HASAN, M.J., ISLAM, M.M.M. and KIM, J.-M. 2021. Multi-sensor fusion-based time-frequency imaging and transfer learning for spherical tank crack diagnosis under variable pressure conditions. *Measurement* [online], 168, article 108478. Available from: https://doi.org/10.1016/j.measurement.2020.108478

Multi-sensor fusion-based time-frequency imaging and transfer learning for spherical tank crack diagnosis under variable pressure conditions.

HASAN, M.J., ISLAM, M.M.M. and KIM, J.-M.

2021

Multi-Sensor Fusion-based Time-Frequency Imaging and Transfer Learning for Spherical Tank Crack Diagnosis Under Variable Pressure Conditions

Md Junayed Hasan[1*], M M Manjurul Islam[2*], and Jong-Myon Kim[3*]

Affiliation:

*School of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan (44610), South Korea

Address:

*Building No. 7, Room 308, 93 Daehak-ro, Nam-gu, Ulsan (44610), South Korea

Email:

[1]junhasan@gmail.com
[2]m.m.manjurul@gmail.com
[3]jmkim07@ulsan.ac.kr

Corresponding Author:

Jong-Myon Kim,
School of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan (44610), South Korea.
Email: jmkim07@ulsan.ac.kr
Tel: +82-52-259-2217
Fax: +82-52-259-1687

Abstract

Due to their structurally efficient shape, spherical tanks are widely used in various industries to store massive amounts of compressed gas or fluid at high pressure. Cracks in such tanks can cause significant financial losses and human casualties. To diagnose the state of spherical tanks at an early stage, existing diagnostic frameworks include manual feature analysis from time, frequency, or time-frequency domains. However, these types of analyses require extensive domain expertise and the statistical feature extraction models are very sensitive to variable operating conditions. To address these issues, several deep learning-based approaches where the feature analyses can be performed automatically have been introduced. Nevertheless, construction of these algorithms requires a substantial amount of prior knowledge and time to establish an optimal diagnostic model. To solve these problems, a crack diagnosis framework is proposed that combines a new signal-to-imaging technique and transfer learning-aided deep learning framework to automate the diagnostic process. The objective of the signal-to-imaging technique is to convert one-dimensional (1D) acoustic emission (AE) signals from multiple sensors into a two-dimensional (2D) image to capture information under variable operating conditions. In this process, a short-time Fourier transform (STFT) is first applied to the AE signal of each sensor, and the STFT results from the different sensors are then fused to obtain a condition-invariant 2D image of cracks; this scheme is denoted as Multi-Sensors Fusion-based Time-Frequency Imaging (MSFTFI). The MSFTFI images are subsequently fed to the fine-tuned transfer learning (FTL) model built on a convolutional neural network (CNN) framework for diagnosing crack types. The proposed diagnostic scheme (MSFTFI+FTL) is tested with a standard AE dataset collected from a self-designed spherical tank to validate the performance under variable pressure conditions. The results suggest that the proposed strategy significantly outperformed existing methods with average performance improvements of 5.39 - 10.82%.

Keywords: Acoustic emissions, convolutional neural network, fault diagnosis, multi-sensors, transfer learning, spherical tank

1. Introduction

Spherical tanks are commonly used in refineries, nuclear plants, and chemical industries to store massive amounts of gas or fluid at high pressure due to their structurally efficient shape [1]. A small accident in a spherical tank may result in the loss of millions of dollars and interruptions in operations, while a large accident can lead to the devaluation of a company's stock price and bankruptcy [1,2]. The prevention of such hazardous incidents is of utmost concern since spherical tanks are mostly installed in environments with coarse operating conditions. Failures in these environments can occur due to stress, corrosion, fatigue cracks, leakage, lightning, and open flames [3–5]. Fatigue cracks are one of the leading causes of spherical tank failure, and such an incident can lead to the spillage of enclosed substances [6–8]. Therefore, in this paper, a data-driven framework is proposed that combines a new signal-to-imaging technique for feature representation and deep learning to automate feature extraction and the crack classification process.

Acoustic emission (AE) is a popular non-destructive test (NDT) for monitoring the structural health of spherical tanks since it can capture a very low-energy signal with information regarding a crack [9,10]. Existing crack diagnosis methods consist of three main steps: (1) handicraft feature extraction by analyzing acquired signals with information regarding structural health, (2) feature subset selection to determine the best set of features, and (3) detection and diagnosis of health states using shallow machine learning algorithms [9,10]. These approaches are mostly built upon the manual feature extraction process and traditional machine learning approaches that require an expert level of domain knowledge [9,10]. In [11], Morofuji *et al.* developed an AE-based technique to identify corrosion in a tank by analyzing the fundamental characteristics of AE waves. In [2], Sohaib *et al.* analyzed the statistical properties of time- and frequency-domain AE signals to extract crack features with information on storage tank health conditions; the extracted features were further applied to a support vector machine (SVM) to diagnose crack types. In previous studies, the authors improved the formation of health indicators for cracks based on a statistical analysis of AE signals in the time and frequency domains [12,13]. However, these approaches have two main limitations: (1) handicraft features are sensitive to variable operating conditions (e.g., different pressures, temperatures, etc.) and (2) the use of shallow machine learning techniques is restricted to the automation of a general diagnostic framework such as in the case of spherical tanks operating in an industrial setting.

To address the first issue mentioned above, a new signal-to-imaging technique is developed that transforms a one-dimensional (1D) AE time-domain signal into a two-dimensional (2D) image based on a short-time Fourier transform (STFT). STFT is an effective time-frequency analysis method that can capture sensitive information changes in the AE signal of spherical tank health due to variable operating conditions [14]. To summarize the proposed signal-to-imaging technique, collected multi-sensor AE signals are first decomposed based on STFT to obtain time vs. frequency information. Next, the time vs. frequency information for each sensor is converted into a gray-scale image. Finally, all gray-scale images for the multi-sensors are fused to form a 2D image about the invariant crack health condition; this process is denoted as Multi-Sensor Fusion-based Time-Frequency Imaging (MSFTFI). MSFTFI produces a 2D image that contains information from different sensors and enables generalization to identify crack types under different pressure conditions.

To address the second issue mentioned above, developed MSFTFI images are fed to a deep learning framework for diagnosing crack types under variable pressure conditions. A popular deep learning framework such as a convolution neural network (CNN) automates the feature extractor and/or diagnostic framework to accommodate the diverse nature of input data [15,16]. However, the CNN often requires that a new optimal network model be established for the classification task under variable operating conditions. Therefore, the CNN requires substantial time to establish an optimal architecture because deciding on the proper structures and training parameters is a complex process that is mostly dependent on prior human experience [17]. Fortunately, transfer learning (TL)-based approaches can inherit improvements from other pre-trained models using developed structures or learning parameters, which can mitigate the need for prior knowledge and save a significant amount of time [18–20]. In this paper, fine-tuned transfer learning (FTL) built on the CNN architecture is applied to automate the final diagnostic process. To build the FTL model, the MSFTFI images from one pressure condition of a collected AE dataset are first used for training; this is defined as a source task. Once the training is complete and the performance of the task is satisfactory, the acquired knowledge is passed to the target task. In the target task, the MSFTFI images for different pressure conditions are fed to the learned CNN in the FTL framework. The proposed diagnostic framework (MSFTFI +FTL) was tested on an AE dataset collected from multiple sensors of a spherical tank to validate performance. The contributions of the proposed scheme can be briefly summarized as follows:

(1) A new signal-to-imaging technique that applies a short-time Fourier transform on 1D AE signals and combines the results of the STFT from multiple sensors to obtain a condition-invariant 2D crack image (i.e., MSFTFI).

