier discrimination (Oh & Yun, 2018; Shao et al., 2017; Tao et al., 2015), have been carried out to identify the degradation of machine health. Furthermore, other unsupervised neural networks were utilized to improve the robustness of autoencoder. For example, Recurrent Neural Network (RNN) with denoising autoencoders (Liu et al., 2018) or categorial generative adversarial network (Tao et al., 2020) were used to discriminate bearing fault conditions, improving robustness from external disturbance and uncertainties over other unsupervised clustering neural networks.

Another recent consideration on machine monitoring is cost-effectiveness demands of small footprint and low-cost sensors for Internet of Things (IoT). For monitoring machine conditions, sound sensing can be more affordable solution for machine monitoring than force or vibration measurement since the total costs for sensors and signal conditioning devices are not required when using numerous sensors at the same time. Several studies have diagnosed machine condition from acoustic signal using various AI techniques mentioned above to classify sounds (Henriquez et al., 2014). In recent studies of machine anomaly detection using acoustic signals, fault detection of railway point machines made of actuators were conducted using Mel-Frequency Cepstral Coefficients (MFCC) and SVM of microphone sounds. It detected slacked nuts and obstructions by ice or ballast in the machine. Autoencoder trained with sounds from microphone was used to detecting the anomaly of Surface-Mounted Devices (Oh & Yun, 2018). From the study, non-greased line was found from normal one by comparing reconstruction error. Continued study (Park & Yun, 2018) used RNN encoder-decoder to find the anomaly and MFCC was compared to STFT that shows trade-off between calculation time and detection performance. However, sound signal using microphone is susceptible to external noise since microphones collect sounds from all directions. Ambient sounds from other machinery and humans can be mixed to the recorded sounds, making it difficult to find the acoustic location. Placing a spherical array of 32 microphones isolated sounds of milling operation from nearby machines (Shaffer et al., 2018). However, if a sensor is attached to the sound source to catch the internal dynamics, less sensors would be used for acoustic location and ambient noise reduction.

In medical science, stethoscope is utilized to diagnose human health by amplifying sounds of interest from human body (Chamberlain et al., 2015; Gupta et al., 2007; Hamidi et al., 2018; Maglogiannis et al., 2009). Inspired from the auscultation, an internal sound sensor using stethoscope was developed, which can be used in machine health monitoring as a cost-effective internal sound sensor (Yun et al., 2020). According to our previous study, the sound signals from the target sites can be obtained and the influence of external noises can be suppressed by attaching the sensors to machine surfaces. However, it is difficult to find any significant approach using classic stethoscope in manufacturing, although the stethoscope has some fair attributes of cost-effective sound sensor.

The stethoscope-based internal sound sensor targets on manufacturing systems. Especially, industrial robots have been widely used not only for pick-and-place and machine tending but also sophisticated processes like machining, additive manufacturing, surface finishing, welding, and quality measurements (Mikolajczyk, 2012a, 2013; Liu and Zhang 2014). In the monitoring and controlling these processes, machine vision and force sensors integrated with custom robot control have been utilized (Mikolajczyk, 2012b, 2015; Zhu et al., 2020), and the system requires other types of sensors to measure the complexity. The developed acoustic sensor in this paper would propose an alternative way of the robotic applications with its affordability and the property of noise reduction.

In this paper, a framework to identify abnormal behavior of industrial robot arm is proposed using autoencoder neural network and stethoscope-based custom internal sound sensor. The sound signals were collected using two stethoscope-based internal sound sensors at different locations of an industrial robot arm. Spectrograms were derived from raw signals for visualization and feature extraction. Meaningful low frequency information was sorted out from the spectrograms after the system identification of the internal sound

sensor. The autoencoder-based framework was selected as the main structure of the anomaly detection algorithm in this work. The lack of faulty status data was assumed; hence the normal status data were utilized for training the autoencoders. The conditions were set by hanging different weights in the end effector while the excessive weights are assumed as anomalous conditions. The conditions were discriminated by applying thresholds of the autoencoder's reconstruction errors. In addition, it is discovered that as the position of a stethoscope is close to the axis, the success rate of the discrimination is increased.

Chapter 2 describes the concepts of autoencoder-based anomaly detection. Chapter 3 introduces experimental setup of hardware and autoencoder framework. Chapter 4 discusses the results of the framework for external noise reduction and the accuracy of anomaly detection.

Concept and principles

Concept of autoencoder-based anomaly detection

Based upon the preliminary knowledge of autoencoder (Kramer, 1991), the autoencoder NN-based anomaly detection using internal sound signals is schematically depicted in Fig. 1, which can be summarized as follows:

- *Sound signals measurement.* The number of stethoscopes and their attachment sites (equivalently, target sites) must be predefined. Sound signals are measured and collected during robot operation, and they are transformed into spectrograms in frequency domain using short-term Fourier transform (STFT). The spectrograms are compressed into features by filtering, which are spectrogram within the frequency range of interest. The spectrogram which is sound signal magnitude versus frequency is one-dimensional image (i.e. array), and it is extended into two-dimensional image by concatenating in order of time. The features, spectrogram image in this study, are inputs of autoencoder, also they play the role of reference outputs for training the autoencoder. The dimension of input and the depth of the autoencoder must be predefined.
- *Training an autoencoder.* The signals, which are considered as "normal" or "acceptable", are collected to train an autoencoder. To select the structure of autoencoder, changing hyperparameters are performed. Features are extracted from sound spectrograms generating 2D images, and then fed into the autoencoders. For the convenience, the sound signals are recorded longer than the input dimension, then the images are divided into several features. The input dimension is decided for each joint with several numbers ($n = 4, 8, 16, 32$) and the one with minimum loss is accepted. The result is summarized in the "Appendix 2".
- *Testing the trained model and making decision (i.e. detecting anomaly).* The signals under "normal" and "anomalous" conditions are collected for testing. Again, the features are extracted in the same way, then fed forward into the trained autoencoder. Here, different weights are loaded on the end effector of robot. In normal conditions, the weight less than the allowable level of the robot is loaded, whereas heavier loads attached to the robot are regarded as anomalous state in this study. In this regard, feasible thresholds ε are preliminarily set between "normal" and "abnormal" status in each axis. Here, RE is the difference between reconstructed features, which are the output of autoencoder (i.e. the reconstructed spectrogram image) and its corresponding input. After the thresholds are set for each joint, the model is used to detect the anomaly by comparing the RE with its corresponding threshold.

In the section of experimental setup, the details of each step such as feature extractions, autoencoder designs, and threshold selections are described.

Feature extraction using STFT

STFT, which is one of Fourier-related transformations, analyzes the frequency and phase content of any segment of a signal from a time-varying system (Gribonval, 2008), therefore STFT provides the spectrogram, which is spectrum change of a signal with respect to time. STFT can be expressed as follows:

$$V(\omega)|_{t=\tau} = \sum_{n=-\infty}^{\infty} v(n)w(n-\tau)e^{-i\omega n} \tag{1}$$

where v, w, and V_τ are the sound signal at time n, the window function, and discrete Fourier transform of windowed signal centered about time τ, respectively. The spectrogram vector $PSD_\tau(\omega)$ at time τ can be obtained by squaring the magnitude of V, therefore, it corresponds to the power spectral density (PSD) of V as shown in (2).

$$PSD(\omega)|_{t=\tau} = \left|(V(\omega)|_{t=\tau})\right|^2 \tag{2}$$

Sound signal obtained at sampling frequency of f_{sampling}, which is high enough to cover the frequency range of interest $[0, f_{\text{interest}}]$ in monitoring with satisfying Nyquist frequency, are converted into spectrograms using STFT with windowing at every second, therefore the frequency resolution of the spectrograms is 1 Hz. The features are extracted by following procedure:

- The power spectral densities (PSDs) are filtered up to f_{interest} Hz. In this step, the first f_{interest} points in each

