



Source localization on large-scale canisters for used nuclear fuel storage using optimal number of acoustic emission sensors

Li Ai^a, Vafa Soltangharai^a, Mahmoud Bayat^{a,*}, Bruce Greer^b, Paul Ziehl^a

^a Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC 29208, USA

^b Electrical Power Research Institute, USA

ABSTRACT

The dry cask storage system (DCSS) canisters have been used for the storage of high-level nuclear for decades. The inspections are needed to ensure that structural integrity is maintained. One mechanism of degradation on DCSS canisters that is of interest is stress corrosion cracking (SCC). Acoustic emission (AE) is a nondestructive technique that can be employed as an inspection approach since it can offer real-time degradation detection. This paper presents the approaches that can localize SCC sources by minimal acoustic emission (AE) sensor. To achieve this goal, three machine learning techniques (artificial neural network, random forest, stacked autoencoders) were adopted to improve the conventional source localization approach. In this paper, source localization is treated as a classification problem. The testing specimen was divided into multiple zones and located the AE signals to their corresponding zones. The AE signals were processed to create two datasets: a dataset consisting of AE parametric features and a dataset consisting of AE waveforms. Source localization approaches using artificial neural networks, random forest, and stacked autoencoders were trained and tested based on the datasets. The results show all three machine learning techniques can learn to map AE signals to their sources. Among them, stacked autoencoders have the best performance (97.8% accuracy of stacked autoencoders versus 91.5% of random forest, and 80.0% of ANN), demonstrating that it could be a potential approach to localize SCC events on DCSS canister.

1. Introduction

Nuclear power generation has been widely applied in the United States for decades (Xie and Zhang, 2015). Currently, spent fuel is stored in cooling pools and dry cast storage systems (DCSS). DCSS use was initiated in the 1970 s. Spent fuels are placed in stainless-steel canisters, then water and air are removed and replaced by an inert gas. These dry storage systems were originally licensed for an operation period of 20 years. With the continued delays in the opening of a functional repository for the storage of these materials, the systems that are currently in operation now will be required to be operational for a significantly longer time. To extend the operation license, the inspections of the DCSS canister are needed to ensure the structural operability. Stress corrosion cracking (SCC) has been identified as the main degradation mechanism of concern on DCSS canisters because of the high salinity and humidity in the coast region where those DCSS canisters are placed (Hill, 2018; Yeom et al., 2020; Wu, 2020). Therefore, the in-situ examination for the detection of SCC defects is desirable.

Acoustic emission (AE) is a nondestructive structural health monitoring technique (Li, 2002; Ono, 2011; Anay et al., 2018; Soltangharai et al., 2018, 2020; Li et al., 2018, 2020; Ai et al., 2019, 2020). It has been

widely used to detect cracks in the infrastructures such as bridges (Anay et al., 2020), dams (Wang et al., 2019), and nuclear facilities (Nozawa et al., 2014; Véronique et al., 2015; Baek et al., 2018). Recently, applications of AE have been investigated to detect stress corrosion cracking. Soltangharai et al. (Soltangharai et al., 2020) utilized AE and pattern recognition to identify the AE signal signatures caused by the propagation of SCC in DCSS. A small-scale 304 stainless steel plate was employed instead of testing on the real-scale DCSS canister. The results indicated that AE monitoring has a good capability to detect and identify SCC events.

However, there is a problem when applying AE monitoring in the realistic DCSS canister. The canisters are very large and are stored in a concrete overpack. Fig. 1 shows a cutaway mockup of a DCSS canister.

The available area for sensor attachment is limited. In a recent study about the AE monitoring of DCSS canister, sensors were placed on the bottom support structure of the canister (EPRI, 2016). However, SCC crack would usually occur on the canister far away from the bottom. It is difficult to employ an array of four or more AE sensors around the cracking region to detect the location of a crack using the time of arrival approach. Machine learning techniques could be alternative methods to solve the source localization problem in this situation.

* Corresponding author.

E-mail addresses: ail@email.sc.edu (L. Ai), vafa@email.sc.edu (V. Soltangharai), mbayat@mailbox.sc.edu, mbayat14@yahoo.com (M. Bayat), bgreer@epri.com (B. Greer), ziehl@cec.sc.edu (P. Ziehl).

<https://doi.org/10.1016/j.nucengdes.2021.111097>

Received 30 August 2020; Received in revised form 11 January 2021; Accepted 20 January 2021

0029-5493/© 2021 Elsevier B.V. All rights reserved.

Artificial neural network (ANN) is a field of interest in machine learning (Basheer and Hajmeer, 2000). Soltangharai et al. (Soltangharai et al., 2019) utilized a back-propagation neural network to localize the impact on aircraft components. Acoustic emission events are collected by a single AE sensor during the impact experiment. AE features such as energy, amplitude, and signal strength were adopted as the neural network's input. The outputs were zonal source localization results. The results showed that source localization of AE events using back-propagation neural network has the capability to acquire accurate localization results. However, the problems of local optimal solutions and explosion gradient (Bengio et al., 1994) have become the main bottlenecks of ANN. Moreover, a good dimension reduction or feature selection is usually needed. Therefore, when applying ANN to the analysis of AE data, appropriate parametric AE features are supposed to be extracted from the original AE signals.

Random forest is one of the present state-of-the-art classifiers based on ensemble learning strategy (Breiman, 2001). It has fast training speed, and it is robust to the number of training samples. Moreover, it can provide the ranking of features importance, which means a good feature selection can be obtained instead of manual selection. The random forest has a wide application in fault diagnose. Cerrada et al. (Cerrada et al., 2016) built a robust system for the multi-class fault diagnosis in spur gears using genetic algorithms and random forest. An acceptable diagnose accuracy was obtained based on the real vibration signals. Patel et al. (Patel and Giri, 2016) presented a random forest classifier as an approach for the classification of bearing fault and feature selection. The most important features of vibration signals were selected and assigned to the random forest model. Results indicated the random forest is turn out to be a suitable approach for fault diagnosis of any rotating machine.

Stacked autoencoders is a deep learning algorithm composed of multi-layer autoencoders. (Bengio et al., 2007). The concept of deep learning stems from the study of artificial neural networks. The most attractive advantage of deep learning is that no feature selection or dimension reduction is needed (Arel et al., 2010). Raw data can be utilized as input. Deep learning combines low-level features to form more abstract high-level features to discover the distributed feature representations of data. A stacked autoencoder has been utilized for AE source localization. Karvelis et al. (Karvelis et al., 2020) studied the structural health monitoring of ship hulls using the acoustic emission method. A stacked autoencoder neural network was employed to obtain

the locations of AE events induced by SCC on ship hulls. The source localization approach was validated, and the results indicate that the method can be very effective and efficient.

The goal of this paper is to propose the machine learning-based source localization approaches for large-scaled steel structures like DCSS canister. The Source localization approaches based on ANN, random forest, and stacked autoencoder are proposed and results compared.

2. Test setup and experimental procedure

The primary objective of the experiments was to examine the capability of the proposed source localization techniques on DCSS canister when a single AE sensor is employed on the bottom edge. To simulate a realistic experimental environmental situation for the test setup, a test specimen was made of similar length and thickness (Industry Spent Fuel Storage Handbook, 2010) to what is typical of DCSS canister shells and of the greatest width that could be managed. This resulted in a specimen size of $5029 \times 1524 \times 16$ mm. The plan view of the specimen is provided in Fig. 2. The specimen is fabricated from 304/304H stainless steel.

AE is a physical phenomenon related to stress waves generated by the rapid release of elastic energy when cracks or damage are formed in materials (Wadley and Mehrabian, 1984; Grosse et al., 2008). By attaching AE sensors to the surface of an object, AE signals can be detected and recorded. The technique of collecting and analyzing AE signals to diagnose the status of an object is referred to as AE monitoring (Scruby, 1987). The Hsu-Nielsen pencil lead break (Hsu, 1977) is one of the widely used artificial sources to generate AE signals by conducting pencil lead break on the object to which AE sensor is attached. In this paper, A Hsu-Nielsen pencil lead break test was conducted to simulate the cracks that usually initiate on the canister. 135 points were set up on the specimen (marked as red dots in Fig. 3). The Hsu-Nielsen pencil lead break was repeated 30 times on each of these points. The AE sensor was attached to the upper left corner of the specimen to simulate the circumstance that the sensor is placed on the bottom support of a DCSS canister. 4050 AE events were collected during the experiment.

To determine the appropriate AE sensor applied in the study. An attenuation test has been conducted on a resonant sensor (type R3I-AST) with a frequency response range of 10–40 kHz, a resonant sensor (type R6I-AST) with a frequency response range of 40–100 kHz, and a broadband sensor (type WDI-AST) with a frequency response range of



Fig. 1. Cutaway mockup of simulated dry storage cask system.

