electronics

MDPI

Article

Transformer-Based GAN with Multi-STFT for Rotating Machinery Vibration Data Analysis

Seokchae Lee [1], Hoejun Jeong [1] and Jangwoo Kwon [2,*]

[1] Department of Electric Computer Engineering, Inha University, Incheon 22212, Republic of Korea; sclee98@inha.edu (S.L.); lilmae@inha.edu (H.J.)
[2] Department of Computer Engineering, Inha University, Incheon 22212, Republic of Korea
* Correspondence: jwkwon@inha.ac.kr

Abstract: Prognostics and health management of general rotating machinery have been studied over time to improve system stability. Recently, the excellent abnormal diagnosis performance of artificial intelligence (AI) was demonstrated, and therefore, AI-based intelligent diagnosis is now being implemented in these systems. AI models are trained using large volumes of data. Therefore, we propose a transformer-based generative adversarial network (GAN) model with a multi-resolution short-time Fourier transform (multi-STFT) loss function to augment the vibration data of rotating machinery to facilitate the successful learning of deep learning models. We constructed a model with a conditional GAN structure, which is transformer based, for learning the feature points of vibration data in the time-series domain. In addition, we applied the multi-STFT loss function to capture the frequency features of the vibration data. The generated data, which adequately captured the frequency features, were used to augment the training data to improve the performance of a deep learning classifier. Furthermore, by visualizing the generated vibration data and comparing the visualizations to those of the vibration data obtained from real machinery, we demonstrated that the generated data were indistinguishable from the actual data.

Keywords: vibration; rotary machine; deep learning; GAN; generative model

check for updates

Citation: Lee, S.; Jeong, H.; Kwon, J. Transformer-Based GAN with Multi-STFT for Rotating Machinery Vibration Data Analysis. *Electronics* **2024**, *13*, 4253. https://doi.org/10.3390/electronics13214253

Academic Editors: Hao Xue and Du Huynh

Received: 20 September 2024
Revised: 28 October 2024
Accepted: 28 October 2024
Published: 30 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rotating machinery is one of the most widely used types of machinery in the industrial field. It is primarily utilized in devices such as automobiles, power plants, turbines, and pumping machines. Therefore, the failure of such machines can lead to system shutdowns due to mechanical malfunctions. To avoid such problems, various approaches have been developed to enhance the reliability of rotating machinery to prevent these issues [1,2]. In particular, approaches that integrate technologies such as digital instrumentation and control systems have emerged [3].

Recently, the field of prognostics and health management (PHM) has advanced, focusing on the development of diagnostic and prognostic models based on data-driven approaches [4,5]. In artificial intelligence (AI), machine learning-based anomaly detection methodologies, such as support vector machines, have been widely used for a long time [6–8]. Additionally, research on PHM models based on deep neural networks, such as recurrent neural networks and long short-term memory models, has been extensive [9–11].

One remarkable advancement is the emergence of Transformer networks. Originally developed for natural language processing, Transformers recently showed excellent performance in processing time-series data, including vibration analysis [12,13]. The self-attention mechanism of Transformers effectively captures long-range dependencies in vibration data, making them particularly useful for analyzing the complex patterns in rotating machinery's vibration signals. Several studies demonstrated that Transformer-based models can perform more accurate fault prediction and condition diagnosis than traditional RNN- or

CNN-based models [14–16]. These advancements create new possibilities for improving the accuracy and efficiency of PHM systems.

However, effectively applying AI to PHM systems requires large amounts of data to achieve robust feature identification and a high learning accuracy [17]. This requirement presents a significant challenge in implementing AI-based PHM systems, where the quantity and quality of data and the development of relevant data-processing techniques are considered essential for advancement. In learning-based methodologies, particularly deep learning, a dataset balanced between normal and abnormal data is needed [18]. For example, in plants, abnormal vibration data are relatively scarce compared with normal operation data, leading to a data imbalance within datasets [19]. Such datasets exhibit what is called a "long-tail distribution", where the scarcity of information about abnormal conditions can degrade the performance during algorithm training, which, in turn, can negatively affect the model accuracy and reliability in real operational environments.

Thus, addressing data imbalance is a critical consideration in the development of deep learning-based PHM systems. The most intuitive solution to data imbalance is developing techniques that use generative models to generate data with patterns similar to those of abnormal vibrations [20,21]. This approach effectively extends the dataset of abnormal-state data, thereby providing a balanced dataset and improving the algorithm training efficiency and performance. Data augmentation using generative models is expected to become an essential strategy for strengthening machine learning and deep learning algorithms, especially in scenarios where limited data are available.

Recently, data augmentation utilizing AI has garnered attention through the use of generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs) [22,23]. These techniques have been successfully applied in various domains, including images, time-series data, and text. While VAEs learn latent representations of data to generate new samples, GANs generate high-quality synthetic data through adversarial training between a generator and a discriminator network [24]. GAN-based approaches have shown promising results in augmenting complex time-series patterns, such as vibration data, and can significantly contribute to addressing the data imbalance issue in the PHM field [25]. Recently, GAN techniques that reflect frequency characteristics have also been developed to generate more realistic vibration data, further improving the fault detection performance in PHM systems [26].