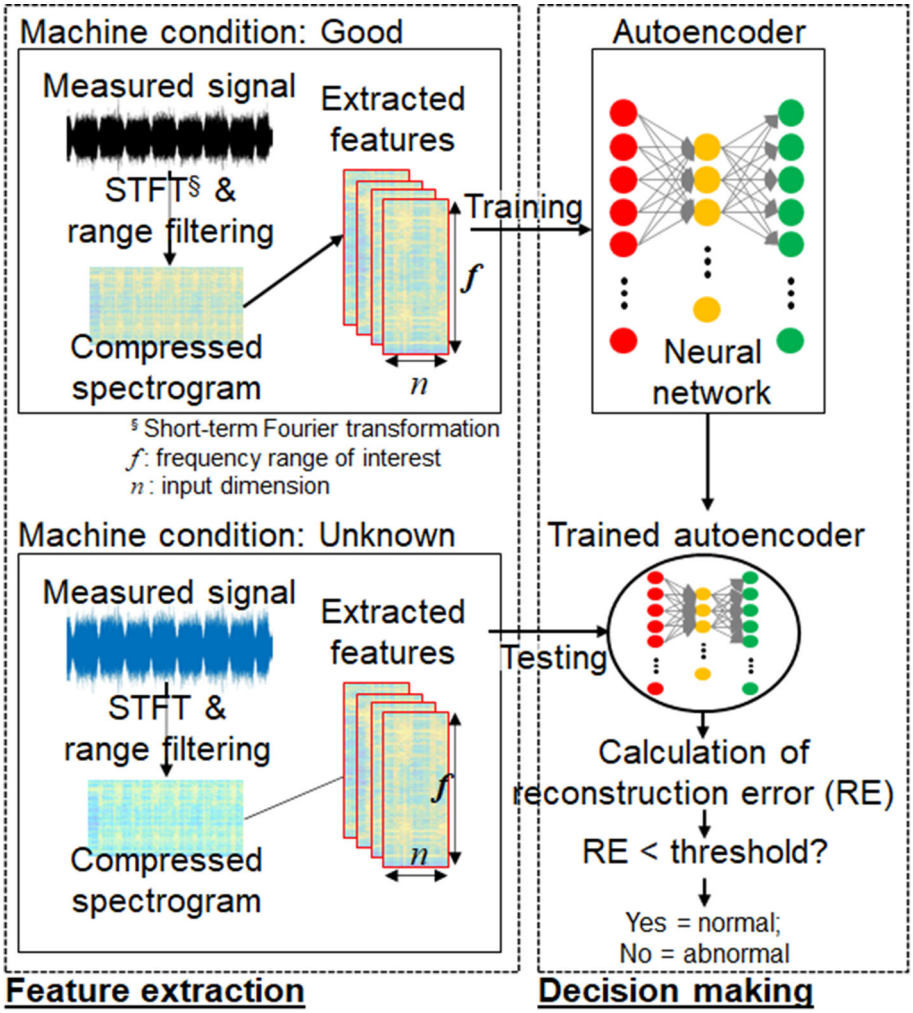


Fig. 1 Overview of autoencoder-based anomaly detection framework using internal sound sensor signals

PSD vector are selected for achieving the bandwidth up to f_{interest} Hz, since the frequency resolution of the raw spectrogram is 1 Hz. After this, every single PSD has a length of $[f_{\text{interest}} + 1]$.

- The filtered PSDs are normalized to have values within 0 and 1 using (3), in which $PSD^k(i)$ is i^{th} component in k^{th} PSD of spectrogram. This step is required since a sigmoid activation function is used in the output layer of the autoencoder, which is bounded between 0 and 1.

$$PSD^k_{norm}(i) = \frac{PSD^k(i) - min\left(PSD^k\right)}{max\left(PSD^k\right) - min\left(PSD^k\right)} \tag{3}$$

- After normalization, successive PSDs are concatenated horizontally along the time axis. Since training autoencoder with one-dimensional (1D) PSDs may lead to confusion between normal and abnormal conditions, the number of n PSDs are combined into two-dimensional (2D) PSD sequences to construct a 2D input for a certain time interval for the autoencoder. The concatenation of 1D PSDs results in 2D images with a size of n. Therefore, k^{th} 2D feature $F(k)$ can be configured through (4), where m is the amount of overlap. The k^{th} feature starts from k^{th} PSD to make features overlap with each other. The overlaps are assigned among the features to cover the entire range since it is difficult to make the recordings synchronized with the robot arm operations.

$$F(k) = \left[PSD^{mk}_{norm} \middle| PSD^{mk+1}_{norm} \middle| \ldots | PSD^{mk+n-1}_{norm}\right] \tag{4}$$

Autoencoder neural network

An autoencoder is one of semi-supervised learning based NN architectures, which is a popular approach in image reconstruction and denoising (Vincent et al., 2008). The term "semi-supervised" comes from the aspect that an autoencoder makes use of inputs as targets for reference. Encoding stage finds a compressed original input (PSD), and decoding stage produces an output that mimics the original input. In this framework, the general type of stacked autoencoder is used.

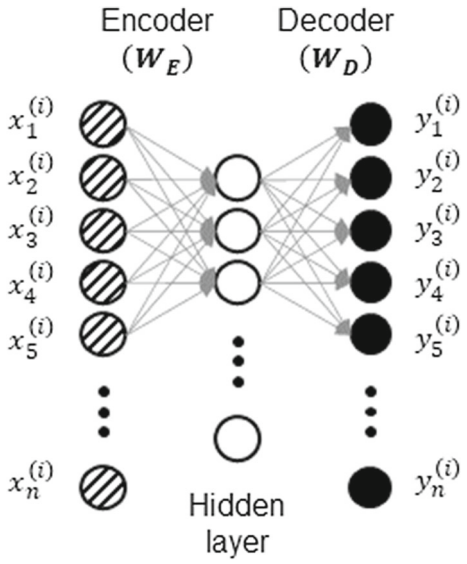


Fig. 2 Schematic representation of single layer autoencoder

It is assumed that there is a sequence composed of n-dimensional vectors $X^{(i)}$ s, $\{X^{(1)}, X^{(2)}, X^{(3)}, \cdots\}$, where $X^{(i)} \in R^n$. The autoencoder tries to adapt output $Y^{(i)}$ close to the original input $X^{(i)}$. Figure 2 portrays the simplest case of a single layer autoencoder. The feedforward process is as follows:

$$(HiddenLayer)Z^{(i)} = \sigma_{enc}\left(W_E X^{(i)}\right)$$
$$(OutputLayer)Y^{(i)} = \sigma_{dec}\left(W_D Z^{(i)}\right) \quad (5)$$

where W_E and W_D are the weight arrays of encoder and decoder, σ is the activation function, and Z is the output of hidden layer, respectively. The error between $X^{(i)}$ and $Y^{(i)}$ is named as reconstruction error (RE) which is also represented as a loss function $Loss\left(x^{(i)}, y^{(i)}\right)$ of the autoencoder. RE is computed after calculating output layer, and the "learning" yields minimizing the loss so that the reconstructed images resemble the original inputs (i.e. spectrogram images). In this work, the activation functions and a loss function are designed as follows:

$$(Activation)\sigma_{enc}(x) = \begin{cases} 0, x < 0 \\ x, x \geq 0 \end{cases}$$
$$(Activation)\sigma_{dec}(x) = \frac{1}{1+e^{-x}}$$
$$(RE)Loss\left(x^{(i)}, y^{(i)}\right) = \frac{1}{n}\sum_{k=1}^{n}\left(x_k^{(i)} - y_k^{(i)}\right)^2 \quad (6)$$

where the activation functions $\sigma_{enc}(x)$ and $\sigma_{dec}(x)$ are also known as the Rectified Linear Unit (ReLU) and the sigmoid function, respectively. In this study, $\sigma_{enc}(x)$ is used for encoding (data compression), and $\sigma_{dec}(x)$ is employed for decoding (input reconstruction). For the loss function, mean-squared error is assigned.

Next, the network parameters W_E and W_D are updated by back-propagation algorithm. The parameters are adjusted where the loss function defined in Eq. (6) is minimized for all training examples. This framework uses Adaptive moment estimation (Adam) optimizer (Kingma and Ba 2014), which is recommended for faster optimization than other methods such as Momentum optimization or Nesterov Accelerated Gradient (Géron, 2019).

RE has been used for the anomaly detection algorithm (Oh & Yun, 2018; Qi et al., 2014; Zhou & Paffenroth, 2017). To classify anomalous signals from normal signals by REs, the autoencoder should be trained purely with normal signals. After training without abnormal signals, the autoencoder produces larger RE when "unseen" data from abnormal status are fed in as input.

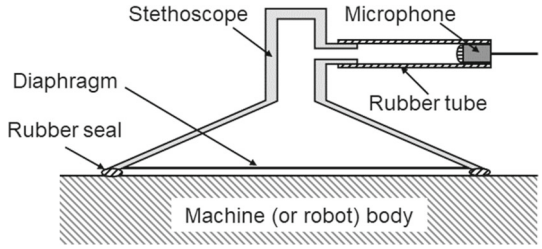


Fig. 3 Structure of internal sound sensor using stethoscope and USB microphone

Experimental setup

Sensor installation and sound recording

In this study, stethoscopes are used to collect internal sounds from a 6-axis industrial robot because stethoscopes can effectively filter out the external sounds owing to its concealed structure. To record sounds, a universal-serial-bus (USB) microphone pickup is connected to the stethoscope via a rubber tube. Figure 3 illustrates the structure of the proposed internal sound sensor. The chest piece with diaphragm is tightly stuck to machine surface by rubber seal to minimize the effect of external sounds. Signal conditioning devices such as amplifier and analog-to-digital converter are not required except a personal computer (PC) that can obtain sounds signal via USB communication. Therefore, the cost of this sensor is around USD 100.00 which is more affordable price than acoustic emission sensors or accelerometers.