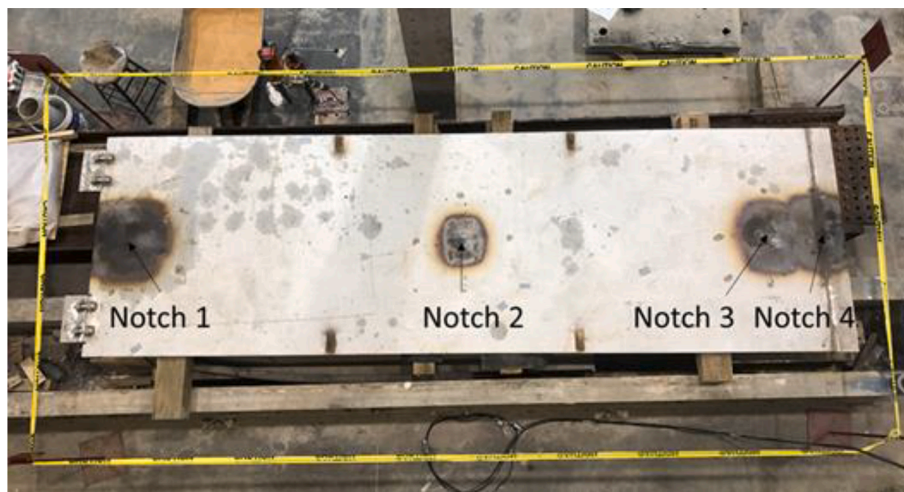


Fig. 2. Plan view of the specimen.

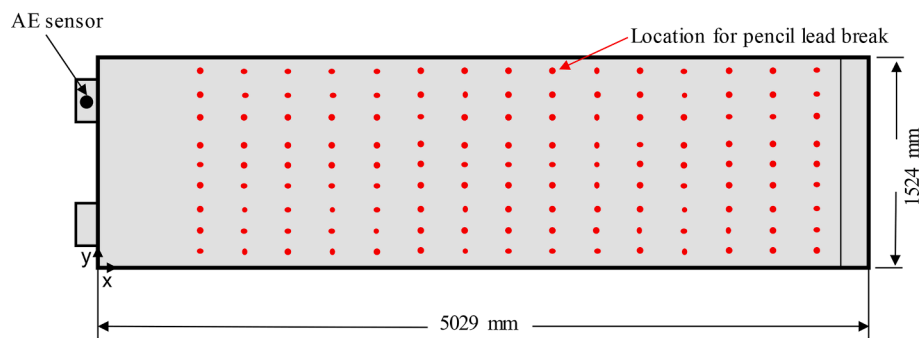


Fig. 3. Locations of AE sensor and pencil lead breaks.

100–900 kHz. Pencil lead breaks were conducted on the specimen with a distance to the sensor from 0–5000 mm. The results are provided in Fig. 4.

It can be observed, R3I-AST sensor has the highest sensitivity in long distance. Therefore R3I-AST sensor was employed in this study to ensure the events far away from the sensor can be detected. The calibration of the AE sensor was conducted by applying Hsu-Nielsen pencil lead break beside the AE sensor when it was attached on the steel plate. The standard of a reliable and sensitive AE sensor is that the acquired amplitude of the waveform is close to 80 dB for the broadband sensor, and 90 dB for resonating sensor. It can also be observed in Fig. 4, when the distance is close to 0 mm, the amplitude of the resonate sensors and the broadband

sensor reached 90 dB, and 80 dB. The Data collected from the sensor was obtained using a 16-channel PCI digital signal processing (DSP) system (manufactured by Mistras Group, Inc. of Princeton Junction, New Jersey). The pre-trigger time, a setting in the software, which recovers acoustic waveforms prior to crossing the threshold, was set to 256 μ s. The sampling rate was set to 1 MHz (or 1,000,000 samples per second). The time from threshold crossing to peak amplitude, peak definition time, was set to 200 μ s. The hit definition time, which determines when to stop recording a hit, was set to 400 μ s. Its value typically twice the peak definition time. The hit lockout time, which minimized the recording of late arrival signals and reflected hits, was set to 200 μ s (Laksimi et al., 1999).

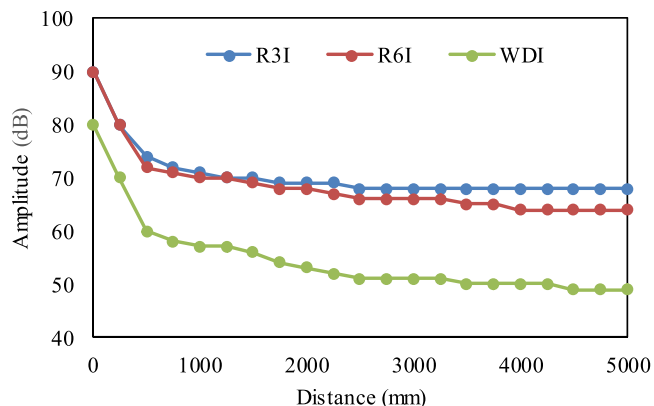


Fig. 4. Attenuation curves.

3. Data collection

3.1. Zone code

The source localization approaches proposed in the paper belong to the family of the paper are the zonal localization method, and AE events will be localized to their corresponding zones. In this paper, the AE events collected in the experiment were divided into five zones. Fig. 5 shows the zonal divisions. From left to right, the AE events on the specimen are divided into zones 1, zones 2, zones 3, zones 4, and zones 5. There are 810 AE events in each of the zones.

3.2. Feature-based data

AE features are used to reduce the amount of information carried by the AE signal to a specific value (Ali et al., 2019). Those features can

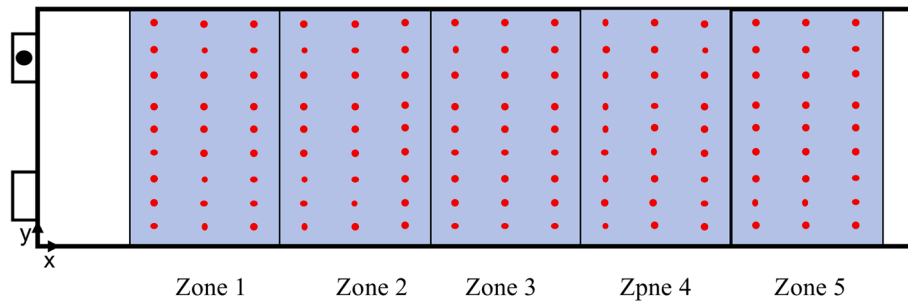


Fig. 5. Zone codes.

represent the characteristics of AE signals in several aspects. For example, “Amplitude” refers to the maximum amplitude at the peak, “Counts” indicates the number of threshold crossings, “Rise time” represents the time interval between the first threshold crossing and maximum amplitude. Fifteen features were extracted from the AE events collected in the experiment. The names and descriptions of the features are provided in Table 1. The AE data after feature extraction forms the feature-based dataset, which contains 4050 data, and each data has 15 sample points. The feature-based data was utilized as the input for source localization approaches using ANN and random forest, which are introduced in Section 4.1 and Section 4.2. The zone codes introduced in Section 3.1 were utilized as labels during training and testing processes.

3.3. Waveform-based data

During AE data acquisition, the sampling rate was set to 1 MHz, and the duration was set to 2000 μ s. An AE waveform collected during the impact test, therefore has 2000 sample points. A dataset consisting of original AE waveforms was constructed based on AE events from the experiment. In the waveform-based dataset, there are 4050 AE waveforms, and each one is a one-dimensional series with 2000 sample points. This dataset was utilized as the input for the source localization approach using stacked autoencoders. The zone codes introduced in Section 3.1 were utilized as labels during training and testing processes. In each of the five zones, 810 AE signals of pencil lead break were collected. A single signal was randomly selected from the 810 AE signals

Table 1
Descriptions of the input features for random forest.

Features	Descriptions
Amplitude (dB)	The maximum amplitude at the peak
Count	The number of threshold crossings
Rise time (μ s)	Time interval between first threshold crossing and maximum amplitude
Duration (μ s)	Time between first and last threshold crossing of signal
Average frequency (kHz)	Counts/Duration
Root mean square (RMS) (V)	The effective voltage with a characteristic time T_{RMS} for average ranging from 10 to 1000ms
Average signal level (ASL) (V)	The effective voltage with a characteristic time T_{ASL} for average ranging from 10 to 1000ms
Energy	The measure of the electrical energy measured for an AE signal
Absolute energy	The absolute measure of the electrical energy measured for an AE signal
Peak frequency (kHz)	Frequency of maximum signal contribution
Reverberation frequency (kHz)	Frequency after the peak
Initial frequency (kHz)	Frequency before the peak
Signal strength	A parameter to evaluate the AE source strength
Frequency centroid (kHz)	A parameter to characterize the overall frequency content of an AE signal
Counts to peak (PCNTS)	The number of threshold crossings from the first threshold crossing to the peak

of each zone and shown in Fig. 6. It can be observed that the AE waveforms in different zones have different data distribution patterns. The stacked autoencoders can learn the patterns of waveforms and classify them into the corresponding zones.

All waveforms in the waveform-based dataset were normalized by amplitude to a range of -1 to 1 . The specimen is large, and the amplitude of the waveforms therefore varies greatly in different zones due to attenuation. If a dataset without normalization is used, the network will focus on the amplitude while other important characteristics may be ignored, significantly impacting the results. To minimize dimensional influences between the waveforms, normalization processing is therefore needed (Jayalakshmi and Santhakumaran, 2011). After the original waveforms are normalized, all characteristics are in the same order of magnitude for comprehensive comparative evaluation.

4. Methods

4.1. Back-Propagation neural network

Artificial neural networks (ANNs) are information processing systems that mimic how the human brain processes information (Hassoun, 1995). The neural network adopted in this section is a back-propagation (BP) network; it consists of an input layer, hidden layers, and an output layer, each layer has many processing elements, called neurons, and each neuron is connected to each other. The number of neurons in the input layer and the output layer corresponds to the number of variables and the number of outputs. Fig. 7 shows a simple three-layer artificial neural network consisting of layer j , i , and k . The number of neurons is m for layer j , n for layer i , and l for layer k . $W_{(ij)}$ and $W_{(kj)}$ are weights between layers. The values of m and l are related to the problem for solving, and n is determined by the network designer.