(2) The 2D MSFTFI images are further applied with transfer learning built on a convolutional neural network architecture to automate the feature extraction and classification processes.

(3) A fine-tuned transfer learning (FTL) model to enhance the classification performance under variable pressure conditions. FTL transfers learned parameters among the CNN models to obtain a fine-tuned model through a knowledge sharing process. The FTL diminishes the need for adjusting CNN architectures with different parameters for various working conditions.

The remainder of this paper is organized as follows: Section 2 provides background knowledge of the STFT, CNN, and FTL that is relevant to the proposed diagnostic framework. The overall scheme of the proposed approach and the self-designed experimental test setup are described in Section 3. Experimental results are discussed in Section 4, and conclusions are presented in Section 5.

2. Preliminaries

This section highlights technical information regarding the short-time Fourier transform, convolutional neural network, and the fundamentals of transfer learning.

2.1 Short–Time Fourier Transform

Time-domain or frequency-domain analysis is frequently utilized to observe the health state of different industrial equipment. However, neither of these methods can portray signal variations in the association between time and frequency domains. In practical cases, most signals acquired from equipment (e.g., spherical tanks) are non-stationary in nature. Therefore, time-frequency analysis meets the challenges of evaluating such signals in the shape of an image [17]. Furthermore, the features of an image directly impact the final detection accuracy of deep learning-based algorithms. Hence, it is of immense significance to investigate methods for time-frequency based analysis.

STFT is a time-frequency-based decomposition technique that is effective for analyzing non-stationary time-varying signals. Such an analysis allows 1D heath condition signals to be transformed into 2D matrices. Thus, STFT contributes to the processing of deep learning-based algorithms by providing 2D matrices [14]. The key concept of STFT is to utilize a static-length window function to capture the total time-varying signals in a smaller time, t, and process each of the captured parts with a Fourier transform to obtain a local spectrum. This way, the feature spectrum of the STFT contains information from the time and frequency domains. The basic formula of the 2D STFT function can be expressed as follows:

$$STFT\{x(t)\}(t,w)=\int_{-\infty}^{\infty} x(fr)F(fr-t)e^{-jwt}ds \tag{1}$$

where $x(t)$ is the signal from the time domain, t and fr indicate time, w is the frequency, and $F(fr-t)$ is the window function. Here, $F(fr-t)$ regulates the time and frequency resolution of the calculated spectrum. A longer window length describes the spectrum with a higher frequency and lower time resolution after calculating the Fourier transform. Thus, the window size must be carefully chosen for a better analysis [14].

2.2 Convolutional Neural Network Architecture

CNN is a deep neural network primarily constructed with an input layer, several convolution layers, pooling layers, a few fully connected layers, and one final classification layer [19,21]. One of benefits of CNN is the sparse number of attributes, which decreases the number of learning parameters (i.e., weights and biases) when compared to conventional artificial neural networks (ANN) [21]. Several optimization constraints, including dropout, batch normalization (BN), and rectified linear units (ReLUs), are also utilized for incorporation into the main architecture of the basic CNN to improve classification performance [22–24].

2.2.1 Convolution Layers

Several convolution strategies have been presented in the literature. All types of convolution operations were primarily utilized for feature mapping to extract the attributes of an input image to the network through their shared weight properties. A valid convolution, which is a convolution operation without any kind of padding on the provided input to the network, is frequently preferred in CNN architectures [17].

Padding is known as the preprocessing step before the convolution operation. For example, if the network has an input A of an $m \times m$ image and there is convolution filter F with a size of $f \times f$, the output matrix of the valid convolution can be calculated as:

$$\left(A * F\right) = \left[\frac{n-f}{s}\right] + 1 \times \left[\frac{n-f}{s}\right] + 1 \tag{2}$$

where s is the number of vertical and horizontal steps that the filter F takes over the provided input image A, and the $(*)$ operator represents the convolution operation. Finally, the Rectified Linear Unit (ReLU) activation function is adopted to finalize the output of the CNN. The overall process of this convolution operation can be expressed as:

$$Y_{cn} = f\left(X * W_{cn} + B_{cn}\right) \tag{3}$$

$$f\left(x\right) = \begin{cases} x, \text{when}\left(x \geq 0\right) \\ 0, \text{when}\left(x \leq 0\right) \end{cases} \tag{4}$$

Here, $X \in R^{A \times B}$ refers to the input of the convolution layer, where the dimension of the input image is $A \times B$; W_{cn} and B_{cn} are the weight matrix and bias, respectively, and f is the ReLU activation function. Thus, the cn th feature map output Y_{cn} is obtained.

2.2.2 Pooling Layers

A down-sampling layer, known as a pooling layer, generally supports each convolutional layer. The objective of the pooling architecture is to decrease both the number of spatial factors and the computational load. Therefore, it is useful to lessen the over-fitting probability. In this study, max pooling [25], which can yield a maximum value of the convolutional output Y_{cn}, is adopted as follows:

$$PL_{cn} = max\left(Y_{cn}\right) \tag{5}$$

The CNN usually incorporates numerous sequences of convolution and pooling layers. Consequently, many fully connected layers proceed layer by layer, which transforms the matrix in a filter to a column or row.

Lastly, a SoftMax [26] function is applied to approximate the probability of every target in the final output layer.

2.2.3 Objective Function

The main objective of the CNN is to reduce the training error. That is to say, the difference between the actual output Y_k and predicted output Y_{Ak} must be minimized by the network. To minimize error, the following cost function is adopted:

$$E(p) = \frac{1}{2}\sum_{k=1}^{m}\left(Y_k^p - Y_{Ak}^p\right)^2 \tag{6}$$

where m is the quantity of neurons, and p denotes the p -th iterative steps. The objective of the CNN is to lessen the cost function $E(n)$ via backpropagation and stochastic gradient descent (GD) [27].

2.3 Transfer Learning with CNN

The main goal of transfer learning is to transfer the knowledge acquired from a specific source task to another new (different but relatively similar) target task in order to improve the performance of the target task [18,20]. Fine-tuning-based transfer learning (FTL) is one of the key factors in this process [17]. In the FTL approach, the parameters, weights, and network structure of one specific task are transferred to the related task. For example, in the source task, the final output of the CNN after completion of the training with source data can be attained as follows:

$$CNN_s = f_s\left(I_s, E_s\right) \tag{7}$$

where I_s refers to the input source data, E_s is the objective function, and f_s is the mapping function of the source task. Similarly, the output of the target domain can be obtained as follows:

$$CNN_T = f_T\left(I_T, E_T\right) \tag{8}$$

where I_T refers to the input source data, E_T is the objective function, and f_T is the mapping function of the source task. In this approach, the network learns the correlated properties in the source task by first attaining the mapping function f_s. It then transfers f_s to the target domain and obtains f_T. Hence, the main objective of FTL is to improve the learning process of the target domain by utilizing the acquired knowledge of the source domain.

