Nuclear Engineering and Technology 55 (2023) 3716–3724

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

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

Original Article

Discrimination model using denoising autoencoder-based majority vote classification for reducing false alarm rate

Heonyong Lee [a,*], Kyungtak Yu [b], Shiu Kim [c]

[a] Department of Mechanical Engineering, Seoul National University, Seoul, South Korea
[b] 1103, Ace High End Tower, 14, Seongsui-ro 10-gil, Seongdong-gu, Seoul, South Korea
[c] 846, Hongnong-ro, Hongnong-eup, Yeonggwang-gun, Jeollanam-do, South Korea

ARTICLE INFO

Article history:
Received 27 December 2022
Received in revised form
20 June 2023
Accepted 21 June 2023
Available online 22 June 2023

Keywords:
Loose part monitoring
Deep learning
Denoising auto-encoder
Impact signal
False alarm

ABSTRACT

Loose parts monitoring and detecting alarm type in real Nuclear Power Plant have challenges such as background noise, insufficient alarm data, and difficulty of distinction between alarm data that occur during start and stop. Although many signal processing methods and alarm determination algorithms have been developed, it is not easy to determine valid alarm and extract the meaning data from alarm signal including background noise. To address these issues, this paper proposes a denoising autoencoder-based majority vote classification. Training and test data are prepared by acquiring alarm data from real NPP and simulation facility for data augmentation, and noisy data is reproduced by adding Gaussian noise. Using DAEs with 3, 5, 7, and 9 layers, features are extracted for each model and classified into neural networks. Finally, the results obtained from each DAE are classified by majority voting. Also, through comparison with other methods, the accuracy and the false alarm rate are compared, and the excellence of the proposed method is confirmed.

1. Introduction

Loose parts mean materials disengaged by mechanical damage or corrosion from structure or tools left inside during construction or maintenance. As an impact caused by loose parts occurs inside the reactor during all operation procedures (starting, operating and stopping), it can cause serious safety problems by blocking a flow path, nuclear fuel damage, and driving obstruction. In order to avoid these problems in advance, the loose part monitoring system(LPMS) should monitor and analyze the impact signal to prevent accidents caused by loose parts [1–3].

The LPMS uses accelerometers mounted on the outside of the reactor system to detect signals that exceed a certain level and classify them according to a specified algorithm. Currently, most LPMS are designed to use a complex algorithm that uses the ratio and time difference of the signals, called the event screen algorithm, to determine whether the signal is an actual impact signal or not. The event screen algorithm performs several types of tests independently to determine the authenticity of an alarm, and if all of these tests result in an actual, the detected signal is judged as an alarm. Therefore, LPMS judges all signals as normal signals except those signals that are determined to be alarms, as shown in Fig. 1(a), and in the event of an alarm, actions such as emergency shutdown are taken to resolve the problems occurring in the NPP.

The entire signal acquisition and discrimination process at the current facility is shown in Fig. 2. Firstly, if the signal exceeds a certain threshold set in the first determination, signal acquisition is started. The acquired signal is judged by an independent algorithm using signal ratio and time difference by the second discrimination, the event screen algorithm, to determine the integrity of the signal. This signal is interpreted as an impact signal from a loose part and triggers an alarm. In practice, however, there are both valid alarms, which are caused by actual loose parts, and false alarms, which are not caused by loose parts. In other words, the existing LPMS has a very high false alarm rate because it determines both these valid and false alarm signals as alarms.

Fig. 1 shows the valid and false alarm signal that the LPMS currently identifies as alarm signals. All of the signals represented in Fig. 1 are currently considered alarm signals by LPMS, but as mentioned above, Fig. 1(a) and (b) are valid alarms caused by a loose part, while (c) is a false alarm caused by something other than a loose part. This false alarm is a signal that the existing algorithm has identified as an alarm, but is not actually caused by a loose part.

* Corresponding author.
E-mail address: heonyong2@gmail.com (H. Lee).

https://doi.org/10.1016/j.net.2023.06.037
1738-5733/

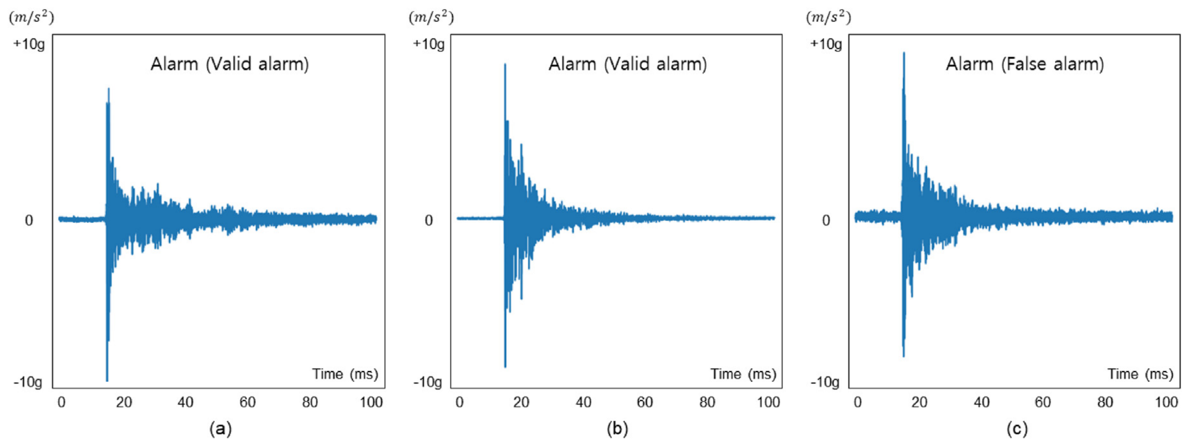

Fig. 1. Signal pattern (a) Valid alarm signal in NPP (b) Valid alarm signal in ¼ scale facility (c) False alarm signal.

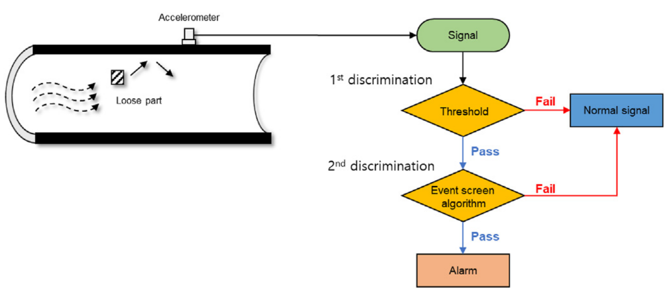

Fig. 2. Signal discrimination flow.