The performance of an ANN model depends on the configuration of the network, including the number of neurons in hidden layers and activation functions for each layer. The number of neurons in hidden layers is determined using a trial-and-error method (Sun et al., 2020). But some guidelines have been developed to decide the upper limit without losing fidelity in approximating. One commonly used method in determining cap for neurons in hidden layers is:

$$N_{hidden} \leq 2N_{input} + 1 \quad (1)$$

where N_{hidden} is the number of neurons in the hidden layers and N_{input} is the number of input variables. However, to avoid overfitting the training data, the number of neurons should also be determined with consideration of the training sample size. Rogers et al. (Rogers and Dowla, 1994) commend the following relationship:

$$N_{hidden} \leq \frac{S_{train}}{N_{input} + 1} \quad (2)$$

where S_{train} is the sample size of training data. Here, we determine the upper limit for the number of hidden layer neurons as the smallest of the values for N_{hidden} calculated by Eq. (1) and (2).

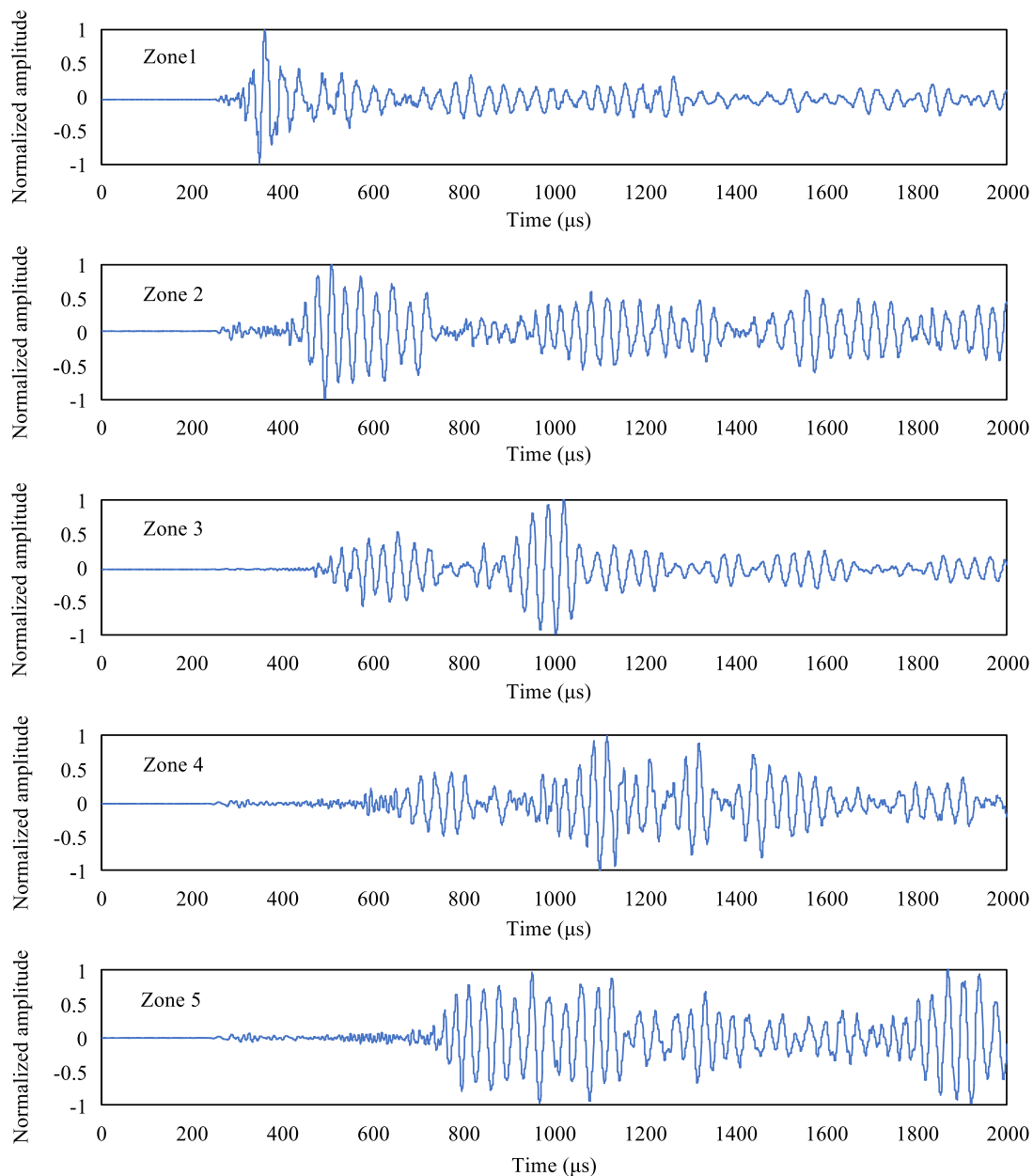


Fig. 6. The waveform in each zone.

A trial-and-error test was conducted for the selection of ANN model configuration. The process is shown in Table 2. The input data of the network is the feature-based AE data introduced in Section 3.2. The zone codes are presented in Section 3.1 were utilized as the labels of the input. It is noticed that the highest accuracy is 80.1%, whose corresponding configuration is two hidden layers with 20 neurons for each of the hidden layers while the lowest (76.6%) accuracy was observed when the corresponding configuration was one hidden layer with 10 neurons. There is some evidence of lower accuracy when neuron numbers reach the upper limit, and generally, networks with 2 hidden layers have better performance than counterparts in one hidden layer model. In this study, two hidden layers with 20 neurons for each of the hidden layers were selected as the configuration of the ANN in AE source localization.

4.2. Random forest

Random forest (RF) is an ensemble learning method containing multiple decision trees (Liaw and Wiener, 2002). Decision tree models

are trained to be independently, and the results produced by these models are put together, with the final prediction receiving the most votes. At present, the mainstream decision tree algorithm includes C4.5 and classification and regression trees (CART). C4.5 is a decision tree algorithm proposed by Quinlan et al. (Quinlan, 2014). C4.5 builds decision trees from a set of training data using the concept of information entropy. In a C4.5 tree, each node can be branched into multiple sub nodes, the combination of features is not supported in the nodes. C4.5 can only be used for classification problems (Sathyadevan and Nair, 2015). For the CART decision tree, each node is branched into only two sub nodes, supporting a combination of features, and can be used for classification and regression problems (Rutkowski et al., 2014). In the random forest, the CART decision tree is a commonly used approach. In this paper, Gini impurity was adopted as the criterion for the branching CART decision tree.

The steps to create a random forest are as follow:

(1) Sample randomization: Assume there is an original dataset named T , which has N samples. By using the bootstrapping method, N

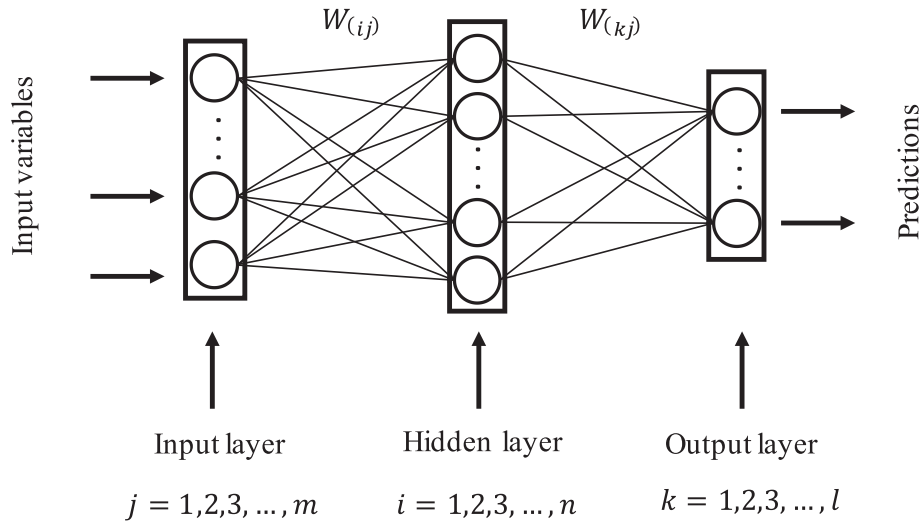


Fig. 7. Three-layer artificial neural network.

Table 2
Description of Artificial Neural Networks Configuration Selection Process.

Hidden layers	Neurons		Accuracy (%)
	1st hidden	2nd hidden	
1	10	N/A	76.6
	20	N/A	78.9
	31	N/A	77.7
2	10	10	76.9
	20	20	80.1
	31	31	79.1

samples are taken from the original dataset T with replacement and form a new subset. These N samples in the new subset may contain a sample that has been taken many times or a sample that has never been taken. The probability that a sample has never been taken can be obtained by Eq. (3):

$$h(N) = \left(1 - \frac{1}{N}\right)^N \quad (3)$$

The limit of the probability can be calculated by Eq. (4):

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = 0.368 \quad (4)$$

According to Eq. (4), nearly 36.8% of the data in the original data set may not appear in the new subset. This unselected data is called out-of-bag (OOB) data, which can be used to test the decision tree's generalization performance.