3. Proposed Methodology

The main purpose of this study is to diagnose crack types in spherical tanks under variable pressure conditions based on a new signal-to-imaging technique and a fine-tuned transfer learning algorithm. Figure 1 presents a block diagram of the proposed methodology.

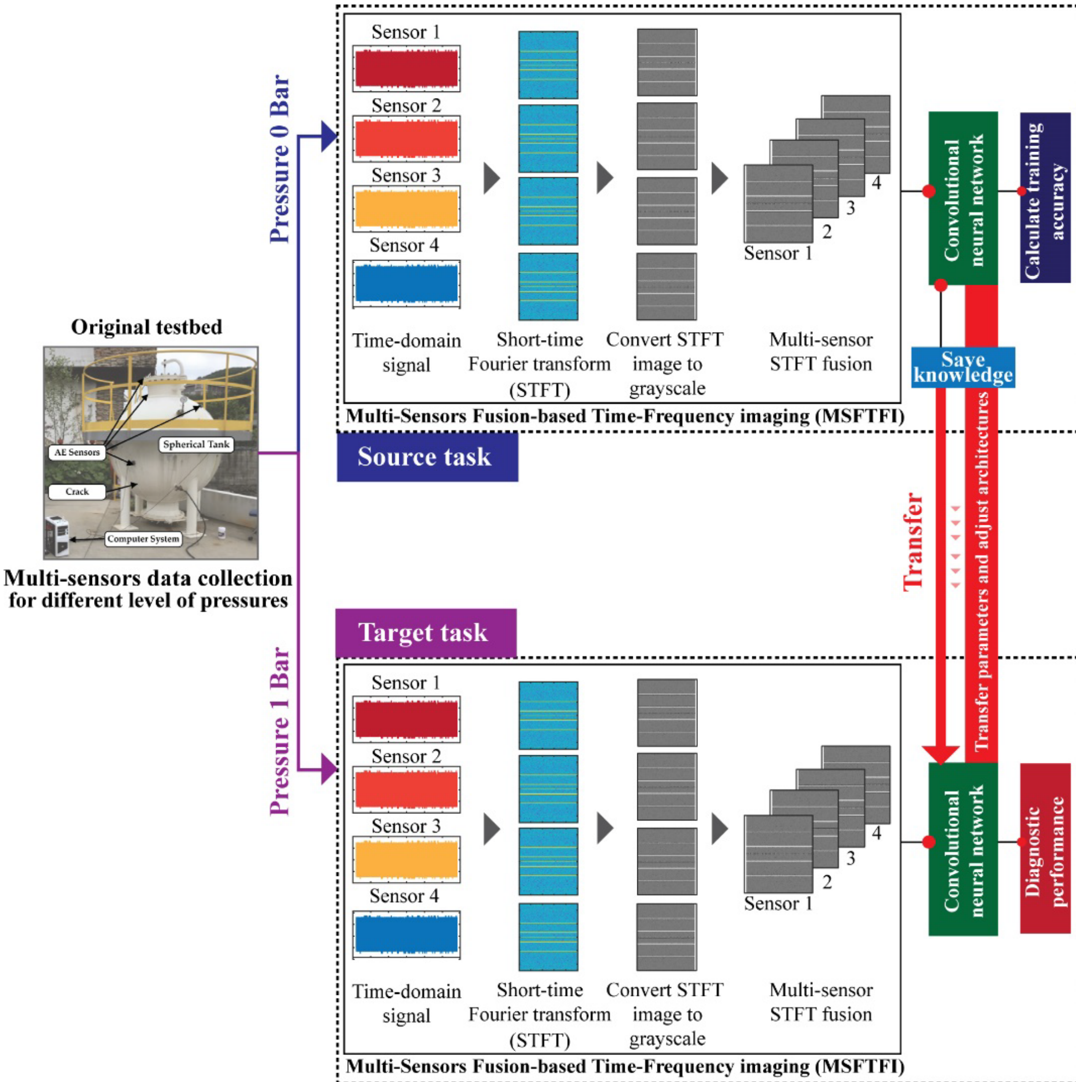

Figure 1: The proposed diagnostic framework for a spherical tank based on transfer learning.

As depicted visually in the figure, the proposed method is composed of three main steps:

(1) Source task: From the original testbed of the spherical tank, data for a pressure of zero (0) bar are collected. The collected 1D acoustic emission signals from multiple sensors are transformed into a 2D image (MSFTFI). MSFTFI allows for visualization of the health conditions of the spherical tank, which are later passed on to the CNN in the FTL framework for optimization of network parameters.

(2) Transfer block: The knowledge assembled from the source task mainly passes to the target network for boosting the performance of the target task by optimizing the parameters of the target network.

(3) Target task: From the testbed, the AE signals for 1 bar of pressure are converted into an MSFTFI image for testing the final model used for classifying different health states.

3.1 Self-Designed Testbed for AE Data

An experiment is performed on a self-designed test bed to collect AE signals from multiple sensors of a spherical tank. The data acquisition (DAQ) system is developed according to industrial standards offered in the American Society of Mechanical Engineers (ASME) Boiler & Pressure Vessel Code (BPVC) – version 2015. A carbon steel (model A283, grade C) spherical tank is utilized to collect AE signals. Here, 4 WDI-AST [28] sensors with -25 dB peak sensitivity are attached to the carbon steel tank. The location of a 3 mm pinhole crack and locations of the AE sensors are presented in Figure 2.

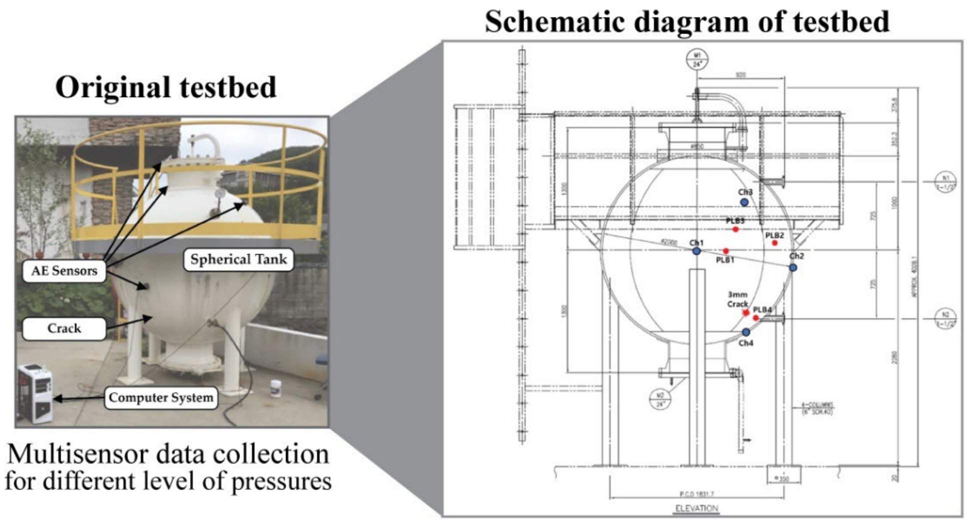

Figure 2: Schematic diagram of the self-designed test rig of a spherical tank.