Experiments were performed using a compact high-working speed 6-DOF industrial robot (KR6 R700, KUKA)

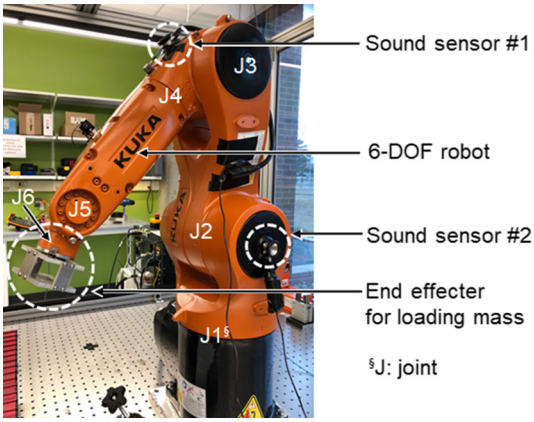


Fig. 4 Experimental system setup using an industrial robot and two sound sensors

Table 1 Angular velocity of each joint at 50% feed override

Joint number	#1	#2	#3	#4	#5	#6
Speed (deg/s)	180	150	180	190.5	194	307.5

with the maximum payload of 13.2 lb (6 kg). For sound signal acquisition, two stethoscopes (Classic II, 3 M Littmann) with the frequency range up to 500 Hz were attached on the wrist and the base of the robot. The sound signals were transferred to a desktop PC by connecting USB microphones (K503, FIFINE) to the outlet of the stethoscopes as shown in Fig. 4. The sound signals from the microphones were collected using sound acquisition software at sampling frequency of 48 kHz and saved as Windows Media Video (WMV, Microsoft) format. The detailed development and identification of frequency range are described in a previous study (Yun et al., 2020).

The anomaly was introduced using mass objects attached to the end effector of the robot arm. The normal condition for the payload were set 0, 1.25, and 2.5 lb while the masses of 5.0, 7.5, 10.0, and 12.5 lb, which are not recommended to cover the full movement range, were considered abnormal. After the hardware setup, the robot arm was programmed to operate six distinct single joint rotations with constant speed (50% feed override), which accounts for the self-diagnosis stage of robot arm. Table 1 describes the angular speed of each joint at the feed override.

Training autoencoder neural network

As shown in Fig. 5, the signals gathered under "normal" conditions (0–2.5 lb) without external noise were used to train and validate the autoencoders. The training and validation data sets were gathered at different days in period #1. Also, the signals under "normal" condition with external noises were collected to investigate the susceptibility of decision-making algorithm to external noise. Noise sounds were collected from real machine shop environment with hammering, riveting, and machining. It was played back at a sound level of 80 dB using an analog speaker while the sensors recorded the sounds from the robot under normal conditions. In period #2, the signals collected under both "normal" and "abnormal" conditions were used to evaluate the detection performances of autoencoders. The data were collected in both in calm and noisy atmospheres to investigate the influence of environmental noise on the performance of autoencoders. Since the robot has six joints as the monitoring targets, six different autoencoders based on two internal sound sensors attached on the wrist and the base of the robot are required in total for each noise condition.

The feature groups were configured for training and validating dataset from the first preparation step. Since each joint motion has different angular velocity for a given speed override, the number of features for each autoencoder are also different from a single trajectory. The autoencoders were trained with the help of Adam optimizer (Kingma and Ba 2014) at a learning rate of 0.0005. The total number of epochs and the batch size were fixed at 100 and 50, respectively. The sizes of training and validation datasets were controlled by input dimension $256 \times n$ points, as shown in "Appendix 1". Keras and Tensorflow packages (version 1.13) in Python 3.6 were used to implement the autoencoder.

For inputs of the autoencoder, the sounds were obtained at sampling frequency of 48 kHz and then converted into spectra by 48,000-point STFT with 50% overlapped Hann window at every second. The frequency resolution of the spectrum is 1 Hz. Then, the first 256 points (input height f) in each spectrum vector were selected for achieving the bandwidth up to 255 Hz. After normalization, the successive spectra are concatenated horizontally along the time axis. Particularly, four candidate autoencoder NN structures with different input width n (4, 8, 16, and 32 widths) were prepared for input. Therefore, the size of input is 1-D vector of $f \times n$ length. Autoencoder was learned with training input data sets, and then generalization of the trained autoencoder was evaluated through validation step. Next, the structure of autoencoder was established by changing the input dimension and the depth of hidden layers. In addition, the training and validation losses of autoencoders were compared to select appropriate structure. Therefore, the number of features was also changed when the dimension of input was modified. Especially, the variations of input dimensions and the depths of hidden layers in this study are specified as in Fig. 6. Four diversities were imposed on the input dimension (here, 4, 8, 16, and 32), and 3 diversities in the depth of hidden layer (i.e. 1, 3, and 5). In encoding phase, the dimension of each layer decreased quarterly at compression process. On the other hand, the dimension was multiplied by four in every layer at decoding phase to recover the same dimension as original input fea-

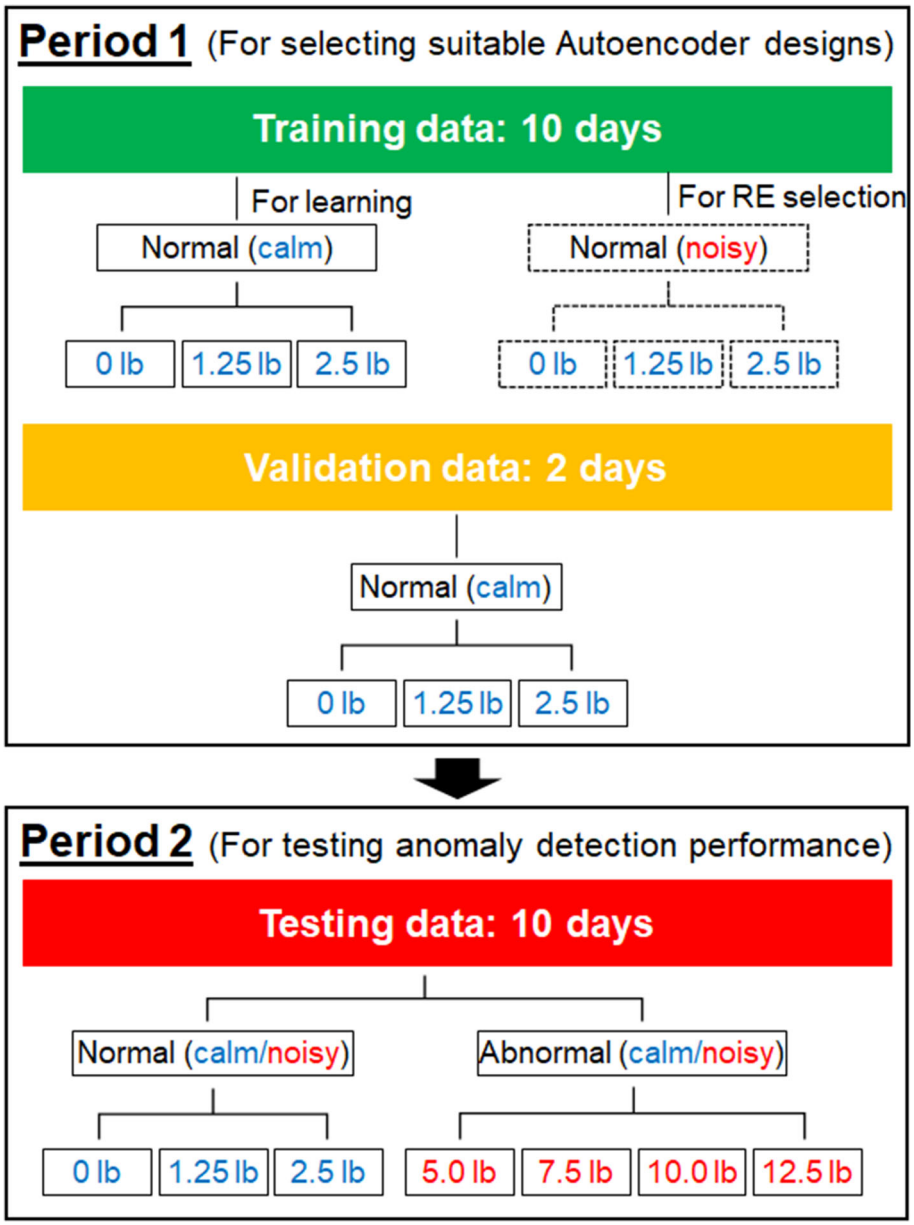


Fig. 5 Data collection for training, validation, and testing

ture. The output of autoencoder is the reconstructed version of input data. In this study, the autoencoder was trained with input data below the load limit of the robot. When the load is over the design limit, it is considered as anomaly state, and the reconstructed output will be more different from input data than the normal state, where the reconstruction error (RE) is defined in Eq. (6).