Compared to an alarm caused by a loose part, it is not easy to distinguish it from a valid alarm because the signal pattern is similar. Also, the contamination of the signal due to background noise generated in the operating environment makes it difficult to determine and detect the alarm signal. This problem increases maintenance cost as well as decreases the reliability of the LPMS.

There are two main reasons for generating false alarms. The first is that it is difficult to accurately classify alarm signals that occur during the start and stop for operating the NPP. Former research has been conducted steadily to detect loose parts in nuclear reactors [4–12]. Among studies, most of them use comparison of RMS values to determine alarm signals by comparing with empirical threshold or extracting signal characteristics such as frequency from the original signal [13–16]. Although these algorithms show good results in normal operation of the NPP, they generate a large amount of false alarms in other environments, such as start and stop section of the NPP when temperature and pressure increase or decrease rapidly, aggravating the task of operators and making it difficult to detect impact signals quickly and accurately.

To reduce false alarms, Kim(Kim et al., 2002) proposed a method to find the optimal value of the rising time, half period, and maximum amplitude to evaluate the impact signal using the neural network [17], and Cao(Cao et al., 2012) proposed an approach to reduce the false alarm rate by combining linear predictive coding(LPC) and support vector machine(SVM) [18]. In addition, Min(Min et al., 2014) attempted to reduce the false alarm rate by introducing the frequency ratio(FR) method using the different ratio of power spectrum density(PSD) between the impact and false alarm signal [19].

Second, there is a contamination of alarm signals at acquisition due to noise. In actual NPP, noise may occur due to the start and stop of facilities, and the acquisition signal may be affected by changes in the surrounding environment. Therefore, efforts to make the signal clearer and reduce the false alarm rate by removing noise mixed with the signal acquired from the NPP have also been consistent. Yang(Yang et al., 2016) used eigenvector algorithm(EVA) to remove superimposed noise from the impact signal [20], and Meng(Meng et al., 2020) eliminates noise through a method called objective function method(OFM), confirming that it performs better than noise removal through conventional wavelet transformations [21]. In addition, Kim(Kim et al., 2001) effectively eliminated noise in the high frequency band using moving average filtering [22]. However, most of them require additional processing, and there is a limitation in that the original signal cannot be used as it is.

In recent years, with the development of deep learning technology, several deep learning methods concerning classification have been applied to the fault diagnosis of machines. In order to perform the denoising described above, some studies introduced stacked denoising autoencoder(DAE) [23] after extracting signal features, and then applied them to deep learning techniques to classify the signal [24]. Also, a study was conducted on a model for classifying images of acquired signals using time-frequency analysis [25]. Meanwhile, research is underway to perform fault diagnosis using deep learning techniques as they are in the raw signal rather than methods which require additional processing such as extracting signal features or noise removal. Haidong (Haidong et al., 2017) explained how to detect signals differently from conventional feature extractions by performing feature learning using the autoencoder(AE) method on rotating machines [26]. Jia (Jia et al., 2018) proposed further from the underlying AE and detected the anomaly signal using values that normalized the weights of the arbitrary segment signals through two techniques, normalized sparse autoencoder(NSAE) and local connection network(LCN) [27]. Zhang and Yang performed fault detection on the raw bearing signal based on convolutional neural network(CNN). Zhang(Zhang et al., 2018) showed noise removal and improvement of general-purpose results on datasets of training models through kernel dropout and small batch training [28]. Yang(Yang et al., 2019) proposed a method performing fault diagnosis using multi-channel signals as opposed to single-channel signal data [29].

The purpose of this study is to develop a model that reduces false alarm rates through the correct classification of alarm signals and accurately determines contaminated signals due to background noise. Through this, it is expected to be a starting point to overcome the shortcomings of monitoring facilities that had to rely on related knowledge and experience.

The outline of this paper is organized as follows. The preliminaries are presented in Section 2. Then Section 3 mentions the proposed method and evaluation index. The experiments and performance results are described in Section 4 and 5. The conclusion and future work of this paper is briefly summarized in Section 6.

2. Preliminaries

2.1. Problem definition

Accelerometers installed on the surface of the facility detect vibrations during operation in real time and transmit them to the LPMS. If the received signal exceeds a certain threshold, it detects the anomaly, captures the signal and determines the authenticity of the alarm signal. In this case, some of the signals identified as alarms are valid alarms caused by the impact, while others are false alarms caused by no impact. In summary, the problem is that the alarms generated by LPMS contain both valid and false alarms, and the current algorithm cannot distinguish between them, resulting in a large number of false alarms.

The inability to distinguish between the two alarms is due to signal characteristics and configuration. The actual acquired valid and false alarm signals are shown in Fig. 1, and the signal data acquisition is described in Section 3. All signals are acquired under the same conditions as in a real nuclear power plant environment. The signal is captured at a sampling rate of 200 K and the acceleration of the vibrations generated by the plant is measured using accelerometers. From the point where the signal exceeds the threshold(4 g), the signal is measured for a period of 15 ms forward and 85 ms backward. Fig. 1(a) and (c) are the valid and false alarms obtained from the NPP, and it can be seen that the shapes of the two are similar to the naked eye. In addition, the characterization of the signals by signal processing showed that the distributions of the two signals overlap significantly. Therefore, we should ultimately reduce the false alarm rate by developing a model that clearly distinguishes between the two alarms.

2.2. AE-based anomaly detection and classification

Classifying is one of the oldest and most common problems encountered in industry. A variety of statistical techniques and algorithms are used, with appropriate models depending on the type and nature of the data. In the past, research has focused on the preprocessing of data to extract features and to select classifiable criteria, or on the grouping of the same features according to their distribution. However, these studies are time consuming and require specialized knowledge to apply additional pre-processing or to select different attributes depending on the nature of the data.