(2) Feature randomization: Assume each sample in the new subset has n features. t features ($t \leq n$) are randomly selected and transfer to the decision tree. By calculating the information contained in each feature, a feature with the most classification ability is selected for node branching. Genuer et al. (Genuer et al., 2010) recommend the following relationship between n and t :

$$t \approx \sqrt{n} \quad (5)$$

(3) Create a decision tree: By repeating steps (1) and (2) for m times, m subsets that contain t features in each sample can be obtained. Each subset is transferred to an individual decision tree. In other words, m decision trees are created.

(4) Form the random forest: the m trees are formed into a random forest. The decision trees inside the random forest generate their own classification results. The final results are determined according to the number of votes.

An advantage of the random forest is the ability to assess the importance of features (Menze et al., 2009). The calculation requires the help of the Gini impurity. By adding noise to one of the features. The new Gini impurity is obtained and compare with the value of the old Gini impurity before adding the noise. The difference between the Gini impurities is utilized as a measure of the importance of this feature.

In this paper, a random forest classification model contains 800 decision trees was adopted for AE data source localization. The feature-based AE dataset, which was introduced in Section 3.2, was utilized for training and to test this approach. By using the Bootstrapping method, 800 subsets were drawn from the original AE dataset. Each sample in the new subsets has 4 features that were randomly selected from the original 15 features. The number of features in the new subset was determined by Eq (6). The subsets were assigned the decision trees. These decision trees inside the random forest work independently and generate their own localization results. The final result is given by voting. Fig. 8 shows the random forest used in this paper. The output of the random forest model is the zoning code of the corresponding AE event.

Generally, the error of random forest will decrease with the increase of the number of decision trees. Once the number of decision trees

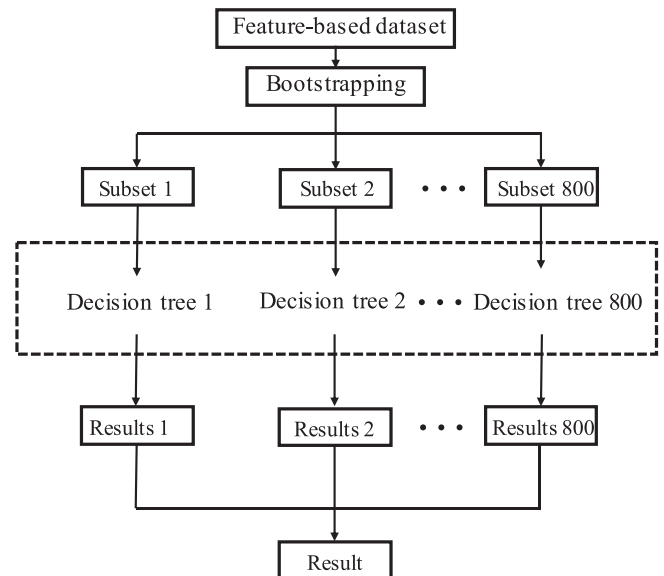


Fig. 8. The architecture of the random forest.

increased to a certain number, the error of the model will converge to a certain value. Continue to increase the number of decision trees will not reduce the error but increase the computational cost of the entire random forest. Therefore, an appropriate number of decision trees is critical to the optimization of the random forest. To determine the number, the random forest model was tested with tree numbers varying from 200 to 2000, with an interval of 200. The zoning code introduced in Section 3.1 was utilized as the labels of the input feature-based AE data. The determination process of the number of trees is shown in Table 3. The classification mistake rate keeps decreasing before the number of trees reaches 800, while no significant change of classification error can be observed when the number of trees continually increased to 2000. However, a small fluctuation can still be observed, and this is because, in a random forest, each decision tree randomly selects training samples in a given training dataset based on bootstrapping. The decision trees within two random forests (e.g., random forests with 800 trees and 1000 trees) are different, which results in a small fluctuation when the tree number increase from 800 to 2000. While the general trend is towards a plateau. Comprehensively considering all the above, the number of decision trees is defined as 800 in this paper.

4.3. Stacked autoencoder

Autoencoder is a three-layer neural network model with an input layer, hidden layer, and output layer (Ng, 2011). The network structure is shown in Fig. 9. The input layer and the output layer have the same dimension, while the hidden layer has a smaller dimension.

Assuming the input data is a n dimensional vector $\{x^1, x^2, x^3 \dots x^n\}$, the process of mapping the input data to the m ($m < n$) dimensional vector $\{h^1, h^2, h^3 \dots h^m\}$ in the hidden layer through the nonlinear encoding function E is named as encoder stage. The process of mapping the m dimension vector in the hidden layer to the n dimension vector $\{x'^1, x'^2, x'^3 \dots x'^n\}$ in the output layer through the decoding function D is named as decoder stage. The encoding function E and the decoding function D are presented in Eqs. (6) and (7):

$$h = E(x) = S_{\theta}(wx + b) \quad (6)$$

$$x' = D(h) = S_{\theta'}(w'h + b') \quad (7)$$

Where, $\{\theta, \theta'\} = \{w, w', b, b'\}$ is the mapping parameter set in the autoencoder. w and w' are the weights of encoding and decoding stages; both are $m \times n$ dimensional matrix. b and b' are the n dimensional bias vectors of encoding and decoding stages. S is the activation function. The activation function for the autoencoder in this paper is a sigmoid function (Eq. (8)):

$$f(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

The process of transferring the input vector $\{x^1, x^2, x^3 \dots x^n\}$ to the output vector $\{x'^1, x'^2, x'^3 \dots x'^n\}$ is called reconstruction. The training

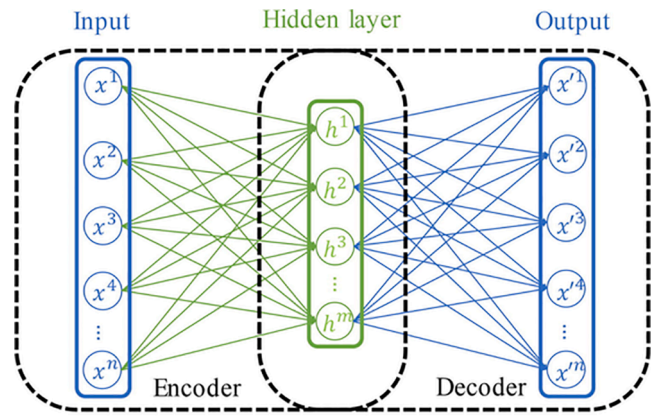


Fig. 9. Scheme of autoencoder.

object of the autoencoder is to minimize the error in data reconstruction by constantly adjusting the mapping parameters set $\{\theta, \theta'\} = \{w, w', b, b'\}$. A training process of N iterations can be expressed by Eq. (9):

$$\theta, \theta' = \operatorname{argmin} \frac{1}{N} \sum_{i=1}^N L(x^i, x'^i) = \operatorname{argmin} \frac{1}{N} \sum_{i=1}^N \|x^i - x'^i\|^2 \quad (9)$$

Where, x^i and x'^i refer to the i th element in the input and output vector. L is the mean squared error between x^i and x'^i .

By minimizing reconstruction errors, the vector in the input vector; in the meantime, the dimension is significantly reduced (Rogers and Dowla, 1994). Therefore, the vector $\{h^1, h^2, h^3 \dots h^m\}$ in the hidden layer can be seen as the feature set extracted from the input vector.

The classification stacked autoencoder consists of multiple autoencoders and a softmax layer (Tao et al., 2015). The first autoencoder extracts the feature set of the input data and takes the obtained feature set as the input data of the next autoencoder for further feature extraction. All the training processes in autoencoders are unsupervised training, and no labels are needed in this stage. The feature set from the last autoencoder is used as the input of the softmax classifier for supervised training. The corresponding classes of each input data are required to be known as the labels. Assume the inputs contain k classes. The final output of a stacked autoencoder is a k dimensional vector; each element represents the probability that the input belongs to this class. The class with the highest probability can be considered as the classification result.

The stacked autoencoder network utilized in this paper has two autoencoders. The first autoencoder has a hidden size of 100, and the second autoencoder has a hidden size of 50. Different from ANN and random forest, the input data of the network is the waveform-based AE data introduced in Section 3.3. The corresponding zone code of the AE data is utilized as the labels in the training of the softmax layer. A source localization result is obtained as the output of this stacked autoencoder neural network. Fig. 10 shows the structure of the stacked autoencoder applied in this paper.

5. Results and discussion

5.1. Source localization using BP-ANN

The feature-based AE dataset was adopted in the scenario of source localization using ANN. In the 4050 data collected during the experiment, 2025 AE data were randomly selected for training, 675 AE data were randomly selected for validation. The remaining 1350 data were utilized to test the performance of the trained ANN. The source localization results are shown in Fig. 11a as a confusion matrix. The numbers

Table 3
Description of the selection process of trees number.