The REs of four candidate autoencoder NN structures were computed and compared one another at the ends of training and validation. To select proper autoencoder structures among given candidates, some guidelines were established as summarized below:

- Higher reconstruction quality; the smaller final loss of training and validation is preferred.
- Less overfitting; The difference between training and validation losses needs to be as small as possible. In other words, the rate of two losses closer to 1 is preferred.
- Lighter computation load; if restrictions guidelines 1) and 2) are similarly satisfied among different structures, the structure with lower depth and dimension is faster.

The results are summarized in "Appendix 2" to choose the best autoencoder structures according to the three guidelines with respect to axis number. The selected designs are highlighted with bold, and underlined characters in "Appendix 2". Table 2 shows the finally selected model parameters from

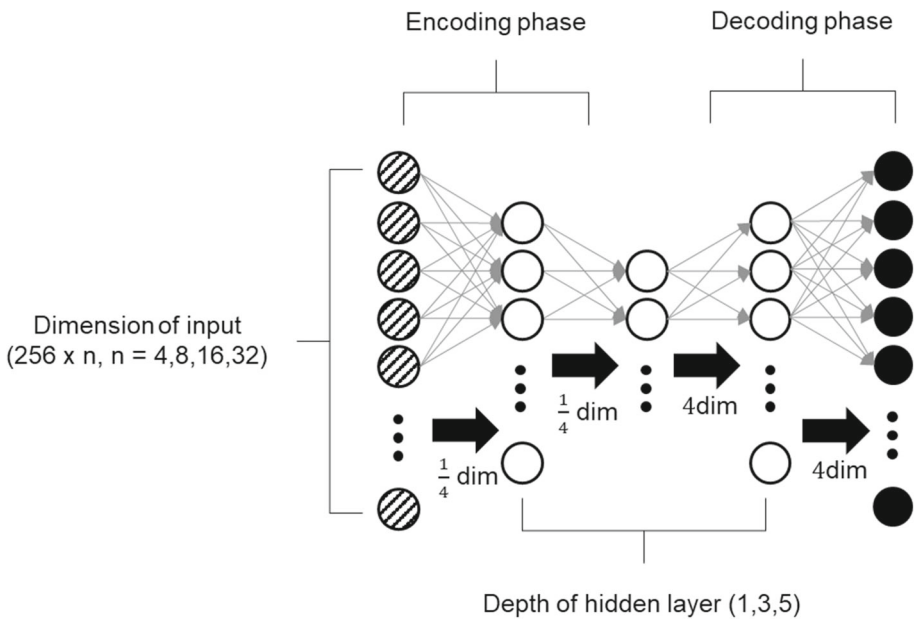


Fig. 6 The structure of autoencoder with tuned parameters

Table 2 Hyperparameters for autoencoder

Sensor	Parameters	Joint number					
		#1	#2	#3	#4	#5	#6
#1 (Wrist)	STFT input feature size	4 × 256	8 × 256	8 × 256	16 × 256	4 × 256	32 × 256
	Depth of hidden layer	5	3	5	5	3	5
#2 (Base)	STFT input feature size	8 × 256	4 × 256	8 × 256	4 × 256	4 × 256	4 × 256
	Depth of hidden layer	3	5	5	1	3	3

Table 3 Training time of autoencoder for different parameters

Input feature size	Autoencoder training time (sec)		
	Hidden Layer: 1	Hidden Layer: 3	Hidden Layer: 5
4 × 256	11.47	13.77	15.18
8 × 256	27.17	35.24	35.90
16 × 256	93.86	113.85	120.37
32 × 256	386.64	527.20	540.99

the guideline. Table 3 lists training time of autoencoder at personal computer (PC) with AMD Ryzen 3600 CPU and Nvidia GTX 1060 graphic card for GPU acceleration in Tensorflow, showing variation of time due to number of weights by hyperparameters. The time is the average of measuring calculation time for three times, and the maximum variation is within 3.1%.

Results and discussions

The distributions of REs in autoencoders under two different noise conditions were investigated to determine proper thresholds for accurate anomaly detection. In this study, RE was used as the critical measure for decision, after ensuring the insusceptibility of RE on external noise. Figure 7 illustrates the average of REs at different loads (0–12.5 lb) for six joints and two sensors. REs obtained under "normal" conditions with external noise, and they were compared with REs under normal and calm conditions. It can be seen that the difference varies from − 14.0 to 38.9%. However, it is difficult to find any tendency in RE difference with respect to external noise and thus the effect of external noise on RE can be ignorable. The difference of RE amounts between joint is from the variance of spectrogram intensity. As a result, the maximum RE value for training data sets of each joint under "normal" conditions without external noise were designated as RE threshold of the autoencoder for each joint. From the comparisons, the feasible thresholds are described in "Appendix 3". The thresholds from training results were utilized for detecting anomalies in testing data sets. The threshold was set by increasing 4th digit higher than the maximum RE.

REs of features from anomalous status are larger than the REs of normal status features, which can be discriminated by the thresholds. Apparent reconstruction failures are noticed in abnormal status feature while the feature from normal status is reconstructed as shown in Fig. 8, which shows the

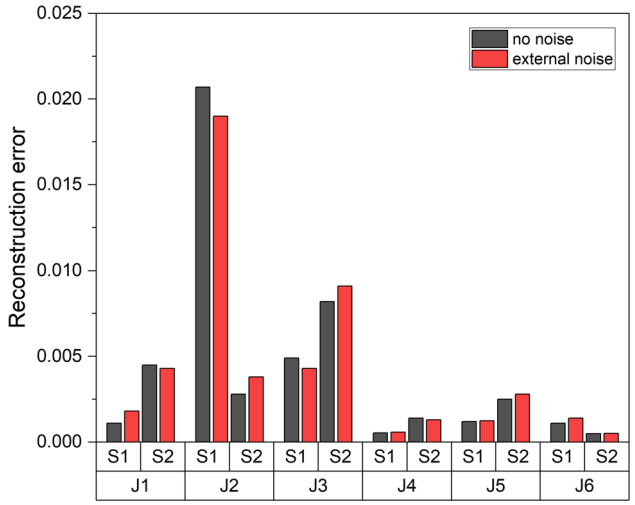


Fig. 7 Average reconstruction error after training the autoencoder (J1–J6: joint; S1–S2: Stethoscopes)

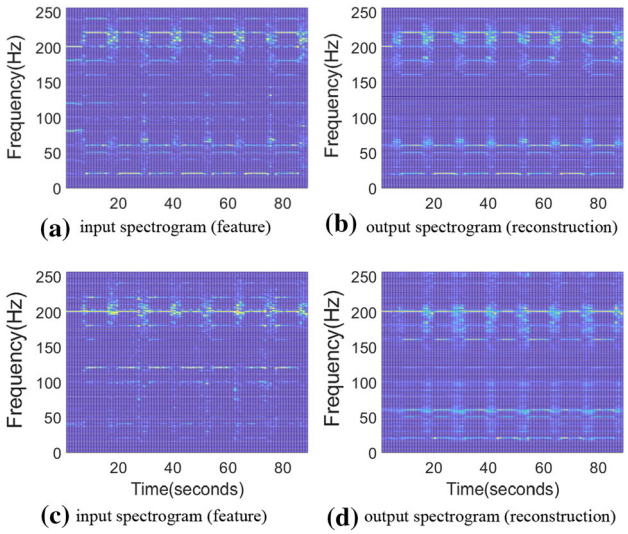


Fig. 8 Reconstruction results of the autoencoder for J1 joint with different load conditions