For collecting data through multiple channels (AE sensors) at two different pressures (0 and 1 bar), a pencil lead test was conducted to produce a guided wave through the structure of the steel tank. A peripheral component interconnect bus (PCI-2) based DAQ device [29] was attached to the AE sensors to record the AE signals for further analysis. Data were collected with a sampling frequency of 1 MHz

3.2 Multi-Sensor Fusion-Based Time-Frequency Imaging (MSFTFI)

Data preprocessing plays an important role in the neural network-based data-driven diagnostic framework [30–32]. Here, MSFTFI is proposed for the preprocessing of AE signals from multiple sensors. In the MSFTFI framework, raw time-domain AE signals from four separate sensors are first decomposed via STFT. Thus, information from both the time and frequency domains are present in the STFT images of each sensor. The resulting matrices of the STFT images are then converted into gray-scale images. These gray-scale images must be compressed to meet the input size constraint of the proposed CNN architecture [17]. Therefore, each gray-scale image from the four separate sensors are compressed into 256×256 dimensions. Finally, the compressed gray-scale images are fused according to sensor to form the final MSFTFI image with a dimension of $256 \times 256 \times 4$. The overall MSFTFI process is displayed in Figure 3.

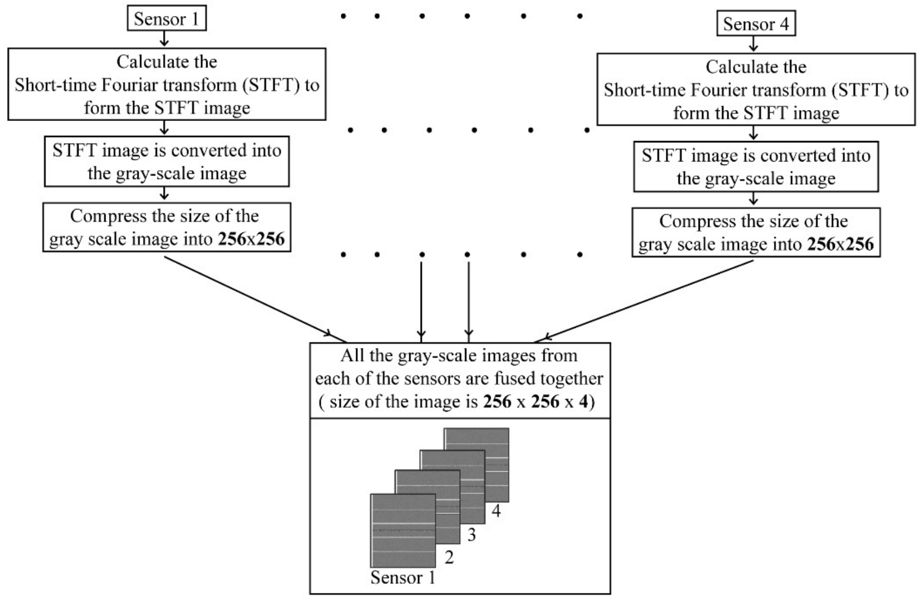

Figure 3: Flowchart of the multi-sensor fusion-based time-frequency imaging process using short-time Fourier transform analysis.

3.3 Convolutional Neural Network Architecture

The prepared MSFTFI images are used as inputs in the proposed CNN architecture. While carrying out the source task (Figure 1), input data are fed to the network to optimize network parameters by minimizing the objective function (Equation 6). The proposed CNN architecture is illustrated in Figure 4.

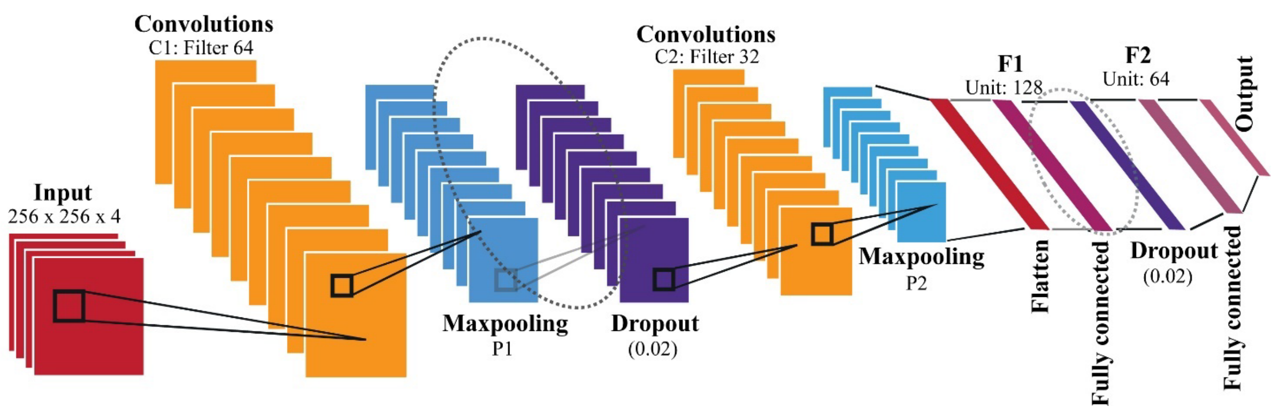

Figure 4: Proposed architecture of the convolutional neural network.

As shown in Figure 4, the proposed architecture has 10 layers: 1 input layer, 2 convolution layers, 2 pooling layers, 2 dropout layers, 2 fully connected layers, and 1 final output layer. The size of the input layer is $256 \times 256 \times 4$ (the size of MSFTFI image), while the size of the convolution kernel is 3×3 to improve model training efficiency by reducing the number of parameters. The C1 and C2 convolution layers have 64 and 32 filters, respectively. The size of the C1 layer is down sampled by pooling layer P1. Similarly, the size of the C2 layers is down sampled by layer P2. The fully connected layer F1 combines all feature maps of the C2 layer into a 1D form. Another fully connected layer F2 helps the output layer classify the input data into desired categories. The valid convolution technique utilized in this study allows the size of the feature maps to remain unchanged. Furthermore, the 2 dropout layers allow the network to generalize data for reducing the over-fitting problem [22,24].

3.4 Fine-Tuned Transfer Learning Framework

FTL built on CNN is adopted for measuring diagnostic performance. As described in Figure 1, the proposed CNN architecture is designed and fine-tuned by minimizing the objective function for the source task. Next, the fine-tuned model with learned parameters and optimized weights are transferred to the target task. Finally, in the target domain the model is adjusted and fine-tuned with the dataset of the target task. This way, the fine-tuned target neural model can attain better diagnostic performance [33]. It is important to mention that both sets of data (data from the source and target domains) are acquired from the same experimental testbed with similar acquisition approaches by varying the pressure conditions. The components of the proposed CNN with specifics regarding the transferrable layers are presented in Table 1.

Table 1. Elements of the proposed CNN with transfer measurements for the target network.