Trees number	Error (%)
200	14.7
400	11.8
600	10.1
800	9.5
1000	9.8
1200	9.9
1400	9.6
1600	9.5
1800	10.1
2000	9.9

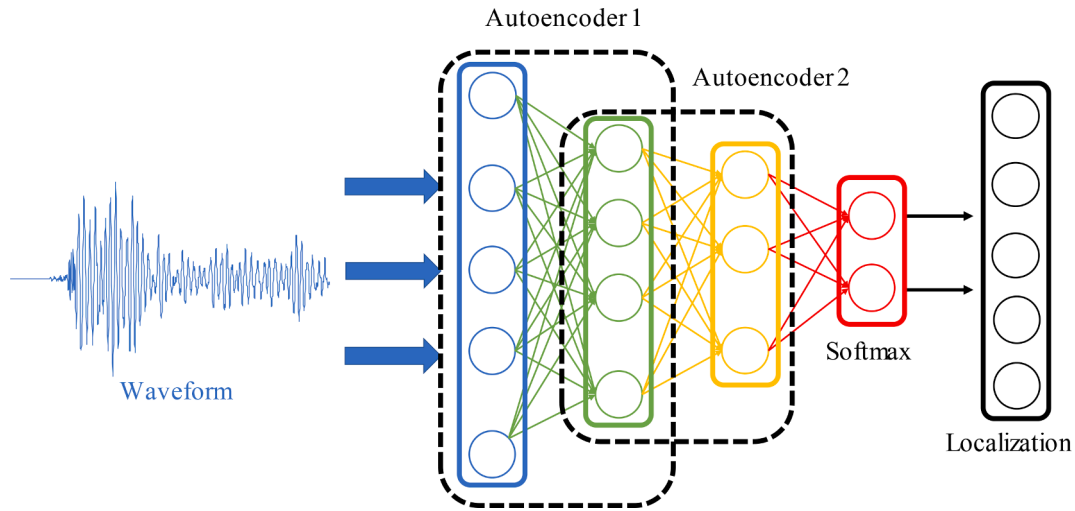


Fig. 10. Stacked Autoencoder network with two autoencoders.

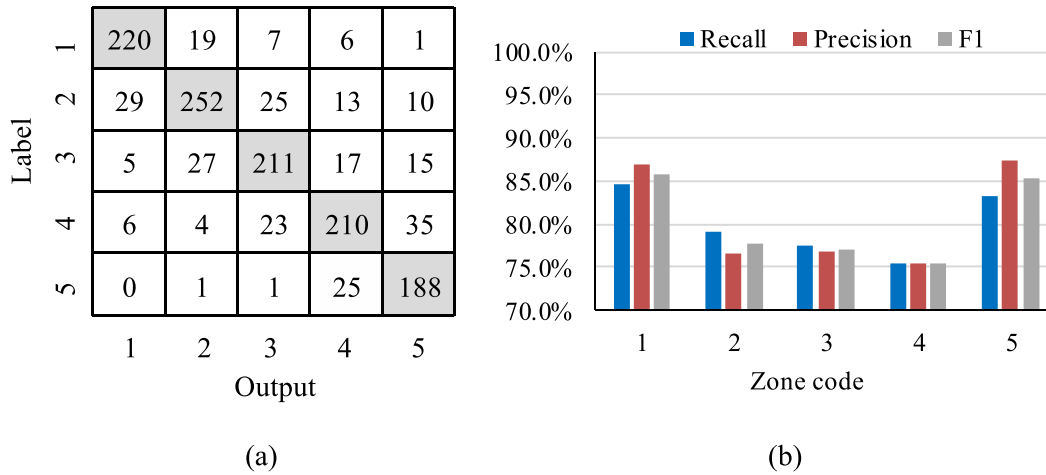


Fig. 11. Performance of each classification using ANN: (a) confusion matrix; (b) evaluation of each zone.

of AE data that are correctly localized in their corresponding zones are shown in the main diagonal of the confusion matrix. There were 1080 AE events correctly located in the corresponding zone, accounting for 80% of the total AE data. In other words, the overall accuracy was 80%. In addition to accuracy, precision, and recall for each class are usually implemented to evaluate the performance of classification in each class (Guo et al., 2020). The value of precision is obtained by Eq. (10):

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Where TP refers to true positives, which means the number of samples that correctly classify to the corresponding class, FP refers to false positives, which is the number of samples that do not belong to the class but are classified into the class by error. Precisions of the five class are respectively 86.9%, 76.6%, 76.7%, 75.5%, 87.4% from zone 1 to 5.

The value of recall can be obtained by Eq. (11):

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

Where FN refers to false negatives, the number of samples that belong to the class but are classified into the other classes by error. Recalls of the five classes are respectively 84.6%, 79.0%, 77.5%, 75.5%, 83.2% from zone 1 to 5.

Precision and recall influence each other. A class with high precision

usually has a low recall and vice versa (Buckland and Gey, 1994). To comprehensively evaluate the efficiency of the classifier in each class, the F1-score can be employed.

F1-score, also referred to as the balanced F score, is defined as the harmonic mean of precision and recall (Zhong et al., 2019). It can be provided by Eq. (12):

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \quad (12)$$

F1 of the five classes are respectively 85.7%, 77.8%, 77.1%, 75.5%, 85.2% from zone 1 to 5. The values of precision recall and F1 for each class are compared and shown in Fig. 11b. Zones on the edge (zone 1 and 5) performed better than the inner areas (zones 2, 3, and 4). This situation occurs because when localizing the events in the inner zone, the misclassification events were almost localized in the nearby zones. While the zones near the edge have one nearby zone, misclassification occurred primarily in that zone only.

The computing time for the training process on an intel i-7 four core CPU was 7.96 s. The computing time required for the testing data running on a trained ANN was 0.02 s.

5.2. Source localization using random forest

The data used for the training of this network were 2,700 feature-

based AE signals randomly selected from all 4,050 data. The remaining 1,350 signals were utilized for testing. Fig. 12a shows the confusion matrix of the random forest source localization result. Overall, the accuracy on the test set was 90.5%. 1,222 AE signals were correctly localized, while 128 were in error. The number of AE signals correctly localized in their corresponding zone is shown in the main diagonal of the confusion matrix (Fig. 12a). Fig. 12b shows the precision and recall for each zone. From zone 1 to 5. Precision of the five classes is respectively 94.8%, 90.1%, 92.6%, 86.6%, 88.8%. Recall of the five classes are respectively 94.8%, 92.0%, 91.2%, 85.7%, 88.8%. F1 of the five classes are respectively 94.8%, 91.0%, 91.9%, 86.1%, 88.8%. Zone 1 performed the best rather than the other zones (zones 2, 3, 4, and 5). This might be caused by the attenuation of signals during the transmission. Zone 1 has the smallest distance from the AE sensor, which results in the signal from zone 1, retaining the complete information.

The computing time for the training process within an intel i-7 four core CPU was 4.03 s. The computing time for testing was 0.02 s.

The importance percentage of AE features can be provided by the random forest model. The names of all 15 features and their corresponding percentages of importance are shown in Fig. 13a. Their ranking descends from left to right. By observing Fig. 13a, it can be noticed that the importance of the features “peak frequency” and “rise time” are significantly higher than the rest. Those two features have a major impact on the source localization results. Features “average frequency”, “ASL” and reverberation frequency hold the lowest importance, meaning their impact on the results is limited. Deleting them will not have a significant influence on localization performance.

The appropriate features can be selected from the feature set by the guidance of the ranking of feature importance when the input has a large dimension. Deleting the features with low importance can reduce the computing time and may increase the classification accuracy. In this paper, features were selected based on the cumulative importance of features. The cumulative importance can be obtained by the calculation of one-by-one accumulation by the sequencing of importance, which can be observed in Fig. 13a. The features that have 65%, 75%, 85%, and 95% cumulative importance were respectively selected. The cumulative importance of features was calculated and shown in Fig. 13b. Four subsets were extracted from the feature-based dataset based on the cumulative importance of features: Subset 1 consist of “peak frequency”, “rise time”, “initial frequency”, “amplitude” and “duration”. The sum of the important percentage of the features in this subset is 65%. Subset 2 which in consist of “peak frequency”, “rise time”, “initial frequency”, “amplitude”, “duration”, “PCNTS” and “counts”. The sum of the importance percentage is 75%. Subset 3 which in consist of “peak frequency”, “rise time”, “initial frequency”, “amplitude”, “duration”, “PCNTS”, “counts”, “absolute energy” and “frequency centroid”. The

sum of the importance percentage is 85%. Subset 4 which in consist of “peak frequency”, “rise time”, “initial frequency”, “amplitude”, “duration”, “PCNTS”, “counts”, “absolute energy”, “frequency centroid”, “energy” and “RMS”. The sum of the important percentage of the features in this subset is 95%. The accuracies and computing times of the random forest models attained by inputting the four subsets are plotted in Fig. 14. The accuracy increased as low importance features were deleted. The total time required for training and testing reduces with the deletion of these features. The maximum accuracy (91.5%) was observed when the subset with 75% of the cumulative importance was utilized as the input; meanwhile, the total computing time was 2.98 s, in which the training required 2.97 s and the testing required 0.01 s. Comparing with the case using the original feature-based dataset, the accuracy increased from 90.5% to 91.5%, training time was reduced from 4.03 s to 2.97 s, and testing time was reduced from 0.02 s to 0.01 s. This indicates the accuracy and computing time can be optimized by using the subset with 75% of cumulative importance. Therefore, the random forest model which utilized the subset consisting of “peak frequency”, “rise time”, “initial frequency”, “amplitude”, “duration”, “PCNTS” and “counts” is preferred and compared with ANN and stacked autoencoders in Section 5.4.