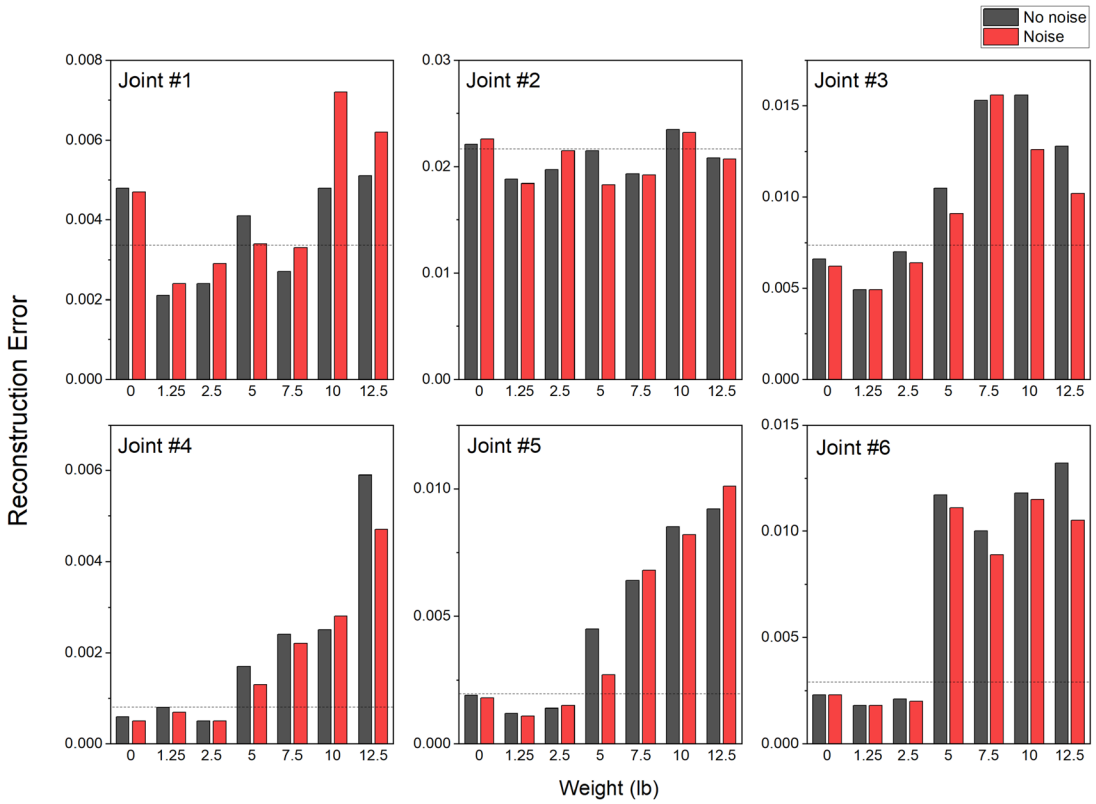
comparison of feature reconstruction between normal and abnormal conditions. Figure 8a, b show an input spectrogram for the joint #1 and its reconstructed one under normal condition when the mass attached to the robot was 0 lb (i.e. normal condition), while Fig. 8c, d are the spectrograms when the load was 12.5 lb. The reconstructed spectrogram reproduced the original one well, however some peaks (i.e. brighter colored dots) were not presented, which meant that the trained autoencoder represented the systematic characteristics of the target robot well. RE of Fig. 8b for (a) was 0.0031. Meanwhile, the reconstructed spectrogram (Fig. 8d) for the robot operation with overload could not reproduce its original one. In detail, the dynamic behavior at around 120 Hz could not be reproduced, false peaks ranged between 20 and 100 Hz were created. The RE value of Fig. 8d for Fig. 8c was 0.0102. This result shows that an autoencoder can reconstruct learned inputs only, not unseen inputs.

As described earlier in the whole features from the period #2 in Fig. 5 were fed into the trained autoencoders to verify the capability of anomaly detection. The average RE values were computed every day for 10 days. The threshold values are indicated as dotted lines on each graph. Figure 9a shows that the sensor #1 at wrist failed to discriminate anomaly at joint #1 and #2. However, the anomaly was successfully detected by the autoencoders for joint #3-#6, which is closer to the sensor location than other two joints. Similarly, it was noticed that the sensor #2 (base) provided fair separation of normal and abnormal status of the joint #3-#6 as depicted in Fig. 9b. As mentioned before, the overall results between calm and noisy environment also show robustness from external noise. From Fig. 9, it is noticed that in each axis the anomalies (excessive load conditions) are discriminable by at least one sensor location. It is also shown that the single threshold value in each axis can be applied for detection regardless of noise.

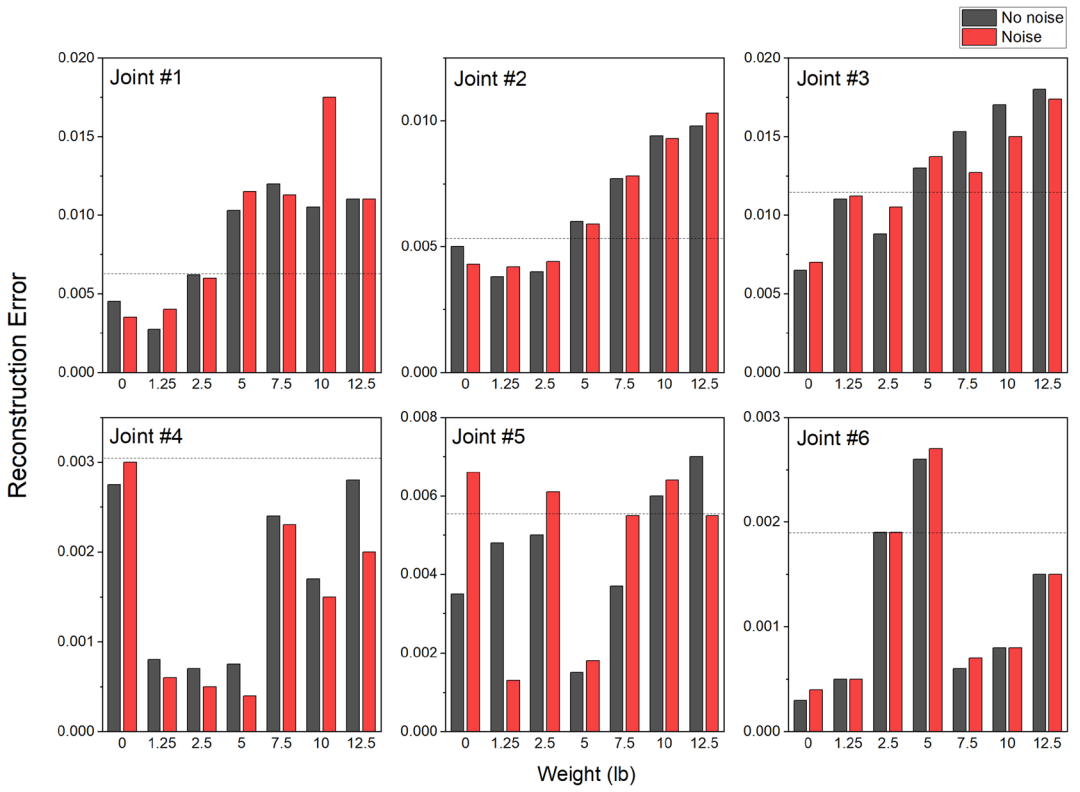
There were three different "normal" and four different "anomalous" load conditions in the experiments each day. In addition, the experiments were performed twice at calm and noisy environments, respectively. Therefore, 60 ($3 \times 2 \times 10$) normal RE groups and 80 ($4 \times 2 \times 10$) anomalous RE groups were generated in period #2 for testing. The success rates of status estimation in each group are summarized in "Appendix 4". Table 4 summarizes the "Appendix 4". The underlined numbers indicate the cases are feasible with success rate at both normal and abnormal conditions above 85%. As described above, the sensor #1 was installed between joint #3 and #4, and sensor #2 was at the joint #2 and near joint #1. The result shows that accuracy of the framework depends on the location of sensor. The accuracy of anomaly detection was increased as the sensor location was closer to joint location.

In summary, the feasibility of the proposed autoencoder-based internal sound signal was analyzed for anomaly detection of 6-axis industrial robot arm. The results provide the discussion results as follows.