The main contributions of this study are as follows:

(1) We propose a novel method for augmenting vibration data using a Transformer-based generative adversarial network (GAN). This approach combines the powerful time-domain feature extraction capabilities of Transformers with the data generation capabilities of GANs.

(2) We applied the multi-resolution short-time Fourier transform (multi-STFT) to effectively capture the frequency characteristics of vibration data and configured them for the learning process of the AI models.

(3) We demonstrated that combining a Transformer-based GAN with a multi-STFT-based loss function allowed for the finer modeling and generation of vibration data's time–frequency characteristics.

(4) We quantified and visualized the similarity between the data generated by the proposed model and the original data to compare and analyze their differences.

(5) We experimentally verified that the data generated by the proposed model contributed to improving the performance of the condition diagnosis for rotating machinery.

This study showed that the Transformer-based GAN could capture the various time-domain and frequency-domain characteristics of vibration data, which generated more realistic and useful data.

2. Related Work

Generative Models for Generating Time-Series Data

Traditionally, data augmentation methods have relied heavily on statistical approaches, such as scaling using the Whiten filter and mathematical transformations (e.g., logarithmic transformations) [27]. Additionally, starting from the late 1970s, statistical models based on the expectation maximization (EM) algorithm have been used for data augmentation [28]. This approach involves exploring the hidden structure of data and supplementing incomplete or insufficient datasets to increase the precision and reliability of analyses. These traditional data augmentation methods have served as the foundation for modern augmentation techniques based on data science and machine learning.

Various methods have been proposed to enhance the performance of statistical models. One such method entails combining parameter expansion and data augmentation to enhance the Gibbs sampling algorithm. In this approach, auxiliary variables are used to expedite the Gibbs sampling and optimize performance. Another notable development is the Monte Carlo EM (MCEM) algorithm, in which the E step of the EM algorithm is implemented using the Monte Carlo method [29]. From the data augmentation perspective, the MCEM provides an approximation of the posterior.

Driven by massive increases in computing power, deep learning has brought about revolutionary changes in data augmentation. As followed Figure 1, variational autoencoder (VAE) is a deep learning generative model that incorporates a probabilistic approach called maximum likelihood into the encoder-decoder structures of autoencoders to learn latent data distributions [30]. By using VAE, it is possible to generate new data not by reconstruction but by learning data distributions. Owing to the efficient data-reconstruction and generation capabilities of VAE, it has been used for tasks such as generating images from imbalanced data distributions and diagnosing anomalies by comparing the generated images to the original data [31]. The ability of VAE to learn and generate data based on probabilistic principles has opened up new possibilities for data augmentation and anomaly detection in various fields, including computer vision and natural language processing [32]. Moreover, the capacity of VAE to generate synthetic data that exhibit the same statistical properties as those the original data has made it a valuable tool in scenarios where data imbalance or scarcity presents challenges.

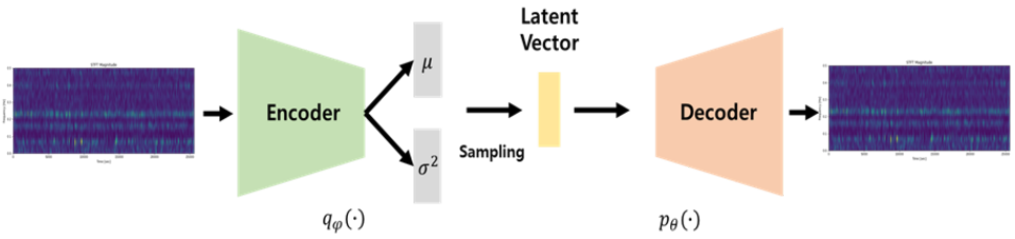


Figure 1. Architecture of Variational AutoEncoder (VAE).

A generative adversarial network (GAN), Figure 2, a highly popular deep learning model for data generation, trains two neural networks competitively [33]. These two neural networks are called the discriminator and generator. First, the generator receives random noise as input and creates data of the same size as the original data. The discriminator, which takes both the generated data and the original data as inputs, distinguishes whether they are from the original dataset or are generated fake data. The generator strives to deceive the discriminator by generating fake data that resemble the original data, while the discriminator aims to accurately differentiate between the fake data generated by the generator and the real data.

The discriminator is trained to output a high probability when given real data x as the input and a low probability when given generated fake data, and it updates itself to effectively distinguish between real and fake data. In other words, it learns to differentiate between the real data and the data produced by the generator. The generator, by contrast, is

trained to generate data G(z) from the random noise input z, which exhibits high probabilities similar to those of the real data when fed into the discriminator. The generator focuses on updating itself to generate data that the discriminator cannot distinguish from real data. In this way, the two neural networks engage in a competitive and adversarial relationship, with the generator gradually learning to produce fake data that are increasingly difficult for the discriminator to differentiate from real data [34].