5.3. Source localization using SAE

The waveform-based AE dataset was utilized for training and testing the performance of the source localization approach using the stacked autoencoder. In this paper, the data acquisition system was set to record 2 million AE samples per second. The recording duration of an AE event was 1 microsecond. Therefore, every wave-based signal in the dataset contains 2000 sample points. The stacked autoencoder in this paper is composed of two autoencoders. The first autoencoder will extract features from the 2000 input sample points, compress them into the first feature set of 100 sample points, and then reconstruct the 2000 input points. The second autoencoder takes the first feature set as an input and compresses it into the second feature set of 50 sample points, and then the autoencoder reconstructs the 100 input points. By comparing the data before and after reconstruction, the feature extraction effect of the autoencoder can be evaluated. Fig. 15 show the input data and their reconstructions of randomly selected data from the waveform-based AE dataset. Fig. 15a shows the input data with 2000 sample points entering the first autoencoder and the reconstructed data. Fig. 15b shows input data with 100 sample points and data reconstructed by the second autoencoder. It can be observed from the figure that the reconstructed data is very similar to their input. They have similar data distribution patterns. This indicates that the two autoencoders can successfully reconstruct their input. Fig. 16 shows the original input data, and the

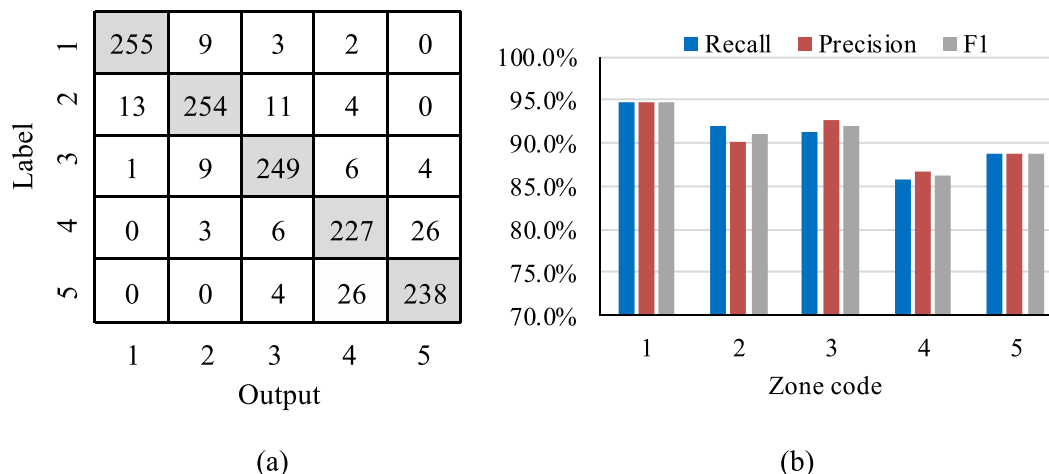


Fig. 12. Performance of each classification using random forest: (a) confusion matrix; (b) evaluation of each zone.

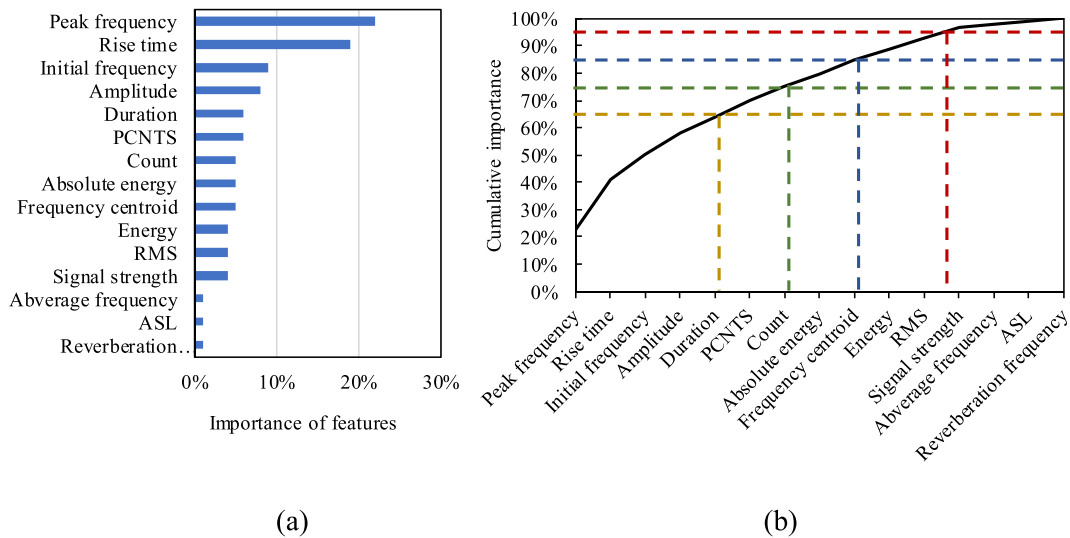


Fig. 13. Importance of features: (a) the ranking; (b) cumulative importance.

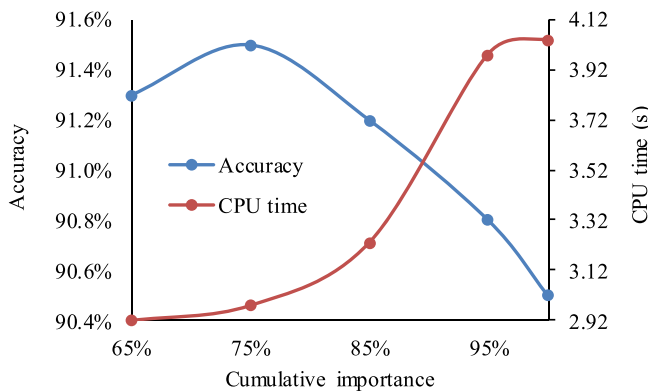


Fig. 14. Accuracies and computing times of random forest using inputs with different cumulative importance.

feature sets extracted by the first and the second autoencoders. In the end, the input waveform-based AE data with 2000 sample points was compressed to the feature set with 50 sample points, and most of the effective information was saved in the feature set. The condensed feature set was employed by a softmax layer for classification.

The data used for the training of this network contained 2700 waveform-based AE signals randomly selected from all 4050 data. The

remaining 1350 signals were utilized for testing. Fig. 17a shows the source localization results given by the stacked autoencoder. The network can correctly localize 1320 AE signals to their corresponding zone. The overall accuracy is 97.8%. To be more specific, the number of the AE data that were correctly localized in the corresponding zones is respectively 266 in zone 1, 264 in zone 2, 264 in zone 3, 258 in zone 4, 264 in zone 5. The precision, recall, and F1 for each zone are shown in Fig. 17b. From zone 1 to 5. Precisions of the five class are respectively 98.5%, 98.5%, 96.7%, 97.3%, 96.4%. Recalls of the five class are respectively 98.5%, 97.8%, 97.8%, 95.6%, 97.8%. F1 of the five class are respectively 98.5%, 98.1%, 97.2%, 96.4%, 97.1%.

The computing time for the training process on an intel i-7 four core CPU was 352.89 s. The computing time for testing was 0.04 s.

5.4. Discussion

The accuracies and computing times of the localization approaches using three different machine learning algorithms are provided in Table 4. The approach using stacked autoencoders gives the highest accuracy (97.8%), which is much higher than the accuracy obtained by ANN (80.0%), and the accuracy given by random forest is acceptable (91.5%). This is because, deep learning algorithm such as stacked autoencoders automatically extracts the features that are most sensitive to the labels, however, the features input to ANN, and random forest are extracted manually from the AE signals and there may be several useful

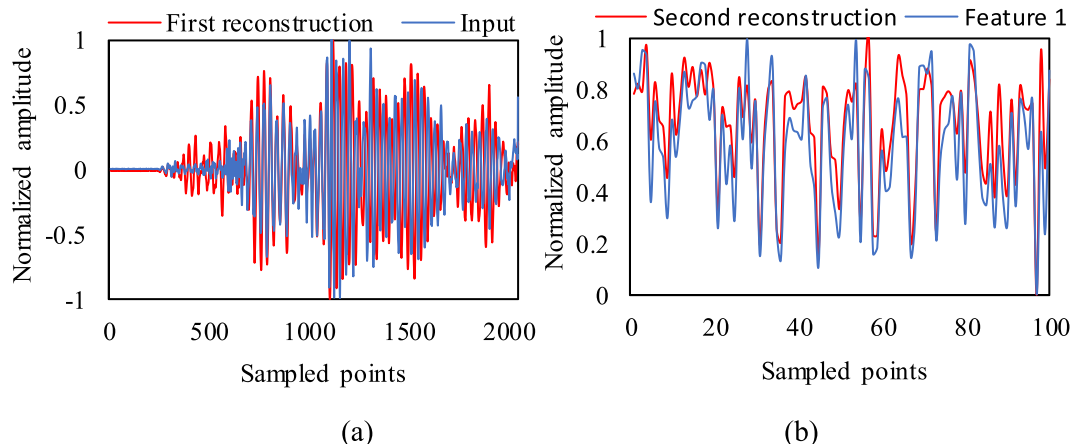


Fig. 15. The input and reconstruction patterns of autoencoders: (a) the first autoencoder; (b) the second autoencoder.