Firstly, the estimation results between normal and abnormal states imply that the predetermined thresholds from past machine experiences can be utilized for future anomaly detections. As stated earlier, the study assumes that fault simulation data is harder to collect than normal state data, hence it is required to train the autoencoder only by normal state data. In real application, the RE values can be gathered in a long term so that it is expected to have Gaussian probability distribution. The trend of RE values will be conducted further on case studies in real systems. After obtaining the distribution, unlikely high RE values can be employed as the



(a) RE distributions: sensor #1 at wrist



(b) RE distributions: sensor #2 at base

Fig. 9 Comparisons of daily RE distributions with respect to different load conditions: **a** mic #1, and **b** mic #2

Table 4 Feasibility of the autoencoder framework

Success rate (%)	Sensor #1 (Wrist)		Sensor #2 (Base)	
	Normal	Abnormal	Normal	Abnormal
Joint #1	66.0	65.0	93.3	100.0
Joint #2	86.7	75.0	95.0	98.8
Joint #3	91.7	100.0	91.7	98.8
Joint #4	91.7	100.0	96.7	0
Joint #5	90.0	100.0	61.7	48.8
Joint #6	88.3	100.0	81.7	25.0

Underline values indicate good sensor performance for both normal and abnormal conditions

warning references for workers. The suitable threshold value would be different by the requirements of applications.

Secondly, it is noticed that mounting sensors at proper locations is a crucial factor for diagnosing target machine. The sensor on the wrist of the robot arm (stethoscope #1) were able to detect load condition change in the joint axis #3, #4, #5, and #6. However, it failed to give clear separation in the joint axis #1 and #2. The sensor at the base (stethoscope #2) detected the changes in the joint axis #1, #2, and #3, but failed to show feasibility in other joints. In another aspect, spectrogram from sound signals may provide the periodic degradations of machine health as time progresses. If self-diagnosis is operated routinely, the change will be noticeable. Therefore, the machine health conditions can be efficiently monitored by combining RE values and sound spectrogram images altogether. The framework is also expected to industrial applications. Robot tending case, for example, a robot can stop picking up objects if their weight is over the limit, preventing severe impact on joints. The framework can be used to robots with various vendors and controller protocols since it uses internal sound sensor instead joint torque sensor or encoder values.

However, there are several limitations for the proposed framework to be applied to practical applications. The framework shown above detected the anomaly when the applied load is over the design limit. Other conditions such as wearing of gear or pulley, servo motor failure, lack of grease, and misalignment after collision can also cause anomaly. Although other studies showed change of sensor data (Jaber & Bicker, 2016) or caught anomaly (Oh & Yun, 2018) in the given conditions, this study could not verify the cases using the proposed framework.

Another limitation of this study is that effective location of internal sound sensors was not theoretically discovered. If the framework is deployed in practical application, the performance would be different for robots or machines with various dimensions and shapes. In addition, external disturbances and uncertainties may affect the autoencoder model's convergence and accuracy. Preliminary study on internal sound sensor (Yun et al., 2020) and the results of this paper verify that the framework is robust to external sound. However, as our previous study showed that impact by human on an industrial robot was detected by the sensor, external vibration or impact can affect the autoencoder's performance. The experiment showed the robustness of the framework with the noise level of 80 dB based on safety rules in the United States. However, certain workplaces may have higher noise level, which can deteriorate the robustness. In combination with other anomalies mentioned above, there are also other uncertainties to reduce the model's performance. In further study, the considerations are required to be verified to improve the robustness of the framework to be applicable in industrial robots at shop floors. Theoretical analysis of the framework can be also possible by comparing with other sensors like accelerometer, strain gauge, temperature sensors, and so on.

The hyperparameters of the framework are width of STFT feature and number of hidden layers. The limitation of the framework is that the hyperparameters should be optimized when sensor location is changed. For practical applications, introducing parameter tuning methods is required. For example, recent study (Sun et al., 2018) shows that particle swarm optimization (PSO) shows higher performance than grid search. In addition, metaheuristic algorithms do not require gradients and convexity of the problem while finding global optimum (Stojanovic et al., 2016), and their usage on hyperparameter tuning has been reported recently (Bibaeva, 2018). Although practical application using autoencoder reviewed in this paper did not demonstrate hyperparameter tuning, the framework in this paper can be improved with appropriate hyperparameter tuning method.

Conclusions

In this paper, a framework for detecting abnormal status of a 6-axis industrial robot arm using internal sound signal and autoencoder neural network is proposed. The sound signals were collected using two stethoscope-based internal sound sensors at different locations of the robot arm. Spectrograms were derived from raw signals for visualization and feature extraction. Meaningful low frequency information under 255 Hz was sorted out from the spectrograms after the system identification of the internal sound sensor. The autoencoder-based framework was selected as the main structure of the anomaly detection algorithm in this work. The lack of faulty status data was assumed; hence the normal status data were utilized for training the autoencoders. The abnormality was assigned by imposing load conditions (0–12.5 lb) on the robot arm. As a result, the excessive load conditions (5.0–12.5 lb), which were not included in training data set, produced comparably higher reconstruction error (RE) than those of acceptable load conditions (0–2.5 lb).

The stethoscope #1 at the wrist of robot predicted correct status of axis #3 (96.4%), axis #4 (96.4%), axis #5 (95.7%), and axis #6 (95%). The stethoscope #2 at the base of robot predicted anomalies of axis #1 (97.1%), axis #2 (97.1%), and axis #3 (95.7%). In addition, it is discovered that as the position of stethoscopes is close to the axis, the success rate of the discrimination is increased. However, the prediction rates of each stethoscope were not satisfactory as the sensor and the target are distant.

It is verified that a classic stethoscope can be an alternative sensing tool compared with other commercial sensors. By implementing a stethoscope as sensing tool, cost-effective solution for machine health monitoring is possible. Especially, robotic manufacturing processes causing vibration and sound will receive benefits for using the developed sensor. For example, robotic drilling and milling for aerospace parts made of aluminum or Carbon-fiber-reinforced polymers (CFRP) can be monitored using the acoustic sensor. The anomalies of sounds such as tool failure, poor machining quality, and severe tool wear would be caught using the proposed sensor and autoencoder. In addition, the focusing effect of stethoscope could be applied for resolving the noise susceptibility of general sound sensors. However, the stethoscope had its limitation of narrow frequency response band. The future works is about developing a sound sensor with broader frequency range, which exploits the focusing attribute of the stethoscope as well. It is also expected to develop some degradations of machine health in real production environments by autoencoder strategy without using any fault condition data.

Acknowledgements The authors acknowledge the support by Wabash Heartland Innovation Network (WHIN) and Indiana Manufacturing Competitiveness Center (IN-MAC). This research was also supported by Kyungpook National University Research Fund, 2019. This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-01519, Key technologies for teleoperation of robot-based manufacturing system for 4IR).

Appendix 1

See Table 5.

Table 5 The numbers of training and validation sets are shown in the table. The number differs by the input dimension

Data size	Joint #1	Joint #2	Joint #3	Joint #4	Joint #5	Joint #6
Training	$3083 - n$	$2119 - n$	$1575 - n$	$4272 - n$	$2801 - n$	$6779 - n$
Validation	$633 - n$	$434 - n$	$314 - n$	$870 - n$	$582 - n$	$1393 - n$

Appendix 2

See Table 6.

Table 6 Training results in each axis are summarized to compare the final losses of autoencoder structure. Selected autoencoder parameters are highlighted with bold, and underlined characters