To address the above issues, this study uses an autoencoder to automatically extract data features from the raw signal and perform classification using the extracted features. Autoencoder-based data classification is a clustering-based method that categorizes data in a space of latent vectors that represent the most important features of the data. This means constructing a new representation space that better encapsulates the features of the data, and separating the data according to the latent vector extracted from that feature map. The idea is to sort the data on the assumption that the data of the same class will be close to each other. So, it is important to find a good representation space where different classes can be well distinguished in order to achieve good classification performance.

Preprocessing the input requires expertise in selecting or extracting features. The disadvantage is that it increases the time required. Therefore, in this study, by using the raw signal as it is, the model automatically selects and extracts features, and performs end-to-end learning without the need for separate preprocessing. Through this, the entire process was automated, and the feature selection and extraction procedure through expert knowledge were excluded.

2.3. Autoencoder

AE consists of an encoding structure that compresses the data to extract only the important features, and a decoding structure that reconstructs it to resemble the original data. This means that the model learns by comparing the input data with the output data after compressing and restoring, and extracts the best compressed and restored latent vectors. Here, the latent vector is the bottleneck layer at the center of the AE, representing values that embody important features of the input data. Important features are data that have been stripped of information that is not valuable as data, such as noise.

It is a symmetric structure with input and output layers on either side and a bottleneck layer in the middle. When data is received as input, AE consists of an encoder to reduce the input to a latent variable and a decoder to play back the latent variable as input. Basically, the two-part structure is symmetric, and the representation of the $(l+1)-$ th layer is as follows.

$$a^{l+1} = f\left(W^l a^l + b^l\right)$$

where $f(\cdot)$ is the activation function, a^l is the output of the $l-$ th layer, and W^l and b^l are weight and bias between the $l-$ th and $(l+1)-$ th layers, respectively.

Here, the encoder extracts essential features from the input, and the decoder reconstructs the input from the output of the encoder. In other words, the purpose of AE is to reduce the input data to a low-dimensional latent vector and to implement the original input as it is in the input data reproduced from this latent vector. The loss function of AE is calculated as follows.

$$l(x,\tilde{x}) = \sum_{j=1}^{n} \left(x_j - \tilde{x}_j\right)^2$$

where x_j is $j-th$ data of input x and $\tilde{x}_j$ is $j-th$ data of input $\tilde{x}$.

AE performs training by minimizing the mean squared error (MSE) or the mean absolute error (MAE) as the mathematical expression as below. Here, MAE is used because it is necessary to classify differences between classes by sensitively responding to local variations in the signal [31].

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{x}_i)^2, L_{MAE} = \frac{1}{N}\sum_{i=1}^{N}|x_i - \widehat{x}_i|$$

where x_i is the $i-$ th sample and N is the number of samples.

In this study, 4 types of DAE models (3, 5, 7 and 9-DAE) are used. The DAEs are designed in accordance with the number of layers in the hidden layer. Once the AE has successfully trained a feature on the input data, it can be said that it is an appropriate AE model to be used for classifying.

2.4. Denoising autoencoder

DAE is the model proposed by Vincent [30]. Basically, it has the same structure as the AE, and when a contaminated input is received, it is an AE model that reproduces the original data. Comparing to the AE, it is almost identical except that the input data is replaced by the contaminated input data. The basic structure

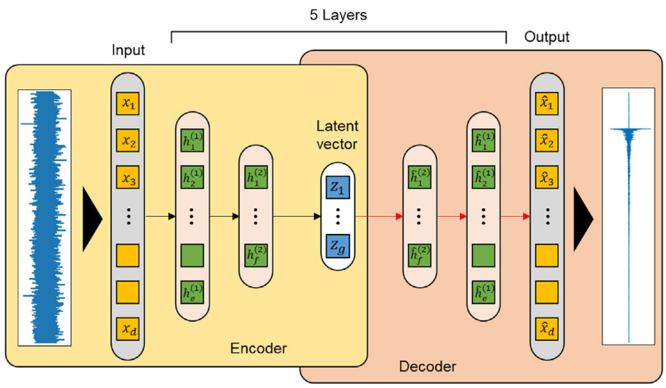

Fig. 3. Architecture of Denoising auto-encoder with 5 layers including latent vector.

of the DAE is shown in Fig. 3. In other words, all configurations are the same except that the input is noisy, contaminated data in comparison to the AE. DAE has the advantage of removing the embedded noise in the input as well as extracting better features while retaining the information in the input data. Therefore, if the noise is strong, the AE will have difficulty extracting the main features that contain data information, but the DAE can extract more accurate data features by removing the noise. This is because the input is a noisy, contaminated signal rather than the original, so the parameters are modified during compression and restoration to extract better features.

3. Proposed model

3.1. Proposed network

To differentiate the alarm signal between valid and false, DAE-MV is now proposed. The entire architecture is shown in Fig. 4. As with the typical classification model process, the proposed model learns and scores through training and test data. Prior to learning, the signal data does not perform any pre-processing to extract signal features, and only prepares the input data set according to the signal size.

In training, we prepare a noise signal to be used as input and a raw signal to compare with the reconstructed result, which noise and raw signals train the model to minimize the discrepancy between them. The model is robust to noisy signals and uses the AE part of the DAE to obtain a well-characterized latent vector from the input signal. The resulting latent vector is mapped to the new representation space and contains the important features of the input data. The latent vector becomes the new input, which is divided into two classes by the neural network. This ensures that each DAE has its own feature map and classification results.

The results of the four classifiers with different number of layers obtained above are then fed back into the input of the majority vote classifier and used for the final classification. Because the features represented in each layer are different, the features you observe will vary depending on the depth of the layer. This is used to classify the data using a majority voting method based on the classification results for each layer. The results for each DAE are weighted differently to prevent the voting results from being identical.

3.2. Procedure of proposed method

The proposed DAE-MV procedure consists of three steps. First, four DAEs with different layers are used to extract the latent vectors of the data to be represented in feature space. A neural network is then used to classify the latent vector using the features extracted from each DAE. Finally, the results of each DAE are used to classify the data classes using a majority voting method. Each step is as follows, and the detailed structure and parameters are shown in Fig. 5.

Step 1. To extract latent vectors representing different features, use a DAE with different numbers of layers. As the input data is compressed and restored, the DAE updates the parameters. The weights are fixed for later classification.
Step 2. The extracted latent vectors are mapped to the feature space. Binary classification is performed using a neural network. The class probabilities obtained by the classifier are used as input for the next step.
Step 3. Based on the information obtained from each DAE, the class of the data is determined by majority vote.