Layers	Parameters	Observations	Height	Width	Depth	Parameters Trainable	Transfer
Input		Preprocessed Signals	256	256	4		
Convolution 1	Kernel Size	Filter	3	3		Yes	Yes
	Padding	Zero					
	Depth	Filter number			64		
	Output		256	256	64		
Pool 1	Kernel Size	Filter	3	3		No	Yes
	Padding	Zero					
	Output		85	85	64		
Dropout	Output		85	85	64	No	Yes
Convolution 2	Kernel Size	Filter	3	3		Yes	Yes
	Padding	Zero					
	Depth	Filter number			32		
	Output		85	85	32		
Pool 2	Kernel Size	Filter	3	3		No	Yes
	Padding	Zero					
	Output		28	28	32		
Dropout	Output		28	28	32	No	Yes
FC 1	Nodes	Flatten into 1D	128			Yes	No
FC 2	Nodes	Flatten into 1D	64			Yes	No

SoftMax	Nodes	Flatten into 1D	2			Classify	No

4. Experimental Verification and Discussion

In this section, the proposed diagnostic framework (MSFTFI+FTL) is validated using data collected from a real-world spherical tank.

4.1 Dataset Description

The standard multi-sensor AE dataset of a spherical tank is utilized to conduct the experimental test. A 0.1 second signal with a 1 MHz sampling frequency is considered [12]. Two different pressure conditions (0 and 1 bar) are employed to record 1000 signals from each health condition (i.e., normal, and faulty). Descriptions of the datasets are provided in Table 2.

Table 2. Specifics of the datasets for the spherical tank.

	Health Type	**Crack Size (mm)**			**Sensors**	**Pressure (Bar)**	**Number of Samples**
		Length (mm)	**Width (mm)**	**Depth (mm)**			
Dataset 1	Normal Condition (NC)	N/A	N/A	N/A	4	0	1000
	Faulty Condition (FC)	3	0.5	0.4	4	0	1000
Dataset 2	Normal Condition (NC)	N/A	N/A	N/A	4	1	1000
	Faulty Condition (FC)	3	0.5	0.4	4	1	1000

4.2 MSFTFI-Based Performance Visualization of Cracks

In Figure 5, the results of the STFT analysis for two datasets with different health conditions are presented. According to the results in the figure, it is observed that in the normal condition (NC) for both datasets, the highest distributions of energy are concentrated into very similar frequency bands for all sensors. Similarly, when a fault occurs in both datasets, a few more significant energies bands are observed in certain frequency ranges. As depicted in Figure 5 for the NC, a strong energy distribution can be observed within a similar range for all sensors under all pressure conditions. Similarly, for the faulty condition (FC) the energy distribution is quite comparable for all sensors. A sample signal from the NC (sensor 1, pressure of 1 bar) is shown in Figure 6(a) to better illustrate the time, frequency, and attained time-frequency domains by STFT. In the frequency domain, two frequencies contain higher energies than the others. Therefore, in the STFT analysis, a strong energy distribution on those specific frequencies is observed with respect to time. As shown in Figure 6(b), a sample from the FC (sensor 1, pressure of 1 bar) has also been considered. From the attained frequency domain of this sample, it is observed that, when compared to the NC, few other frequency bands contain a higher energy distribution with respect to time.

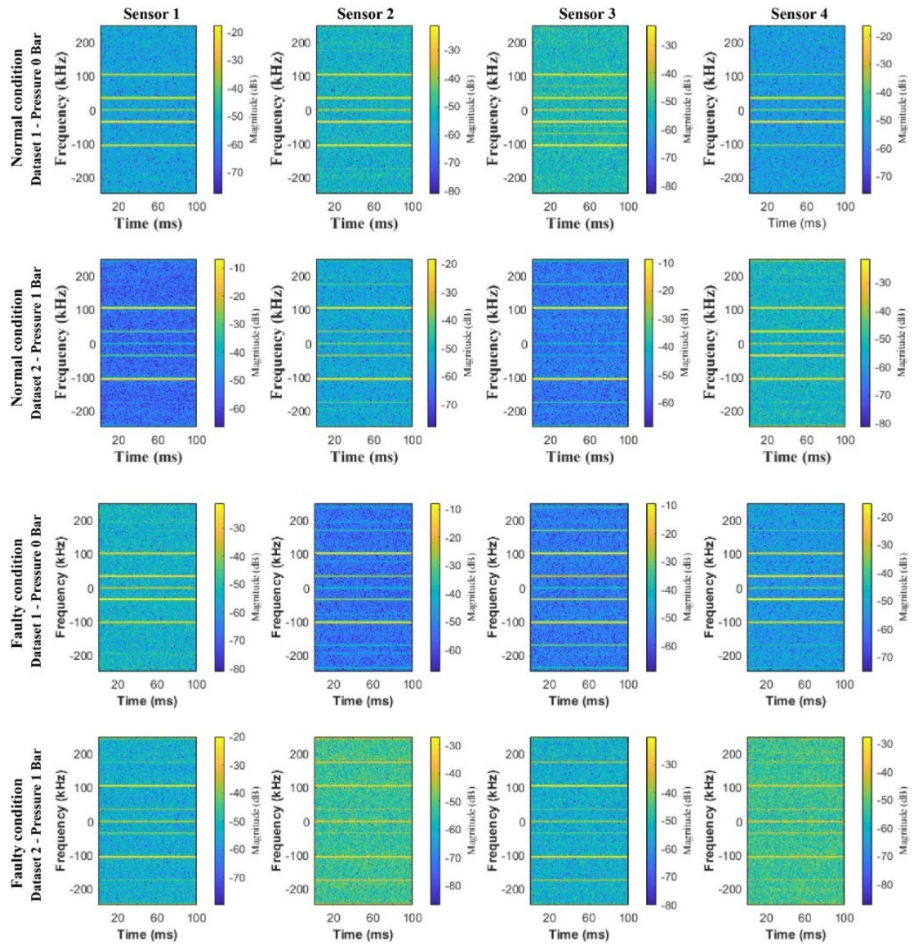

Figure 5: Short-time Fourier transform (STFT) images of different sensors at various health conditions.