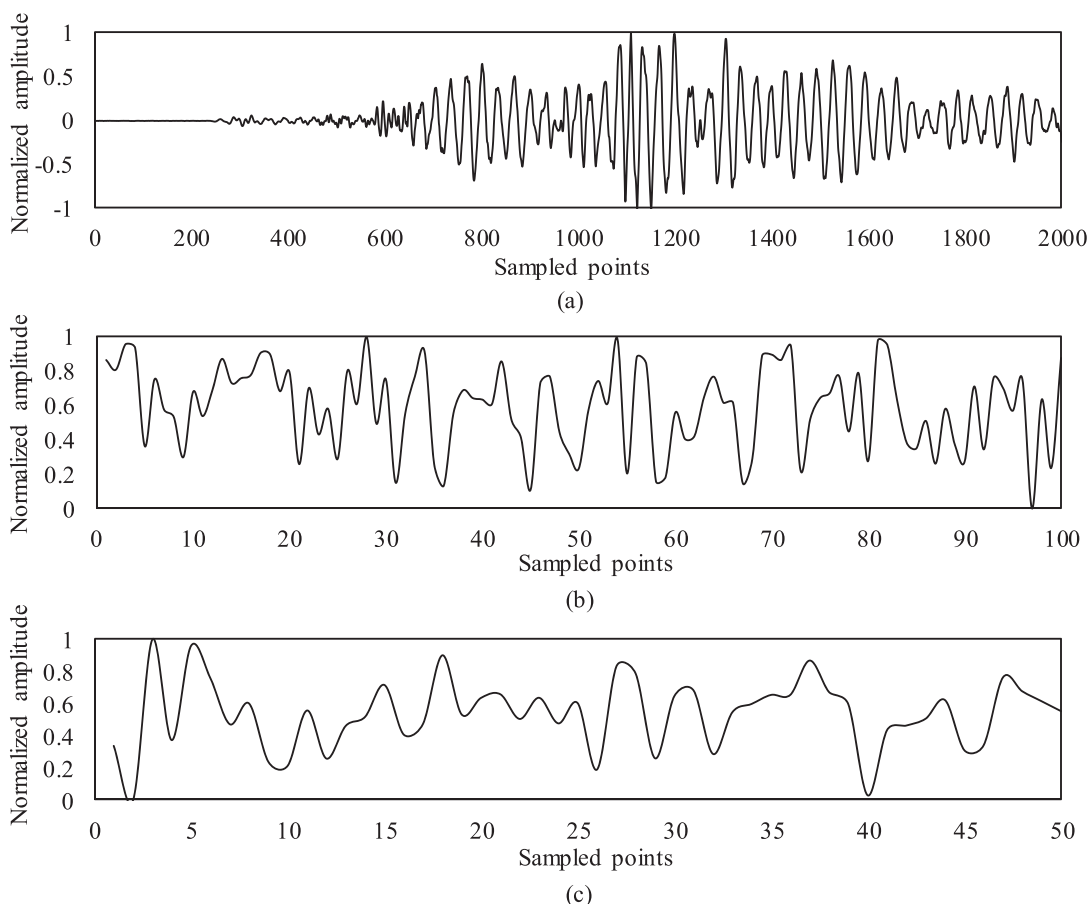


Fig. 16. Patterns of the input waveform and extracted features: (a) input waveform; (b) the first compressed feature set; (c) the second compressed feature set.

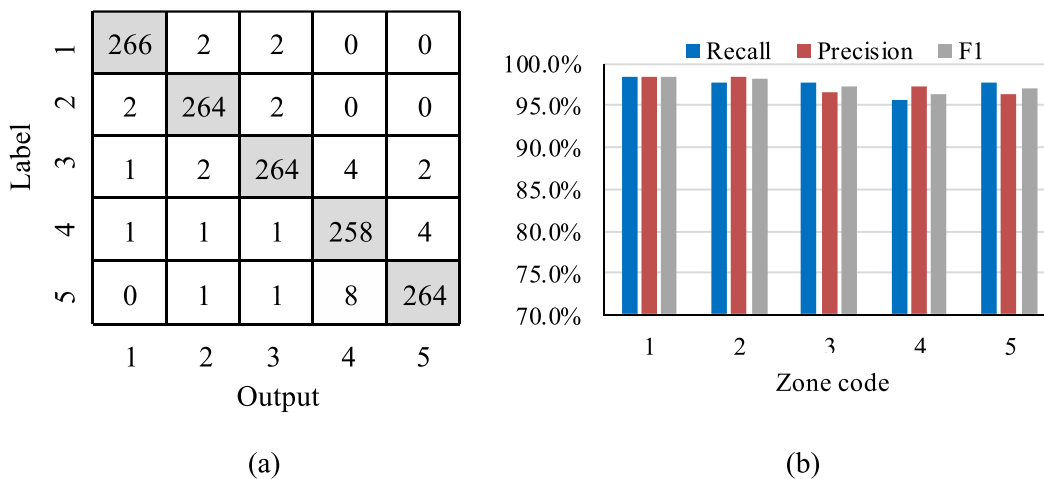


Fig. 17. Performance of classification using SAE: (a) confusion matrix; (b) evaluation of each zone.

Table 4
Accuracies and computing times of three algorithms.

Machine learning algorithm	Accuracy	Training time (s)	Testing time (s)
ANN	80.0%	7.96	0.02
Random forest (75% cumulative importance)	91.5%	2.97	0.01
Stacked autoencoders	97.8%	352.89	0.04

features that are not extracted as input. The training time of stacked autoencoders (352.89 s) is significantly more than ANN (7.96 s) and random forest (2.97 s), while the time for the trained machine learning models to run the testing processes don't show a wide variety. Testing time for stacked autoencoders, ANN, and the random forest is respectively 0.04 s, 0.02 s, and 0.01 s.

The F1-scores obtained by three different algorithms are shown in Fig. 18. It can be observed that the F1 scores of five zones obtained by random forest and stacked autoencoders are relatively stable, while the F1-scores obtained by ANN show a wide variety in different zones.

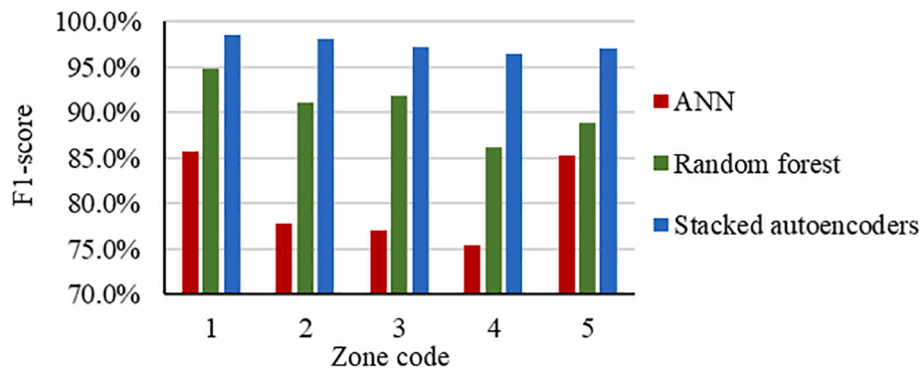


Fig. 18. Comparison of F1-scores.

Moreover, the F1-scores of stacked autoencoders are generally the highest of the three. This means the stacked autoencoder has the best performance for localization in each zone.

To summarize, the source localization approach using stacked autoencoders has the best accuracy but the largest training time, while the testing times do not differ much by using models being trained. If there is no minimum requirement for training time, the source localization approach using stacked autoencoders is the most desirable. The approach using random forest is slightly worse than using stacked autoencoders but much better than using ANN. The approach using random forest is supposed to be an acceptable choice in the case that waveforms are not available, and only AE parametric features can be provided.

6. Conclusions.

This paper considered three machine learning approaches to localize simulated SCC AE sources on a 304 stainless steel specimen, which has a similar length and thickness with the realistic DCSS canister. ANN, random forest, and stacked autoencoders were used. This study aims to detect and localize AE sources with only one sensor attached opposite the source. To collect a sufficient number of AE data for training and testing, AE sources were simulated on the specimen by conducting Hsu-Nielsen pencil lead break tests. The main conclusions are as follows:

- The performance of three machine learning approaches was compared. The stacked autoencoders have the best performance (97.8% accuracy versus 91.5% and 80.0%). Although the training time required for stacked autoencoders is more than the other two (352.89 s versus 0.02 s and 0.01 s), their computing time required for testing is similar.
- Feature selection can be achieved by running random forest. The random forest model indicated that the AE parametric features “peak frequency”, “rise time”, “initial frequency”, “amplitude”, “duration”, “PCNTS” and “counts” made up 75% of the cumulative importance for all 15 features. Using them as the input for random forest leads to increasing accuracy and decreasing computational time.
- The F1-scores indicated that the performance of stacked autoencoders in various zones is the best among the three. Moreover, good performance in terms of stability in various zones was observed in stacked autoencoders while ANN resulted in more variability.

CRedit authorship contribution statement

Li Ai: Conceptualization, Investigation, Writing - original draft, Formal analysis. **Vafa Soltangharai:** Conceptualization, Investigation, Writing - original draft, Formal analysis, editing and review. **Mahmoud Bayat:** Conceptualization, Investigation, Formal analysis, Writing - review & editing. **Bruce Greer:** Methodology, Investigation. **Paul Ziehl:** Conceptualization, Methodology, Writing - review & editing,

Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was partially supported by Electric Power Research Institute (EPRI) under the project number 1-108781.

Data availability

The raw/processed data required to reproduce these findings can be made available upon request and with the written permission of the sponsor.