In a series of steps, the model extracts the latent vectors that best represent the features of the data and uses these values to calculate the probability of classifying the data in each DAE. The probabilities are then used to classify the data according to a majority vote.

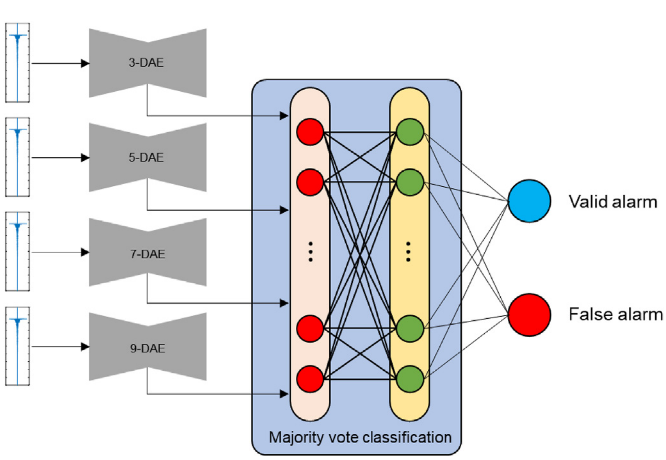

Fig. 4. Architecture of DAE-based majority vote classification network.

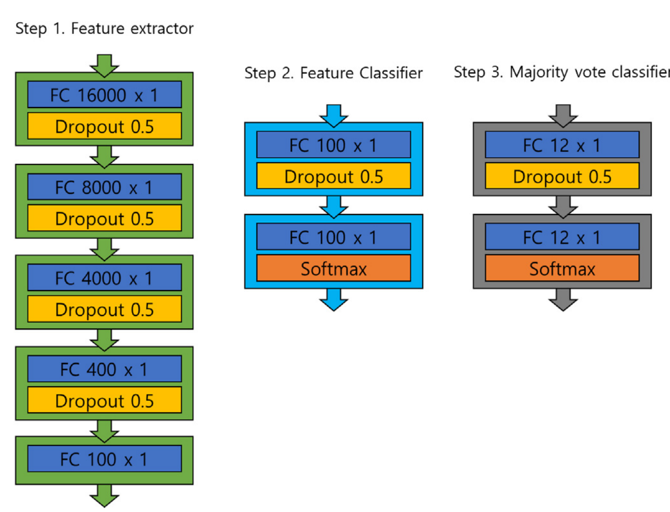

Fig. 5. The architecture and parameters of the proposed meth.

3.3. Performance metrics

To classify the two alarm signals, the AE extracts representation features through training, and the classifier performs binary classification through the extracted representation features. When binary classification is performed, it can be divided into the number of four cases - true positive (TP), true negative (TN), false positive (FP), and false negative (FN) - according to the classification results, and it is possible to evaluate how well the data were classified using the combination of these values. TP refers to a case where a valid alarm is predicted as valid, and TN is a case where a false alarm is predicted as false. FP is predicted to be valid, but it is actually a case of false alarm, and FN is a case where a false alarm is predicted, but it is a case of valid alarm.

This study used accuracy, precision and recall as representative performance metrics among those commonly used as classification metrics, adding false alarm rate to assess frequent false alarm occurrence. Accuracy indicates whether or not valid and false alarms are correctly classified among all signals, and FAR is the proportion of false alarms that the model misidentifies as valid alarms. False alarm rate (FAR) and accuracy are expressed as follows.

$$\text{False Alarm Rate(FAR)} = \frac{\text{FP}}{\text{TP} + \text{FP}} \times 100\ (\%)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\ (\%)$$

On the other hand, precision and recall are also important metrics, given the specificity of NPP and the rarity of impact-induced alarms. Ultimately, a good classification model should be good at both metrics at the same time. There can be no sacrifice of one for the sake of the other. Therefore, to check the performance of recall and precision at the same time, the F1-score is used. Each metric is expressed as below.

$$\begin{aligned}\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\ (\%), \text{Recall} \\ &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\ (\%), \text{F1} - \text{score} \\ &= \frac{2 \times \text{Presicion} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\ (\%)\end{aligned}$$

The above metrics allow to check not only the performance of the model, but also how much the false alarm rate is reduced, and to exclude models that do not detect real alarms, even if they have good classification performance. In this way, an optimal classification model can be selected.

4. Experiments

4.1. Preparation of dataset

In order to conduct this research, the generation and collection of data is essential. Especially for data-driven classification models, data collection is the most important part of the research. Both valid alarms caused by impact and false alarms generated in the field are collected to train and test the model to discriminate between real and false alarms. However, while false alarm data can be obtained during the start-up of the NPP, valid alarm data caused by an impact is not feasible for safety and practical reasons. Therefore, the data was acquired through an impact test using a metal sphere externally, and the same impact test was performed on a 1/4 scale model of the NPP structure to supplement the data. For data acquisition, data is received from 18 sensors at a sampling rate of 200 KHz, as shown in Fig. 6(c), and the signals are acquired for 100 ms. A total of 1225 valid and 1225 false alarms are acquired and the data is used in a 7:3 ratio for training and testing. The acquisition process for each signal is as follows.

4.1.1. False alarm signal at NPP

While most of the false alarm signals are neglected or classified as unnecessary data until now. Not only has the number of false alarm data recorded and stored has been small in quantity, but also the interest in analyzing it has been very low. So, the data management of false alarms has not been done well. It is necessary to record false alarms that occur during the operation of NPP, and for this purpose, signals have been collected during the period of false alarms.

4.1.2. Valid alarm signal at NPP

Accelerometers installed on the surface of the NPP collect the vibration signals generated by the collisions of metal debris in the interior. It is very difficult to obtain a signal by inserting a small metal part inside or hitting the surface during NPP operation. So, it is not easy to obtain a sufficient amount of signal acquired from facility for training and testing models. In this study, an artificial impact is created externally using a metal ball and the vibration is picked up by a sensor to provide an alarm signal from the impact. In Fig. 6(a), four accelerometers are installed on the surface of the steam generator and the signals are acquired by performing an impact test with a 227 g steel ball at the same location as in Table 1, taking into account the safety and design of the facility.