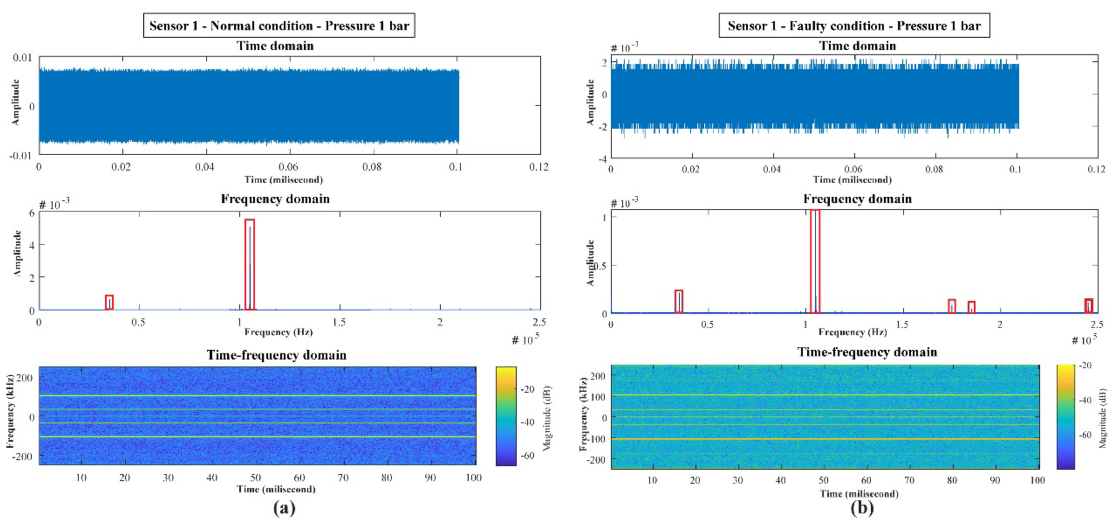

Figure 6: (a) Time, frequency, and time-frequency domain analysis of the sample considered from the normal condition (sensor 1, 1 bar of pressure) and (b) time, frequency, and time-frequency domain analysis of the sample considered from the faulty condition (sensor 1, 1 bar of pressure).

The extracted STFT images from all sensors (depicted into Figure 5) are fused together channel-wise (details are described into Section 3.2) to form the final MSFTFI image. Finally, the MSFTFI images are supplied to the proposed FTL-embedded CNN architecture for a final diagnosis.

4.3 FTL-Based Diagnostic Performance Analysis

The proposed MSFTFI framework is very useful for visualizing the state of the spherical tank. To further utilize the full benefits of MSFTFI images, FTL is proposed to diagnosis cracks in the spherical tanks under variable pressure conditions. To validate the proposed MSFTFI+FTL method, the dataset is divided into training and testing categories. Two scenarios were employed in this experiment. In scenario 1, dataset 1 is used for training the improved CNN architecture to gather knowledge as the source task. For this case, 70% of the data is utilized for training, 20% is employed for validation, and 10% is used for testing network performance before sharing the acquired knowledge with the target task. Next, dataset 2 is fed to the target task for final diagnosis using the shared knowledge learned from the source task. From dataset 2, 20% of the data is first used for adjusting the target network with shared knowledge from the source task. Consequently, the remaining 80% of the data is passed to the network for diagnostic purposes. Similarly, for scenario 2, dataset 2 is employed for the source task and dataset 1 is considered for the target task. To train, test, and validate all cases, 10-fold cross validation is used to remove bias from the diagnosis result [19]. For measuring diagnostic performance, the sensitivity score (SN) and average class sensitivity (avcSN) are calculated. The SN is calculated as follows [19]:

$$SN = \frac{true\_positive}{true\_positive + false\_negetive} \times 100\% \tag{9}$$

where the term "true positive" refers to correctly classified samples from the provided test data to the network at every iteration, while the term "false negative" refers to the number of samples from a class that are wrongly classified. The avcSN can be obtained as follows:

$$avcSN = \frac{\sum SN}{total\_classes} \tag{10}$$

where $\sum SN$ is a summation of the class-wise accuracy for the target dataset. To clarify the pressure variation situation, 2 scenarios are considered to measure diagnostic performance. In scenario 1, dataset 1 is considered as the source task and dataset 2 as the target task. The improved CNN is first trained and validated with dataset 1. After attaining 100% accuracy for both the training and validation data considered from dataset 1, the acquired knowledge is transferred to the target task. The performance of this training stage is illustrated in Figure 7(a), while the performance of the target task with dataset 2 is demonstrated in Figure 7(b) over 600 epochs. In a similar way, for scenario 2, dataset 2 is first considered as the source task to train and validate the network architecture, and the network parameters and architecture are subsequently used in the target domain to verify the diagnostic performance with dataset 1. The diagnostic performances of the proposed framework are listed in Table 3. According to the results in Table 3, the diagnostic performances are 100% both scenarios.

Table 3. Diagnostic performance of the proposed framework.

Scenario	Target Domain	Source Domain	Sensitivity (SN) %		Average Class Sensitivity (avcSN) %
			NC	FC	
1	Dataset 1	Dataset 2	100	100	100
2	Dataset 2	Dataset 1	100	100	100
Average (%)			**100**	**100**	**100**

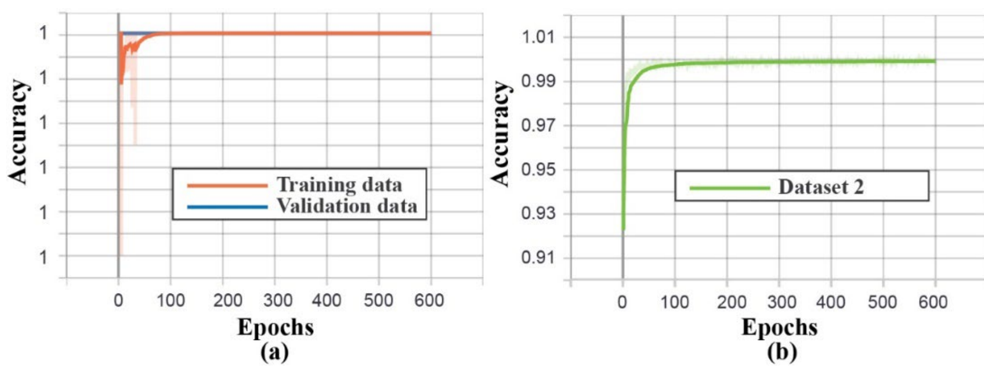

Figure 7: (a) Training vs. validation accuracy for the source domain (scenario 1: dataset 1) and **(b)** testing accuracy with transfer learning for the target domain (scenario 1: dataset 2).

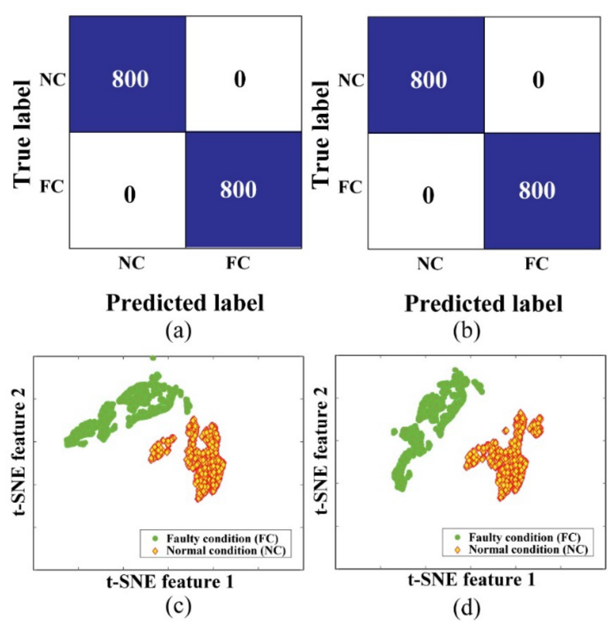

Figure 8: (a) Confusion matrix of scenario 1 (target task 1: dataset 2), **(b)** confusion matrix of scenario 2 (target task 2: dataset 1), **(c)** learned feature space of the proposed network in the target task (target task 1: dataset 2), and **(d)** learned feature space of the proposed network in the target task (target task 2: dataset 1).