References

- Xie, Y., Zhang, J., 2015. Chloride-induced stress corrosion cracking of used nuclear fuel welded stainless steel canisters: A review. *J. Nucl. Mater.* 466, 85–93.
- Hill, J.W. (2018). Acoustic Emission Detection in 304H Stainless Steel Due To Intergranular Stress Corrosion Cracking.
- Yeom, H., Dabney, T., Pocquette, N., Ross, K., Pfefferkorn, F.E., Sridharan, K., 2020. Cold spray deposition of 304L stainless steel to mitigate chloride-induced stress corrosion cracking in canisters for used nuclear fuel storage. *J. Nucl. Mater.* 152254.
- Wu, X., 2020. On residual stress analysis and microstructural evolution for stainless steel type 304 spent nuclear fuel canisters weld joint: Numerical and experimental studies. *J. Nucl. Mater.* 152131.
- Li, X., 2002. A brief review: acoustic emission method for tool wear monitoring during turning. *Int. J. Mach. Tools Manuf.* 42 (2), 157–165.
- Ono, K., 2011. Application of acoustic emission for structure diagnosis. *Diagnostyka* 3–18.
- Anay, R., Soltangharai, V., Assi, L., DeVol, T., Ziehl, P., 2018. Identification of damage mechanisms in cement paste based on acoustic emission. *Constr. Build. Mater.* 164, 286–296.
- Soltangharai, V., Anay, R., Hayes, N.W., Assi, L., Le Pape, Y., Ma, Z.J., Ziehl, P., 2018. Damage mechanism evaluation of large-scale concrete structures affected by alkali-silica reaction using acoustic emission. *Appl. Sci.* 8 (11), 2148.
- Li, D., Kuang, K.S.C., Koh, C.G., 2018. Rail crack monitoring based on Tsallis synchrosqueezed wavelet entropy of acoustic emission signals: A field study. *Struct. Health Monit.* 17 (6), 1410–1424.
- Ai, L., Greer, B., Hill, J., Soltangharai, V., Ziehl, R.A.P. (2019, May). Finite element modeling of acoustic emission in dry cask storage systems generated by cosine bell sources. In AIP Conference Proceedings (Vol. 2102, No. 1, p. 130001). AIP Publishing LLC.
- Ai, L., Soltangharai, V., Anay, R., van Tooren, M.J., Ziehl, P. (2020, March). Data-Driven Source Localization of Impact on Aircraft Control Surfaces. In 2020 IEEE Aerospace Conference (pp. 1–10). IEEE.
- Soltangharai, V., Anay, R., Ai, L., Giannini, E.R., Zhu, J., Ziehl, P., 2020. Temporal evaluation of ASR cracking in concrete specimens using acoustic emission. *J. Mater. Civ. Eng.* 32 (10), 04020285.
- Li, D., Wang, Y., Yan, W.J., Ren, W.X., 2020. Acoustic emission wave classification for rail crack monitoring based on synchrosqueezed wavelet transform and multi-branch convolutional neural network. *Struct. Health Monit.*, 1475921720922797.

- Anay, R., Lane, A., Jáuregui, D.V., Weldon, B.D., Soltangharaei, V., Ziehl, P., 2020. On-site acoustic-emission monitoring for a prestressed concrete BT-54 AASHTO girder bridge. *J. Perform. Constr. Facil.* 34 (3), 04020034.
- Wang, S., Liu, Y., Zhou, H., Zhang, Y., Wu, Z., Yang, Q., 2019. Experimental study on failure process of arch dam based on acoustic emission technique. *Eng. Fail. Anal.* 97, 128–144.
- Nozawa, T., Ozawa, K., Asakura, Y., Kohyama, A., Tanigawa, H., 2014. Evaluation of damage accumulation behavior and strength anisotropy of NITE SiC/SiC composites by acoustic emission, digital image correlation and electrical resistivity monitoring. *J. Nucl. Mater.* 455 (1–3), 549–553.
- Véronique, P., Eric, S., François, G., Jean, K., François, R., Michel, C., 2015. In situ high temperature oxidation analysis of Zircaloy-4 using acoustic emission coupled with thermogravimetry. *J. Nucl. Mater.* 461, 365–375.
- Baek, S.H., Shim, H.S., Kim, J.G., Hur, D.H., 2018. Visualization and acoustic emission monitoring of nucleate boiling on rough and smooth fuel cladding surfaces at atmospheric pressure. *Nucl. Eng. Des.* 330, 429–436.
- Soltangharaei, V., Hill, J.W., Ai, L., Anay, R., Greer, B., Bayat, M., Ziehl, P., 2020. Acoustic emission technique to identify stress corrosion cracking damage. *Struct. Eng. Mech.* 75 (6), 723–736.
- Investigation of Acoustic Emission Technologies for Monitoring Inaccessible Regions of Dry Fuel Storage Systems. EPRI, Palo Alto, CA: 2016.
- Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Methods* 43 (1), 3–31.
- Soltangharaei, V., Anay, R., Begrajka, D., Bijman, M., ElBatanouny, M.K., Ziehl, P., Van Tooren, M.J. 2019. A minimally invasive impact event detection system for aircraft movables. In *AIAA Scitech 2019 Forum*, p. 1268.
- Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* 5 (2), 157–166.
- Breiman, L., 2001. Random forests. *Machine Learn.* 45 (1), 5–32.
- Cerrada, M., Zurita, G., Cabrera, D., Sánchez, R.V., Artés, M., Li, C., 2016. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mech. Syst. Sig. Process.* 70, 87–103.
- Patel, R.K., Giri, V.K., 2016. Feature selection and classification of mechanical fault of an induction motor using random forest classifier. *Perspect. Sci.* 8, 334–337.
- Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H. 2007. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, pp. 153–160.
- Arel, I., Rose, D.C., Karnowski, T.P., 2010. Deep machine learning—a new frontier in artificial intelligence research [research frontier]. *IEEE Comput. Intell. Mag.* 5 (4), 13–18.
- Karvelis, P., Georgoulas, G., Kappatos, V., Stylios, C., 2020. Deep machine learning for structural health monitoring on ship hulls using acoustic emission method. *Ships Offshore Struct.* 1–9.
- Industry Spent Fuel Storage Handbook. EPRI, Palo Alto, CA: 2010. 102048.
- Wadley, H.N.G., Mehrabian, R., 1984. Acoustic emission for materials processing: a review. *Mater. Sci. Eng.* 65 (2), 245–263.
- Grosse, C.U., Ohtsu, M. eds., 2008. *Acoustic Emission Testing*. Springer Science & Business Media.
- Scruby, C.B., 1987. An introduction to acoustic emission. *J. Phys. E: Sci. Instrum.* 20 (8), 946.
- Hsu, N.N., 1977. Inventor; Lockheed Corp, assignee. Acoustic emissions simulator. United States patent US 4,018,084.
- Laksimi, A., Benmedakhene, S., Bounouas, L. (1999, July). Monitoring acoustic emission during tensile loading of thermoplastic composites materials. In *Proceeding of ICCM (Vol. 12)*.
- Ali, S.M., Hui, K.H., Hee, L.M., Leong, M.S., Abdelrhman, A.M., Al-Obaidi, M.A., 2019. Observations of changes in acoustic emission parameters for varying corrosion defect in reciprocating compressor valves. *Ain Shams Eng. J.* 10 (2), 253–265.
- Jayalakshmi, T., Santhakumaran, A., 2011. Statistical normalization and back propagation for classification. *Int. J. Comput. Theory Eng.* 3 (1), 1793–8201.
- Hassoun, M.H., 1995. *Fundamentals of Artificial Neural Networks*. MIT press.
- Sun, R., Chen, Y., Dubey, A., Pugliese, P., 2020. Hybrid electric buses fuel consumption prediction based on real-world driving data. *Transportation Research Part D: Transport and Environment*, p.102637.
- Rogers, L.L., Dowla, F.U., 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resour. Res.* 30 (2), 457–481.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2 (3), 18–22.
- Quinlan, J.R., 2014. *C4.5: programs for machine learning*. Elsevier.
- Sathyadevan, S., Nair, R.R. 2015. Comparative analysis of decision tree algorithms: ID3, C4.5 and random forest. In *Computational intelligence in data mining-volume 1* (pp. 549–562). Springer, New Delhi.
- Rutkowski, L., Jaworski, M., Pietruczuk, L., Duda, P., 2014. The CART decision tree for mining data streams. *Inf. Sci.* 266, 1–15.
- Genuer, R., Poggi, J.M., Tuleau-Malot, C. 2010. Variable selection using random forests. *Pattern Recognition Letters*, 31, 2225 doi: 10.1016.J. PATREC14.
- Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., Hamprecht, F.A., 2009. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinf.* 10 (1), 213.
- Ng A. Sparse autoencoder. CS294A Lecture notes. 2011 Jan;72(2011):1–9.
- Tao, S., Zhang, T., Yang, J., Wang, X., Lu, W. (2015, July). Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In *2015 34th Chinese Control Conference (CCC)* (pp. 6331–6335). IEEE.
- Guo, F., Qian, Y., Wu, Y., Leng, Z., Yu, H., 2020. Automatic railroad track components inspection using real-time instance segmentation. *Computer-Aided Civil and Infrastructure Engineering*. <https://doi.org/10.1111/mice.12625>.
- Buckland, M., Gey, F., 1994. The relationship between recall and precision. *J. Am. Soc. Inf. Sci.* 45 (1), 12–19.
- Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* 221, 430–443.