4.1.3. Valid alarm signal at 1/4 scale test facility

Getting a sufficient amount of valid alarm data for learning the model is one of the big problems. In addition, it is necessary to augment the valid alarm data to resolve the imbalance with the number of false alarm data. Therefore, as shown in Fig. 6(b), this study attempted to solve this problem by performing an impact test on a 1/4 scale test facility in KAERI to obtain additional impact signals. An impact signal is detected by sensors at three points excluding the upper position, and the impact test is conducted at a position equivalent to the impact test conducted at the existing power plant, as shown in Fig. 6(a). Considering that the mass and velocity of loose parts generated in the real environment is various, an impact signal is obtained by variously simulating the impact mass(37, 70, 113, 132, 174, and 199 g).

To confirm the association between the signals acquired in the field and test facility, the event screen algorithm of the existing LPMS facility is used to distinguish between the two signals. As a result, it is found that they have similar configurations as shown in Fig. 1(a) and (b), and are identified as alarm signals. To compare the similarity of the signals obtained in the field and in the test facility, the distributions of representative signal characteristics such as frequency, skewness, kurtosis and impulse factor are examined. The probability density functions of the two signal characteristics are shown in Fig. 7. Both signal distributions are contained within the characteristic range of typical impact signals, and are found to have distributions in adjacent ranges. In particular, we can see that the frequency ranges are very similar, which is an important characteristic that distinguishes an impact signal. This means that the impact signals obtained in the test facility are similar to those in the field.

4.2. Generate noise signal

In the real world, the complexity of the peripherals and the variety of noise generated by the environment make it impossible to account for all types of noise. However, it is safe to assume that the noise generated when running continuously will not vary and

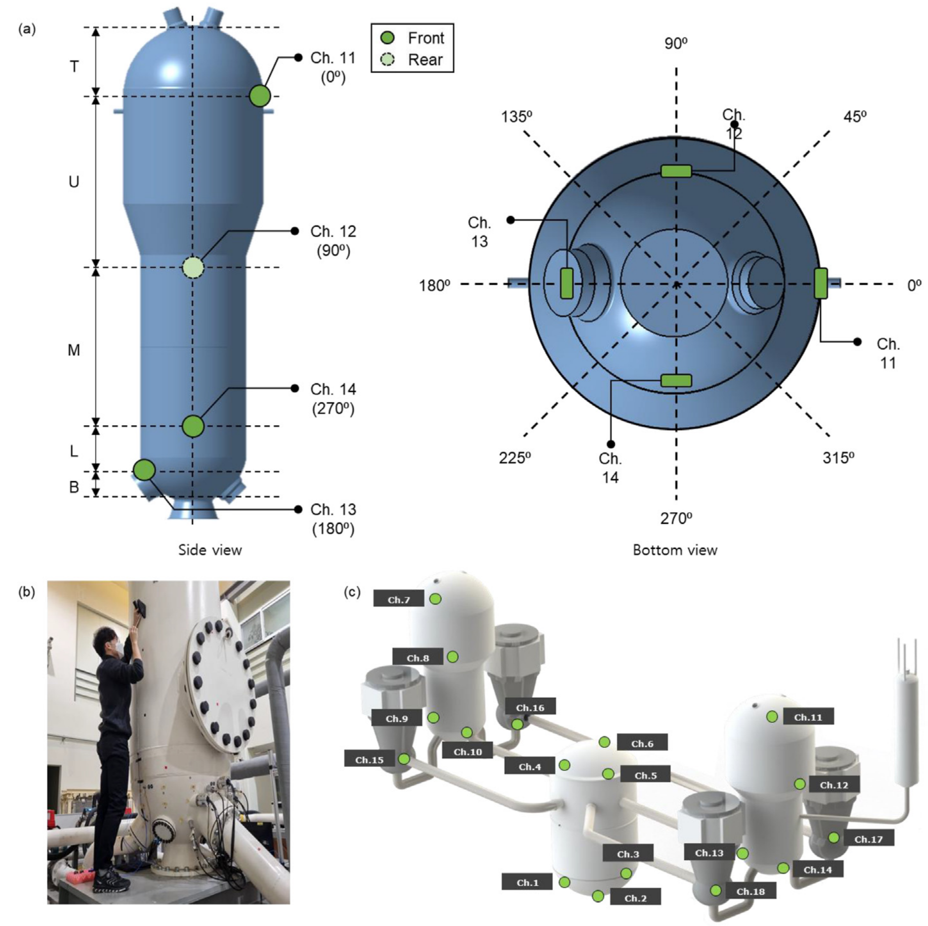

Fig. 6. (a) Schematic diagram of sensor location at steam generator (b) Impact test on ¼ scale steam generator model (c) Schematic diagram of sensor location at NPP.

Table 1
Information of the impact test.

#	Location 1 (Height)	Location 2 (Angle)
1	T (Top)	0/90/180/270 °
2	U (Upper)	0/45/90/135/225/270/315 °
3	M (Middle)	0/45/90/135/180/225/270/315 °
4	L (Lower)	0/45/90/135/180/225/270/315 °
5	B (Bottom)	1/90/270 °

will remain constant. Depending on the assumption, the noise used for the NPP is somewhat random and has a standard distribution. In other words, the noise generated by the NPP is replaced by Gaussian noise, and the noise signal and the original signal are used for training by adding noise of different sizes to the original signal by changing the noise size according to the situation. For a quantitative measure of the added Gaussian noise, the signal-to-noise ratio (SNR) is used, which is defined as follows.

$$SNR_{dB} = 10\ \log_{10}\left(\frac{P_{Signal}}{P_{Noise}}\right)$$

where P_{Signal} and P_{Noise} are the power of the original signal and the added Gaussian noise, respectively.

By learning the DAE with this noisy signal, if the signal acquired in the real environment is also noisy, it can be used in the classification model without additional filtering and robustness can be confirmed.