Train loss validation loss		Stethoscope #1 (wrist)			Stethoscope #2 (base)		
Dimensions		1	3	5	1	3	5
Joint #1	n = 4 (256 × 4)	0.0043	0.0037	**0.0016**	0.0065	0.0052	0.0055
		0.0043	0.0038	**0.0015**	0.0072	0.0059	0.0053
	n = 8 (256 × 8)	0.0033	0.0030	0.0009	0.0063	**0.0045**	0.0050
		0.0035	0.0053	0.0015	0.0063	**0.0046**	0.0048
	n = 16 (256 × 16)	0.0011	0.0020	0.0008	0.0050	0.0042	0.0047
		0.0038	0.0044	0.0014	0.0063	0.0047	0.0050
	n = 32 (256 × 32)	0.0174	0.0020	0.0009	0.0045	0.0041	0.0047
		0.0162	0.0045	0.0013	0.0062	0.0047	0.0050
Joint #2	n = 4 (256 × 4)	0.0082	0.0046	0.0042	0.0065	0.0043	**0.0035**
		0.0075	0.0064	0.0063	0.0065	0.0044	**0.0038**
	n = 8 (256 × 8)	0.0044	**0.0017**	0.0039	0.0059	0.0035	0.0035
		0.0046	**0.0020**	0.0032	0.0065	0.0038	0.0037
	n = 16 (256 × 16)	0.0041	0.0012	0.0031	0.0037	0.0034	0.0009
		0.0046	0.0021	0.0029	0.0049	0.0038	0.0037
	n = 32 (256 × 32)	0.0028	0.0013	0.0029	0.0039	0.0032	0.0010
		0.0044	0.0019	0.0029	0.0046	0.0039	0.0038
Joint #3	n = 4 (256 × 4)	0.0072	0.0068	0.0046	0.0121	0.0081	0.0062
		0.0067	0.0071	0.0068	0.0114	0.0088	0.0092
	n = 8 (256 × 8)	0.0057	0.0064	**0.0047**	0.0102	0.0079	**0.0085**
		0.0067	0.0064	**0.0047**	0.0102	0.0091	**0.0089**
	n = 16 (256 × 16)	0.0053	0.0047	0.0045	0.0074	0.0055	0.0085
		0.0067	0.0048	0.0047	0.0098	0.0088	0.0088
	n = 32 (256 × 32)	0.0046	0.0036	0.0177	0.0052	0.0042	0.0172
		0.0066	0.0050	0.0154	0.0099	0.0089	0.0160
Joint #4	n = 4 (256 × 4)	0.0023	0.0017	0.0005	**0.0015**	0.0008	0.0011
		0.0022	0.0024	0.0026	**0.0018**	0.0030	0.0036
	n = 8 (256 × 8)	0.0004	0.0010	0.0009	0.0014	0.0006	0.0009
		0.0023	0.0011	0.0011	0.0022	0.0030	0.0035
	n = 16 (256 × 16)	0.0007	0.0008	**0.0006**	0.0014	0.0005	0.0032
		0.0036	0.0012	**0.0005**	0.0026	0.0031	0.0065
	n = 32 (256 × 32)	0.0003	0.0008	0.0006	0.0013	0.0002	0.0032
		0.0029	0.0011	0.0005	0.0024	0.0029	0.0065
Joint #5	n = 4 (256 × 4)	0.0006	**0.0012**	0.0020	0.0037	**0.0031**	0.0038
		0.0016	**0.0014**	0.0020	0.0036	**0.0033**	0.0038
	n = 8 (256 × 8)	0.0012	0.0008	0.0017	0.0036	0.0028	0.0034
		0.0013	0.0013	0.0014	0.0036	0.0034	0.0035
	n = 16 (256 × 16)	0.0009	0.0007	0.0074	0.0034	0.0027	0.0075
		0.0013	0.0013	0.0074	0.0036	0.0033	0.0080
	n = 32 (256 × 32)	0.0008	0.0007	0.0074	0.0029	0.0028	0.0075
		0.0012	0.0014	0.0074	0.0037	0.0034	0.0080
Joint #6	n = 4 (256 × 4)	0.0122	0.0111	0.0059	0.0053	**0.0007**	0.0018
		0.0140	0.0105	0.0077	0.0059	**0.0010**	0.0027
	n = 8 (256 × 8)	0.0111	0.0063	0.0047	0.0054	0.0005	0.0016
		0.0113	0.0103	0.0050	0.0055	0.0011	0.0027
	n = 16 (256 × 16)	0.0063	0.0043	0.0036	0.0047	0.0005	0.0044
		0.0065	0.0052	0.0037	0.0055	0.0010	0.0048
	n = 32 (256 × 32)	0.0063	0.0032	**0.0013**	0.0046	0.0005	0.0044
		0.0052	0.0040	**0.0016**	0.0056	0.0010	0.0048

Appendix 3

See Table 7.

Appendix 4

See Table8.

Table 7 The table shows the first threshold values in each axis are settled from the maximum REs

Stethoscope	Values	Axis #1	Axis #2	Axis #3	Axis #4	Axis #5	Axis #6
Sensor #1	Maximum RE	0.0034	0.0220	0.0074	0.0007	0.0019	0.0027
	Initial threshold	0.0035	0.0225	0.0075	0.0008	0.0020	0.0028
Sensor #2	Maximum RE	0.0063	0.0054	0.0121	0.0031	0.0054	0.0019
	Initial threshold	0.0065	0.0055	0.0123	0.0032	0.0055	0.0019

Table 8 Estimation results of entire axes are explained in the table. S1 and S2 means stethoscope #1 (Wrist) and #2 (Base), respectively. Green characters indicate correct predictions, and red characters indicate false-positive and false-negative decisions. In the rightmost column, Applicable solutions are highlighted in blue characters

	Conditions		0lb	1.25lb	2.5lb	5.0lb	7.5lb	10.0lb	12.5lb	Success rate
Joint #1	S1	Normal	0/20	20/20	20/20	8/20	20/20	0/20	0/20	40/60 (66%)
		Abnormal	20/20	0/20	0/20	12/20	0/20	20/20	20/20	52/80 (65%)
	S2	Normal	20/20	20/20	16/20	0/20	0/20	0/20	0/20	56/60 (93.3%)
		Abnormal	0/20	0/20	4/20	20/20	20/20	20/20	20/20	80/80 (100%)
Joint #2	S1	Normal	12/20	20/20	20/20	19/20	0/20	1/20	0/20	52/60 (86.7%)
		Abnormal	8/20	0/20	0/20	1/20	20/20	19/20	20/20	60/80 (75%)
	S2	Normal	19/20	20/20	18/20	1/20	0/20	0/20	0/20	57/60 (95%)
		Abnormal	1/20	0/20	2/20	19/20	20/20	20/20	20/20	79/80 (98.8%)
Joint #3	S1	Normal	18/20	20/20	17/20	0/20	0/20	0/20	0/20	55/60 (91.7%)
		Abnormal	2/20	0/20	3/20	20/20	20/20	20/20	20/20	80/80 (100%)
	S2	Normal	20/20	15/20	20/20	0/20	0/20	0/20	1/20	55/60 (91.7%)
		Abnormal	0/20	5/20	0/20	20/20	20/20	20/20	19/20	79/80 (98.8%)
Joint #4	S1	Normal	20/20	15/20	20/20	0/20	0/20	0/20	0/20	55/60 (91.7%)
		Abnormal	0/20	5/20	0/20	20/20	20/20	20/20	20/20	80/80 (100%)
	S2	Normal	18/20	20/20	20/20	20/20	20/20	20/20	20/20	58/60 (96.7%)
		Abnormal	2/20	0/20	0/20	0/20	0/20	0/20	0/20	0/80 (0%)

Table 8 continued

Joint #5	**S1**	Normal	14/20	20/20	20/20	0/20	0/20	0/20	0/20	54/60 (90%)
		Abnormal	6/20	0/20	0/20	20/20	20/20	20/20	20/20	80/80 (100%)
	S2	Normal	10/20	17/20	10/20	20/20	20/20	1/20	0/20	37/60 (61.7%)
		Abnormal	10/20	3/20	10/20	0/20	0/20	19/20	20/20	39/80 (48.8%)
Joint #6	**S1**	Normal	20/20	20/20	13/20	0/20	0/20	0/20	0/20	53/60 (88.3%)
		Abnormal	0/20	0/20	7/20	20/20	20/20	20/20	20/20	80/80 (100%)
	S2	Normal	20/20	20/20	9/20	0/20	20/20	20/20	20/20	49/60 (81.7%)
		Abnormal	0/20	0/20	11/20	20/20	0/20	0/20	0/20	20/80 (25%)

References

Al-Ghamd, A. M., & Mba, D. (2006). A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical Systems and Signal Processing, 20*(7), 1537–1571. https://doi.org/10.1016/j.ymssp.2004.10.013

Bibaeva, V. (2018). Using metaheuristics for hyper-parameter optimization of convolutional neural networks. In 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1–6). https://doi.org/10.1109/MLSP.2018.8516989

Bittencourt, A. C., & Gunnarsson, S. (2012). Static friction in a robot joint—modeling and identification of load and temperature effects. *Journal of Dynamic Systems, Measurement, and Control.* https://doi.org/10.1115/1.4006589

Chamberlain, D., Mofor, J., Fletcher, R., & Kodgule, R. (2015). Mobile stethoscope and signal processing algorithms for pulmonary screening and diagnostics. In 2015 IEEE Global Humanitarian Technology Conference (GHTC) (pp. 385–392). https://doi.org/10.1109/GHTC.2015.7344001

Chebil, J., Noel, G., Mesbah, M., & Deriche, M. (2009). Wavelet decomposition for the detection and diagnosis of faults in rolling element bearings. *Jordan Journal of Mechanical and Industrial Engineering, 3*(4), 260–267.

Cong, F., Chen, J., Dong, G., & Pecht, M. (2013). Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis. *Journal of Sound and Vibration, 332*(8), 2081–2097. https://doi.org/10.1016/j.jsv.2012.11.029

Dohnal, F., & Sekhar, A. S. (2014). Current signature analysis for unbalance fault detection in a rotor supported by active magnetic bearings. *International Journal of Condition Monitoring, 4*(1), 2–8. https://doi.org/10.1784/204764214813883315

Eren, L., Ince, T., & Kiranyaz, S. (2019). A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *Journal of Signal Processing Systems, 91*(2), 179–189. https://doi.org/10.1007/s11265-018-1378-3

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media.