To further validate the maximum diagnostic performance, the results of the confusion matrix [34] and feature space (visualized by t-stochastic neighbor embedding, t-SNE) final layer of FTL for the target domain in both scenarios are provided. The confusion matrix depicts the classification performance in the form of actual verse-predicted deviation. According to the results in Figure 8, the confusion matrix perfectly classifies all fault types with no error. Furthermore, the t-SNE-based feature distribution for both crack classes (NC, FC) are clearly separable, which also ensures better diagnostic performance. Besides, to confirm the efficiency of the proposed framework, several experiments are carried out. A diagnostic comparison between the FTL-embedded CNN and the CNN without FTL is performed. For this experiment, the CNN is trained with 20% of dataset 2 for scenario 1; the remaining data are used for testing performance. The train vs. test ratio has been kept constant to facilitate a comparison of diagnostic performance on a similar scale. From Figure 9(a), it is shown that the improved CNN without FTL does not perform as well (80.2% accuracy) as the proposed framework in the training phase. As displayed in Figure 9(b), the FTL-embedded CNN (proposed framework) can learn faster during the training phase.

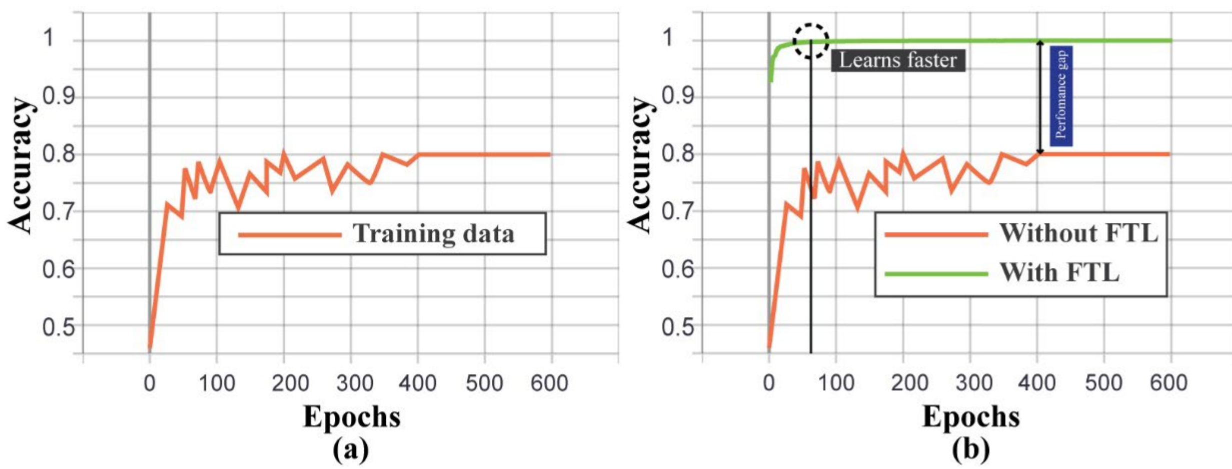

Figure 9: (a) The training accuracy typically achieved with dataset 2 (without TL, where train: test = 20:80) and **(b)** comparison of training accuracy for the two approaches (with and without FTL).

To demonstrate the robustness of the devised framework, the performance of the proposed method is compared to that achieved with two state-of-art approaches, namely (1) RAW+TL-CNN: a TL-based method where a raw 1D signal from a single sensor is fed to the network for measuring diagnostic performance [33] and (2) FE+k-NN: a traditional feature extraction-based approach where statistical features are first extracted from single-sensor data and a k-nearest neighborhood (k-NN) algorithm is used for the final diagnosis after reducing the features by principal component analysis (PCA) [35]. Details regarding the comparison results are presented in Table 4.

Table 4. Diagnostic performance comparison

		Sensitivity (SN) %		

Scenario	Method	NC	FC	Average Class Sensitivity (avcSN) %	Improvement (%)
1	**Proposed**	100	100	100	**-**
	RAW+TL-CNN [33]	93.82	94.91	94.37	**5.64**
	FE+k-NN [35]	89.21	90.23	89.72	**10.82**
2	**Proposed**	100	100	100	**-**
	RAW+TL-CNN [33]	93.62	95.61	94.62	**5.39**
	FE+k-NN [35]	90.57	91.21	90.89	**9.11**

The comparison findings show that the proposed framework (MSFTFI+FTL) clearly outperformed two state-of-the-art methods, yielding average performance improvements of 5.64 -10.82% and 5.39 - 9.11% for scenarios 1 and 2, respectively. The impact of noisy data on the diagnostic performance was also explored. Gaussian white noise with a signal to noise ratio (SNR) of 10 dB is added into test samples of the target task to simulate data with supplementary background noise. All comparable methods and the proposed scheme were first trained on original AE data in the source task. This was followed by testing and validation on noisy data created for the target task. The diagnostic performances of the proposed and comparable methods are listed in Figure 10.

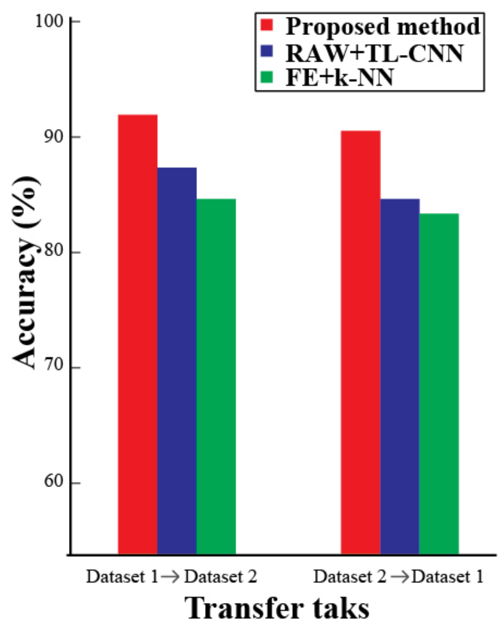

Figure 10: Impact of noisy data on classification performance.

From Figure 10, it can be stated that the diagnostic performance of all methods degrades due to the noisy dataset. However, the performance of the proposed framework is still better than that of the other two approaches considered for comparison.

5. Conclusion

This paper introduced a multi-sensor fusion-based imaging technique combined with fine-tuned transfer learning (FTL) built on a convolutional neural network (CNN) framework that augments a new diagnostic approach for spherical tank structural health monitoring. By incorporating a deep learning-based architecture with short-time Fourier transform (STFT) analysis, the proposed method makes full use of the capability of STFT to process non-stationary multi-sensory acoustic emission (AE) signals and enable an end-to-end diagnosis without handcrafted feature analysis. Data collected from a self-designed test rig are utilized to validate the diagnostic performance of the proposed approach. Experimental findings imply that the proposed approach can significantly enhance diagnostic performance and enable more rapid converging when compared to basic CNN-based models. The experimental results also indicate that the proposed framework (MSFTFI+FTL) clearly outperformed two state-of-the-art methods, yielding significant performance improvements.