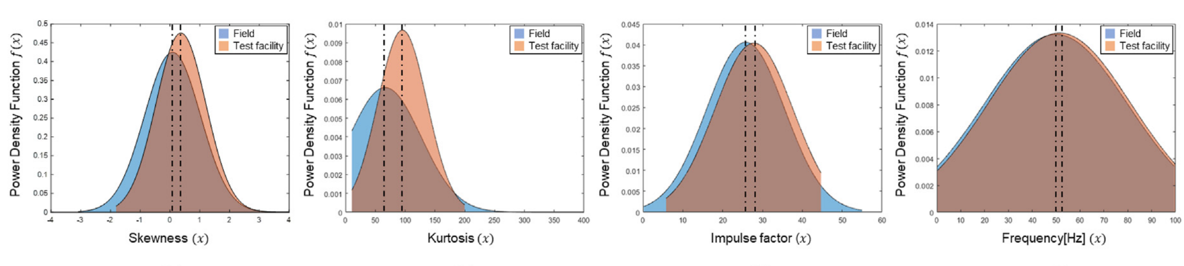

Fig. 7. Compare the characteristics of signals acquired in the field and in test facilities (a) Skewness (b) Kurtosis (c) Impulse factor (d) Frequency.

5. Performance results

5.1. Comparison methods

The focus of this research is the accurate classification of LPMS alarm signals. Using the signals obtained in sections 4.1 and 4.2, we compare them with the results of other classification methods to verify the performance of the proposed model. The other models for comparison are the algorithmically based FR method [16] and common machine learning classification models: SVM, logistic regression classification(LRC) and k-nearest neighborhood(k-NN) approaches. In order to verify the effectiveness of the DAE, we also carry out a comparison of the performance of AE models with the same structure. Table 2 is the details of the models being compared.

5.2. Comparison results of raw and noisy datasets

The accuracy of the proposed model and comparative models are compared using the raw data prepared in section 4.1 and the noisy data in section 4.2, respectively. The size of the input is the same(100 ms) as the signal data obtained from the NPP.

5.2.1. Raw data classification result

Using raw data, the data-driven models perform well overall, with the exception of the algorithm-based FR method. The models that used the AE and DAE are some of the best performing models. Fig. 8(a) shows the results for accuracy performance. On the other hand, in the case of the AE, the performance of the raw data is high regardless of the number of hidden layers, which is likely due to the fact that the raw data shows a good feature, so that even if the layers are shallow, the features can be sufficiently extracted. In Fig. 8(b), we can see the overall classification quality through the F1-score value. It has the same high performance as the accuracy.

5.2.2. Noise data classification result

In Fig. 8, Compared to the raw data results, all models with noise data show lower or similar performance. In particular, AE models showed significant performance degradation. However, in the case of the DAE, the 5-DAE model showed the highest accuracy of 96.73%, and the performance did not deteriorate much compared to the raw signal result in all models. As a result, the excellence of the DAE model can be confirmed for noise data. It was confirmed that the 5-DAE is the best model for raw and noise data.

The results of the noise data classification can be seen in Fig. 8 alongside the raw data results. Overall performance is lower or similar to the raw data results. In particular, AE shows a significant drop in accuracy and F1 score. This is likely due to the fact that AE performs reconstruction on the noiseless data, which allows for good feature extraction on the raw signal, but hinders feature extraction on the noise data.

For the DAE models, on the other hand, the DAE5 model had the highest accuracy at 96.73%, and all models did not degrade significantly with noisy signals. This is because the input is the noisy data and the output is the raw data, so the DAE extracts features while performing the denoising itself. These results demonstrate the robustness of the DAE model to noise data. Furthermore, by confirming that the 5-DAE is the best model for raw and noisy signal data, we can conclude that the feature map of the latent vector extracted from the 5-DAE performs the best. In summary, both AE and DAE perform well when classifying with raw data only, but when classifying with noise data, DAE, which can remove noise, performs better than AE.

Table 2
Description of comparison methods.

Methods	Description
FR method	After converting the data to PSD, compare the area of the PSD graph based on 15 kHz. If the area ratio is greater than 1, it is an impact signal, and if it is less than 1, it is a non-impact signal.
SVM	Classify by selecting the best hyperplane to separate the data, i.e. the hyperplane with the largest margin between the two classes.
LRC	Regression is used to estimate the probability of data falling into a category as a value between 0 and 1, and to classify it into a more likely category based on that probability.
k-NN	Classification into the class assigned to the most frequent class among the k nearest neighbors in the feature space. For majority rule, k is an odd number greater than or equal to 3.
AE	Prepare an AE model with the same structure as the DAE to investigate the effect of denoising. Prepare an AE model with 3,5,7,9 layers, the same as the DAE.

5.3. Robustness of the DAE with noise

To verify the robustness of the DAE model mentioned in Section 5.2, we evaluate the performance of the AE and DAE models on data where the noise is added at the same rate as the raw signal, as shown in Fig. 9(a). For the AE, as the size of the noise increased, a steep decline in the performance of the model is observed.

The DAE, on the other hand, suffers from the same degradation in performance as the AE, but not to the same extent. There is no significant change in performance for the 10 dB noise data, but performance is reduced by 10–20% for the 1 dB noise data. The 5-DAE also performed best when the noise was low, but as the noise increased, the 7-DAE and 9-DAE performed better. This is because the deeper the layer, the more denoising it is and the better the features are extracted.

This is also the case for the AE model, where we can see that the accuracy increases with the number of layers for 10 dB noise data. However, when the level of noise becomes too high, it seems difficult to extract the correct signal features from the data, even as the depth of the layer increases.

5.4. Classification results using DAE-MV

It is confirmed that the DAE model shows the best performance for noise signal discrimination. However, looking at the results of the DAE models, the data misclassified by all DAEs are not the same, and the classification results are different for each model. Therefore, based on the results of the four DAEs (3-, 5-, 7- and 9-), it is expected that the performance improvement can be led through the majority vote classification method.

The performance of the DAE and AE for each number of layers is shown in Fig. 9(b), along with the performance of autoencoder majority vote(AE-MV) classification method and DAE-MV results. DAE outperformed the AE on all models for noise data. The use of majority voting also appears to improve classification performance for both DAE and AE. This means that the DAE-MV performed best, and even the AE model, which performed poorly on noise data, shows a large improvement in performance when classified by the majority voting method of multiple models.

In Fig. 10, the confusion matrix is used to visualize the discrimination results of the FR method, 5-DAE, which performed the best among the single DAE models, and the DAE-MV. The FR method, with its low accuracy, often resulted in incorrectly resolved valid and false alarms. Of the 368 and 367 valid and false alarms respectively, the 5-DAE correctly classified 356 and 355. Finally, DAE-MV correctly distinguished 360 and 364, respectively, with 98.5% accuracy.