Gribonval, R. (2008). Linear time-frequency analysis I: Fourier-type representations. Time-Frequency Analysis: Concepts and Methods, 61–91.

Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing, 240*, 98–109. https://doi.org/10.1016/j.neucom.2017.02.045

Gupta, C. N., Palaniappan, R., Swaminathan, S., & Krishnan, S. M. (2007). Neural network classification of homomorphic segmented heart sounds. *Applied Soft Computing, 7*(1), 286–297. https://doi.org/10.1016/j.asoc.2005.06.006

Hamidi, M., Ghassemian, H., & Imani, M. (2018). Classification of heart sound signal using curve fitting and fractal dimension. *Biomedical Signal Processing and Control, 39*, 351–359. https://doi.org/10.1016/j.bspc.2017.08.002

Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing, 23*(3), 724–739. https://doi.org/10.1016/j.ymssp.2008.06.009

Henriquez, P., Alonso, J. B., Ferrer, M. A., & Travieso, C. M. (2014). Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Transactions on Systems, Man, and Cyber-*

netics: Systems, 44(5), 642–652. https://doi.org/10.1109/TSMCC.2013.2257752

Immovilli, F., Bellini, A., Rubini, R., & Tassoni, C. (2010). Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. *IEEE Transactions on Industry Applications, 46*(4), 1350–1359. https://doi.org/10.1109/TIA.2010.2049623

Jaber, A. A., & Bicker, R. (2016). Fault diagnosis of industrial robot gears based on discrete wavelet transform and artificial neural network. *Insight—Non-Destructive Testing and Condition Monitoring, 58*(4), 179–186. https://doi.org/10.1784/insi.2016.58.4.179

Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., et al. (2016). Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration, 377*, 331–345. https://doi.org/10.1016/j.jsv.2016.05.027

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing, 107*, 241–265. https://doi.org/10.1016/j.ymssp.2017.11.024

Kingma, D. P., & Ba, J. (2017). Adam: a method for stochastic optimization. arXiv:1412.6980 [cs]. http://arxiv.org/abs/1412.6980. Accessed 31 May 2020

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal, 37*(2), 233–243. https://doi.org/10.1002/aic.690370209

Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions, 77*, 167–178. https://doi.org/10.1016/j.isatra.2018.04.005

Liu, Y., & Zhang, Y. (2015). Toward welding robot with human knowledge: A remotely-controlled approach. *IEEE Transactions on Automation Science and Engineering, 12*(2), 769–774. https://doi.org/10.1109/TASE.2014.2359006

Maglogiannis, I., Loukis, E., Zafiropoulos, E., & Stasis, A. (2009). Support vectors machine-based identification of heart valve diseases using heart sounds. *Computer Methods and Programs in Biomedicine, 95*(1), 47–61. https://doi.org/10.1016/j.cmpb.2009.01.003

Martin, H. R., & Honarvar, F. (1995). Application of statistical moments to bearing failure detection. *Applied Acoustics, 44*(1), 67–77. https://doi.org/10.1016/0003-682X(94)P4420-B

Meng, Q., Sen, D., Wang, S., & Hayes, L. (2008). Impulse response measurement with sine sweeps and amplitude modulation schemes. In 2008 2nd International Conference on Signal Processing and Communication Systems (pp. 1–5). https://doi.org/10.1109/ICSPCS.2008.4813749

Mikolajczyk, T. (2012a). Manufacturing using robot. *Advanced Materials Research, 463–464*, 1643–1646. https://doi.org/10.4028/www.scientific.net/AMR.463-464.1643

Mikolajczyk, T. (2012b). System to surface control in robot machining. *Advanced Materials Research, 463–464*, 708–711. https://doi.org/10.4028/www.scientific.net/AMR.463-464.708

Mikolajczyk, T. (2013). Indication of machining area with the robot's camera using. *Applied Mechanics and Materials, 282*, 146–151. https://doi.org/10.4028/www.scientific.net/AMM.282.146

Mikolajczyk, T. (2015). Control system for industrial robot equipped with tool for advanced task in manufacturing. *Applied Mechanics and Materials, 783*, 105–113. https://doi.org/10.4028/www.scientific.net/AMM.783.105

Oh, D. Y., & Yun, I. D. (2018). Residual error based anomaly detection using auto-encoder in SMD machine sound. *Sensors, 18*(5), 1308. https://doi.org/10.3390/s18051308

Pan, Y., Chen, J., & Guo, L. (2009). Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description. *Mechanical Systems and Signal Processing, 23*(3), 669–681. https://doi.org/10.1016/j.ymssp.2008.05.011

Park, Y., & Yun, I. D. (2018). Fast adaptive RNN encoder–decoder for anomaly detection in SMD assembly machine. *Sensors, 18*(10), 3573. https://doi.org/10.3390/s18103573

Peng, Z. K., Tse, P. W., & Chu, F. L. (2005). A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing, 19*(5), 974–988. https://doi.org/10.1016/j.ymssp.2004.01.006

Qi, Y., Wang, Y., Zheng, X., & Wu, Z. (2014). Robust feature learning by stacked autoencoder with maximum correntropy criterion. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6716–6720). https://doi.org/10.1109/ICASSP.2014.6854900

Rai, V. K., & Mohanty, A. R. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. *Mechanical Systems and Signal Processing, 21*(6), 2607–2615. https://doi.org/10.1016/j.ymssp.2006.12.004

Safizadeh, M. S., & Latifi, S. K. (2014). Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. *Information Fusion, 18*, 1–8. https://doi.org/10.1016/j.inffus.2013.10.002

Shaffer, D., Ragai, I., Danesh-Yazdi, A., & Loker, D. (2018). Investigation of the feasibility of using microphone arrays in monitoring machining conditions. *Manufacturing Letters, 15*, 132–134. https://doi.org/10.1016/j.mfglet.2017.12.008

Shao, H., Jiang, H., Zhao, H., & Wang, F. (2017). A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing, 95*, 187–204. https://doi.org/10.1016/j.ymssp.2017.03.034

Stojanovic, V., Nedic, N., Prsic, D., Dubonjic, L., & Djordjevic, V. (2016). Application of Cuckoo search algorithm to constrained control problem of a parallel robot platform. *The International Journal of Advanced Manufacturing Technology, 87*(9), 2497–2507. https://doi.org/10.1007/s00170-016-8627-z

Sun, Y., Xue, B., Zhang, M., & Yen, G. G. (2018). An experimental study on hyper-parameter optimization for stacked auto-encoders. In 2018 IEEE Congress on Evolutionary Computation (CEC) (pp. 1–8). https://doi.org/10.1109/CEC.2018.8477921

Tao, H., Wang, P., Chen, Y., Stojanovic, V., & Yang, H. (2020). An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks. *Journal of the Franklin Institute, 357*(11), 7286–7307. https://doi.org/10.1016/j.jfranklin.2020.04.024

Tao, S., Zhang, T., Yang, J., Wang, X., & Lu, W. (2015). Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In 2015 34th Chinese Control Conference (CCC) (pp. 6331–6335). https://doi.org/10.1109/ChiCC.2015.7260634

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning (pp. 1096–1103). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1390156.1390294

Wang, D., Tse, P. W., & Tsui, K. L. (2013). An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing, 35*(1), 176–199. https://doi.org/10.1016/j.ymssp.2012.10.003

Yun, H., Kim, H., Kim, E., & Jun, M. B. G. (2020). Development of internal sound sensor using stethoscope and its applications for machine monitoring. *Procedia Manufacturing, 48*, 1072–1078. https://doi.org/10.1016/j.promfg.2020.05.147

Zhang, B., Li, W., Hao, J., Li, X.-L., & Zhang, M. (2018). Adversarial adaptive 1-D convolutional neural networks for bearing fault diagnosis under varying working condition. arXiv:1805.00778 [cs, eess]. http://arxiv.org/abs/1805.00778.

Zhou, C., & Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 665–674). Halifax, NS, Canada: Association for Computing Machinery. https://doi.org/10.1145/3097983.3098052

Zhu, D., Feng, X., Xu, X., Yang, Z., Li, W., Yan, S., & Ding, H. (2020). Robotic grinding of complex components: A step towards efficient and intelligent machining–challenges, solutions, and applications. *Robotics and Computer-Integrated Manufacturing, 65*, 101908. https://doi.org/10.1016/j.rcim.2019.101908