At present, the proposed approach is confined to the fixed time-frequency resolution of STFTs. MSFTFI images with adaptive time-frequency resolution will be considered as inputs in future work. While the current framework belongs to the supervised learning paradigm, meaning that health states must be labeled in advance, the unsupervised learning paradigm could be a fascinating direction for future studies. Lastly, an assessment of the usefulness of the developed diagnostic framework will be performed for relevant applications such as boiler tubes, cylindrical pumps, and pipeline fault diagnosis.

Acknowledgment

This work was supported by the Technology Infrastructure Program funded by the Ministry of SMEs and Startups (MSS), Korea.

References

[1] J.I. Chang, C.-C. Lin, A study of storage tank accidents, J. Loss Prev. Process Ind. 19 (2006) 51–59.

[2] M. Sohaib, M. Islam, J. Kim, D.-C. Jeon, J.-M. Kim, Leakage Detection of a Spherical Water Storage Tank in a Chemical Industry Using Acoustic Emissions, Appl. Sci. 9 (2019) 196.

[3] C. Ennaceur, A. Laksimi, C. Herve, M. Cherfaoui, Monitoring crack growth in pressure vessel steels by the acoustic emission technique and the method of potential difference, Int. J. Press. Vessel. Pip. 83 (2006) 197–204.

[4] C.R. Farrar, K. Worden, An introduction to structural health monitoring, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 365 (2007) 303–315.

[5] G. Qian, V.F. González-Albuixech, M. Niffenegger, Effects of embedded cracks and residual stresses on the integrity of a reactor pressure vessel, Eng. Fail. Anal. 90 (2018) 451–462.

[6] P. Rizzo, I. Bartoli, A. Marzani, F. Lanza di Scalea, Defect classification in pipes by neural networks using

multiple guided ultrasonic wave features extracted after wavelet processing, (2005).

[7] P. Dong, Length scale of secondary stresses in fracture and fatigue, Int. J. Press. Vessel. Pip. 85 (2008) 128–143.

[8] M. Islam, M. Sohaib, J. Kim, J.-M. Kim, Crack Classification of a Pressure Vessel Using Feature Selection and Deep Learning Methods, Sensors. 18 (2018) 4379.

[9] K.A. Korkmaz, A. Sari, A.I. Carhoglu, Seismic risk assessment of storage tanks in Turkish industrial facilities, J. Loss Prev. Process Ind. 24 (2011) 314–320.

[10] S.A. Niknam, V. Songmene, Y.H.J. Au, The use of acoustic emission information to distinguish between dry and lubricated rolling element bearings in low-speed rotating machines, Int. J. Adv. Manuf. Technol. 69 (2013) 2679–2689.

[11] K. Morofuji, N. Tsui, M. Yamada, A. Maie, S. Yuyama, Z.W. Li, Quantitative Study of Acoustic Emission Due To Leaks From Water Tanks, Group. 21 (2003) 213–222.

[12] M.J. Hasan, J.M. Kim, Fault detection of a spherical tank using a genetic algorithm-based hybrid feature pool and k-nearest neighbor algorithm, Energies. 12 (2019).

[13] M.J. Hasan, J. Kim, C.H. Kim, J.-M. Kim, Health State Classification of a Spherical Tank Using a Hybrid Bag of Features and K-Nearest Neighbor, Appl. Sci. 10 (2020) 2525.

[14] H. Tao, P. Wang, Y. Chen, V. Stojanovic, H. Yang, An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks, J. Franklin Inst. (2020).

[15] M. M. Manjurul Islam and Jong-Myon Kim, Motor Bearing Fault Diagnosis Using Deep Convolutional Neural Networks with 2D Analysis of Vibration Sign, Springer International Publishing, 2018.

[16] M. Sohaib, C.-H. Kim, J.-M. Kim, A Hybrid Feature Model and Deep-Learning-Based Bearing Fault Diagnosis, Sensors. 17 (2017) 2876.

[17] J. Wang, Z. Mo, H. Zhang, Q. Miao, A deep learning method for bearing fault diagnosis based on time-frequency image, IEEE Access. 7 (2019) 42373–42383.

[18] M.J. Hasan, J.-M. Kim, Bearing Fault Diagnosis under Variable Rotational Speeds Using Stockwell Transform-Based Vibration Imaging and Transfer Learning, Appl. Sci. 8 (2018) 2357.

[19] M.J. Hasan, M.M.M. Islam, J.M. Kim, Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions, Meas. J. Int. Meas. Confed. 138 (2019) 620–631.

[20] M.J. Hasan, M. Sohaib, J.-M. Kim, 1D CNN-Based Transfer Learning Model for Bearing Fault Diagnosis Under Variable Working Conditions, in: Adv. Intell. Syst. Comput., 2019: pp. 13–23.

[21] Y. LeCun, L.D. Jackel, L. Bottou, C. Cortes, J.S. Denker, H. Drucker, I. Guyon, U.A. Muller, E. Sackinger, P. Simard, Learning algorithms for classification: A comparison on handwritten digit recognition, Neural Networks Stat. Mech. Perspect. 261 (1995) 276.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.

[23] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, ArXiv Prepr. ArXiv1502.03167. (2015).

[24] G.E. Dahl, T.N. Sainath, G.E. Hinton, Improving deep neural networks for LVCSR using rectified linear units and dropout, in: 2013 IEEE Int. Conf. Acoust. Speech Signal Process., IEEE, 2013: pp. 8609–8613.

[25] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement. 111 (2017) 1–10.

[26] J. Ma, F. Wu, J. Zhu, D. Xu, D. Kong, A pre-trained convolutional neural network based method for thyroid nodule diagnosis, Ultrasonics. 73 (2017) 221–230.

[27] N. Qian, On the momentum term in gradient descent learning algorithms, Neural Networks. 12 (1999) 145–151.

[28] physicalacoustics - sensors, (n.d.). https://www.physicalacoustics.com/by-product/sensors/WDI-AST-100-900-kHz-Wideband-Differential-AE-Sensor (accessed January 25, 2020).

[29] physicalacoustics - pci 2, (n.d.). https://www.physicalacoustics.com/by-product/pci-2/ (accessed January 25, 2020).

[30] D.-T. Hoang, H.-J. Kang, Rolling element bearing fault diagnosis using convolutional neural network and vibration image, Cogn. Syst. Res. 53 (2019) 42–50.

[31] M. Sohaib, J.-M. Kim, A robust deep learning based fault diagnosis of rotary machine bearings, Adv. Sci. Lett. 23 (2017) 12797–12801.

[32] M. Zhao, M. Kang, B. Tang, M. Pecht, Deep Residual Networks with Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes, IEEE Trans. Ind. Electron. 65 (2018) 4290–4300.

[33] R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer Learning with Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions, IEEE Access. 5 (2017) 14347–14357.

[34] A. Luque, A. Carrasco, A. Martín, A. de las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recognit. 91 (2019) 216–231.

[35] M. Kang, M.R. Islam, J. Kim, J.M. Kim, M. Pecht, A Hybrid Feature Selection Scheme for Reducing Diagnostic Performance Deterioration Caused by Outliers in Data-Driven Diagnostics, IEEE Trans. Ind. Electron. 63 (2016) 3299–3310.