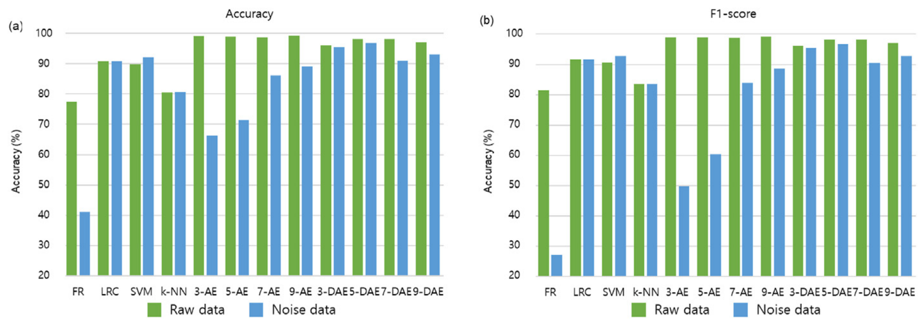

Fig. 8. Performance of classification model (a) Accuracy (b) F1-score.

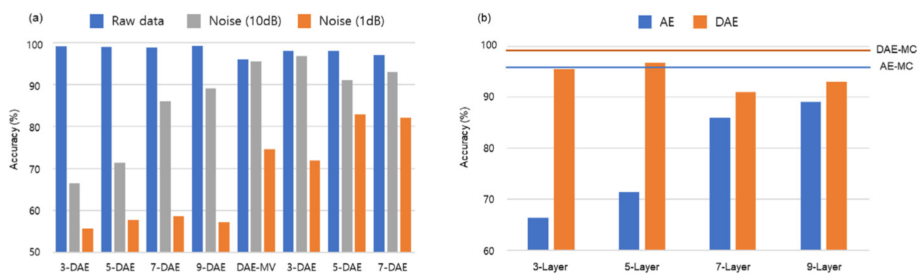

Fig. 9. (a) Accuracy of AE and DAE for confirming the robust of model (b) Comparison of AE and DAE accuracy according to numbers of layers with DAE-MC accuracy.

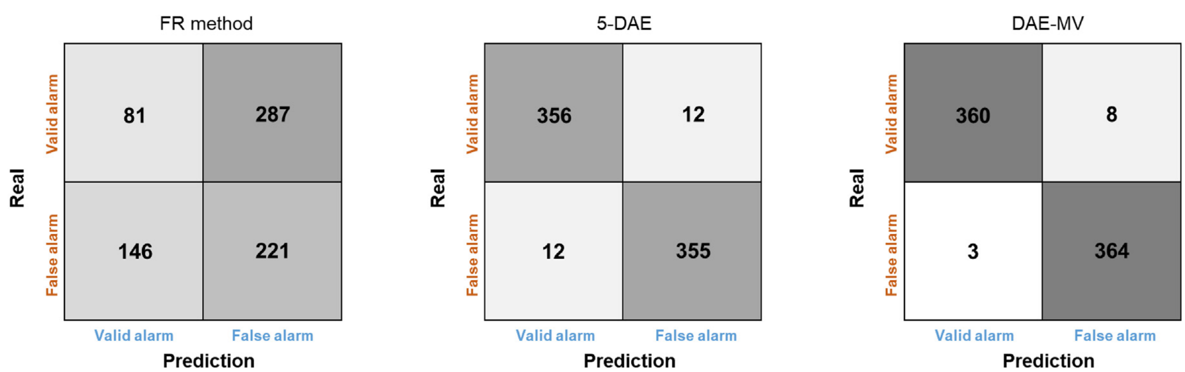

Fig. 10. Confusion matrix of classification models (FR method, 5-DAE, and DAE-MC.

5.5. Performance metric result

Using the performance metrics described in Section 3.3, we check the performance of the proposed model and the comparison model. The results are shown in Table 3, which examines the performance of the model in terms of accuracy, F1 score and FAR. Overall, the FR method is the worst performing method with noise data. The machine learning-based classification models (LRC, SVM, and k-NN) had good accuracy and f1-score, but a relatively high false alarm rate.

In the AE and DAE classification results, AE model performs better as the number of hidden layers increases, but the 9-AE model has a high false alarm rate despite its high accuracy. This means that accurate feature information can be extracted with increasing depth, but the features of valid and false alarm become more similar with depth, which seems to cause false alarm to be misclassified as valid alarm. In addition, the false alarm rate of the 3-AE and 5-AE is low, which appears to be a good model, but the F1 score and accuracy are also low. This is because the data that should have been determined as valid alarm are determined as false alarm. It is therefore not a suitable classification model.

Table 3
Performance metric result (SNR = 10 dB).

Model	Performance metric (%)		
	Accuracy	FAR	F1-score
FR method	41.14	39.78	27.23
LRC	90.75	18.26	91.52
SVM	92.11	15.8	92.69
3-NN	80.54	37.87	83.58
3-AE	66.39	0.54	49.89
5-AE	71.43	0.54	60.38
7-AE	85.99	1.09	83.93
9-AE	89.12	6.27	88.6
3-DAE	95.51	2.18	95.41
5-DAE	96.73	3.27	96.74
7-DAE	91.02	2.72	90.43
9-DAE	93.06	1.91	92.7
AE-MV	96.33	2.72	96.3
DAE-MV	98.5	0.82	98.5

In the case of the DAE, all the models have an accuracy and F1-score above 90% and FAR below about 4%. This is an indication that classification can be very robust to noise data. This allows for very robust classification against noisy signals, and we found that DAE-MV, composed of multiple DAE models, performed well across all metrics. It has a FAR of around 0.82% and an F1 -score of 98.5%, with an accuracy of 98.5%. It can be concluded that the DAE-MV model is the best performing classification model.

6. Conclusion

Under steady-state, no signals other than background noise are picked up by the sensors attached to the NPP. So, when a special signal such as an impact occurs, it makes the right decision and is highly accurate with few false alarms. However, many false alarms can be caused by temperature and pressure fluctuations during power plant start-up and shutdown. It's difficult to distinguish false alarms from valid ones. This is because the comparison values and shapes of the valid and false alarm signals are very similar. In addition, it can be difficult to accurately detect and identify alarm signals due to the background noise generated during operation.

This study proposes a denoising autoencoder-based majority vote network method for alarm signal discrimination, which can distinguish alarm signals contaminated by noise. Through each DAE, the noise signal without preprocessing is used as the input. Also, the feature space that allows the signals of each class to be well distinguished is well extracted by the DAE. Finally, the alarm type is determined using the results of each DAE as input to the majority classification network.

The proposed model classifies valid and false alarms with high probability and is able to distinguish between false alarms that have not been distinguished by existing algorithms and other classification models. Therefore, by reducing the false alarm rate to 0.82%, the proposed model is expected to reduce unnecessary false alarms, improve system performance, and reduce operator fatigue and costs compared to the existing method.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20206510100010).

References

[1] B. Bechtold, U. Kunze, KUES'95-The modern diagnostic system for loose parts monitoring, Prog. Nucl. Energy 34 (3) (1999) 221–230.
[2] C.W. Mayo, Loose parts signal theory, Prog. Nucl. Energy 15 (1985) 535–543.
[3] C.W. Mayo, Loose-part signal properties, Prog. Nucl. Energy 28 (4) (1994) 347–357.
[4] J.S. Kim, I.K. Hwang, J.T. Kim, J. Lyou, Development of automatic algorithm for localizing loose parts with a steam generator, Nucl. Eng. Des. 219 (3) (2003) 269–276.
[5] Y.W. Chang, J.C. Jung, P.H. Seong, Algorithm automation for nuclear power plant loose parts monitoring system, Nucl. Eng. Des. 231 (1) (2004) 99–107.
[6] S.H. Shin, J.H. Park, D.B. Yoon, Y.C. Choi, Mass estimation of impacting objects against a structure using an artificial neural network without consideration of background noise, Nucl. Eng. Technol. 43 (4) (2011) 343–354.
[7] L.X. Fang, T.T. Ji, F. Zeng, W. Zhang, Y.C. Xie, C.H. Wang, K.F. Zhang, A study on the method of impact mass estimation of loose parts, Prog. Nucl. Energy 70 (2014) 242–248.
[8] J. Liska, S. Künkel, Localization of loose Part Impacts on the general 3D surface of the nuclear power plant coolant circuit components, Prog. Nucl. Energy 99 (2017) 140–146.
[9] S. Moon, S. Han, T. Kang, S. Han, M. Kim, Model-based localization and mass-estimation methodology of metallic loose parts, Nucl. Eng. Technol. 52 (4) (2020) 846–855.
[10] S. Moon, S. Han, T. Kang, S. Han, K. Kim, Y. Yu, J. Eom, Impact parameter prediction of a simulated metallic loose part using convolutional neural network, Nucl. Eng. Technol. 53 (4) (2021) 1199–1209.
[11] Y. Choi, J. Park, K. Choi, An impact source localization technique for a nuclear power plant by using sensors of different types, ISA Trans. 50 (1) (2011) 111–118.
[12] S. Shin, J. Park, D. Yoon, S. Han, T. Kang, Markov chain-based mass estimation method for loose part monitoring system and its performance, Nucl. Eng. Technol. 49 (7) (2017) 1555–1562.
[13] T. Tsunoda, T. Kato, K. Hirata, Y. Sekido, K. Sendai, M. Segawa, O. Tsuneoka, Studies on the loose part evaluation technique, Prog. Nucl. Energy 15 (1985) 569–576.
[14] G.L. Zigler, G.A. Morris, R.C. Roberts, On-line loose parts monitoring false alarm discrimination, Prog. Nucl. Energy 15 (1985) 577–581.
[15] C.W. Mayo, H.G. Shugars, Loose part monitoring system improvements, Prog. Nucl. Energy 21 (1988) 505–513.
[16] S. Figedy, G. Oksa, Modern methods of signal processing in the loose part monitoring system, Prog. Nucl. Energy 46 (3–4) (2005) 253–267.
[17] J.S. Kim, I.K. Hwang, J.T. Kim, B.S. Moon, J. Lyou, A study on loose part monitoring system in nuclear power plant based on neural network, International journal of fuzzy logic and intelligent systems 2 (2) (2002) 95–99.
[18] Y. Cao, Y. He, H. Zheng, J. Yang, An alarm method for a loose parts monitoring system, Shock Vib. 19 (4) (2012) 753–761.
[19] M.G. Min, C.G. Jeong, J.K. Lee, S.H. Jo, H.J. Kim, Development of an algorithm to discriminate between valid and false alarms in a loose-parts monitoring system, Nucl. Eng. Des. 278 (2014) 1–6.
[20] J. Yang, B. Yang, M. Liu, Y. Cao, Detection method of loose parts in nuclear reactor based on eigenvector algorithm, Int. J. Fuzzy Logic And Int. Syst. Pro.
[21] J. Meng, Y. Su, S. Xie, Loose parts detection method combining blind deconvolution with support vector machine, Ann. Nucl. Energy 149 (2020), 107782.
[22] J.S. Kim, I.K. Hwang, T.W. Kim, J. Lyou, An automatic diagnosis method for loose parts monitoring system, in: ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 01TH8570), vol. 3, IEEE, 2001, pp. 1971–1977.
[23] C. Lu, Z.Y. Wang, W.L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Process. 130 (2017) 377–388.
[24] M. Gan, C. Wang, Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings, Mech. Syst. Signal Process. 72 (2016) 92–104.
[25] D. Verstraete, A. Ferrada, E.L. Droguett, V. Meruane, M. Modarres, Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings, Shock Vib. (2017).
[26] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Signal Process. 95 (2017) 187–204.
[27] F. Jia, Y. Lei, L. Guo, J. Lin, S. Xing, A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines, Neurocomputing 272 (2018) 619–628.
[28] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Process. 100 (2018) 439–453.
[29] H.B. Yang, J.A. Zhang, L.L. Chen, H.L. Zhang, S.L. Liu, Fault diagnosis of reciprocating compressor based on convolutional neural networks with multisource raw vibration signals, Math. Probl Eng. (2019).
[30] P. Vincent, H. Larochelle, Y. Bengio, P.A. Manzagol, Extracting and composing robust features with denoising autoencoders, in: Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 1096–1103.
[31] C.J. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. 30 (1) (2005) 79–82.