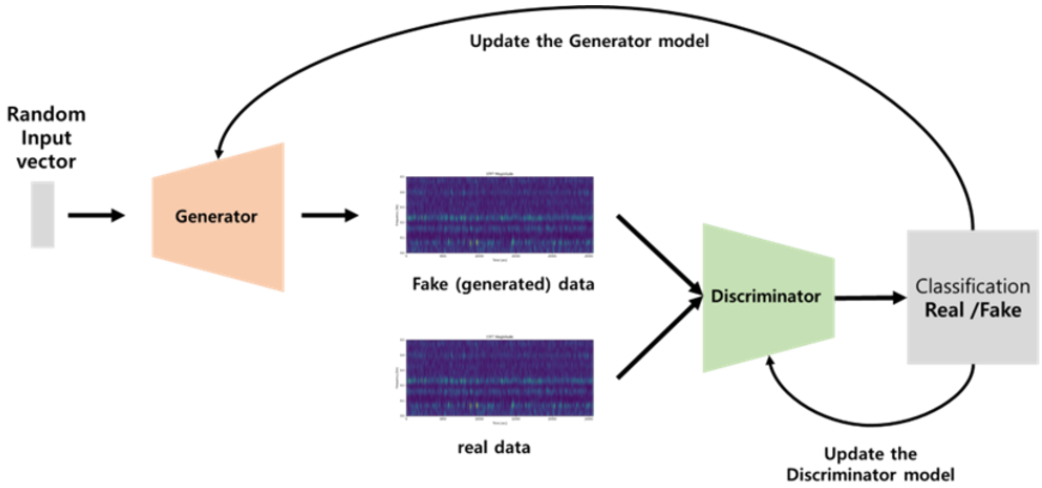


Figure 2. Architecture of Generatve Adversarial Network (GAN).

A GAN is known for its outstanding data generation performance, and its data augmentation capabilities have been demonstrated for addressing data imbalance issues through improvements in image classification accuracy [35]. Furthermore, when used in conjunction with traditional data augmentation techniques, such as masking or rotation, a GAN achieves a high classification accuracy. The application of a GAN for generating one-dimensional (1D) data has progressed at a slower rate compared with that for image generation. By using a type of GAN called a conditional GAN (cGAN) and adapting the generator and discriminator to the data format in terms of dimensions, electrocardiogram data and vibration data from railways have been augmented effectively [36]. A recent study of rotating equipment used biGAN generated data by adding an additional loss function based on the rotating speed [26].

3. Methodology

3.1. Transformer-Based Generative Adversarial Network

In this paper, a GAN-based data generation model is proposed for augmenting vibration data. The proposed model has three main features: (1) Transformer-based network architecture, (2) data generation using a cGAN structure, and (3) the application of multi-resolution STFT for an additional loss. The Transformer architecture is well-suited for processing sequence data, and it has been used extensively in recent studies for analyzing vibration data. This architecture is essentially a neural network algorithm that tracks the relationships within sequential data, such as words in a sentence, to learn the context and meaning [37]. The most crucial components of the transformer algorithm are the multi-head attention mechanism and positional encoding. The multi-head attention mechanism, as the name suggests, processes the attention mechanism in parallel with multiple heads (represented as "i"). It calculates attention as follows:

$$head_i = Attention(Q_i, K_i, V_i) = softmax(\frac{Q_i K_i^T}{\sqrt{d_k}})V_i \quad (1)$$

The attention mechanism employs three components: query, key, and value. The query of each input data element represents how the element interacts with other elements, thereby allowing for the computation of similarity with keys. The computed similarities are then used

as weights of the values, and ultimately, the output is generated as a weighted sum. Positional encoding assigns positional information to each element of an input sequence. It uses sine and cosine functions to compute the positional encoding values of each position. The calculated values are added to the embeddings of each input token. This allows each token to carry information about its position, aiding the model in understanding the order of the tokens. The transformer model, with its multi-head attention mechanism and positional encoding, has demonstrated exceptional performance, not only in natural language processing but also in fields such as image object segmentation and time-series data prediction. It continues to be researched and developed actively owing to its outstanding capabilities. Building on the advantages of the transformer architecture, in recent years, research was conducted to generate and synthesize human voices and vibration [38].

However, research on the synthesis or generation of the vibrations encountered in industrial facilities remains lacking. In this study, by using the transformer network, we constructed a generator and discriminator, as illustrated in Figure 3 below. The overall feature extraction structure utilizes only the encoder part of the transformer, and the structures of the two networks are quite similar. To enable the model to capture the time-domain characteristics, the raw input signal is fed directly into the network. Additionally, the part for upscaling the features extracted by the generator to generate the data uses deconvolution layers, while the discriminator uses linear layers to output a single value for identifying the source of the input signal.

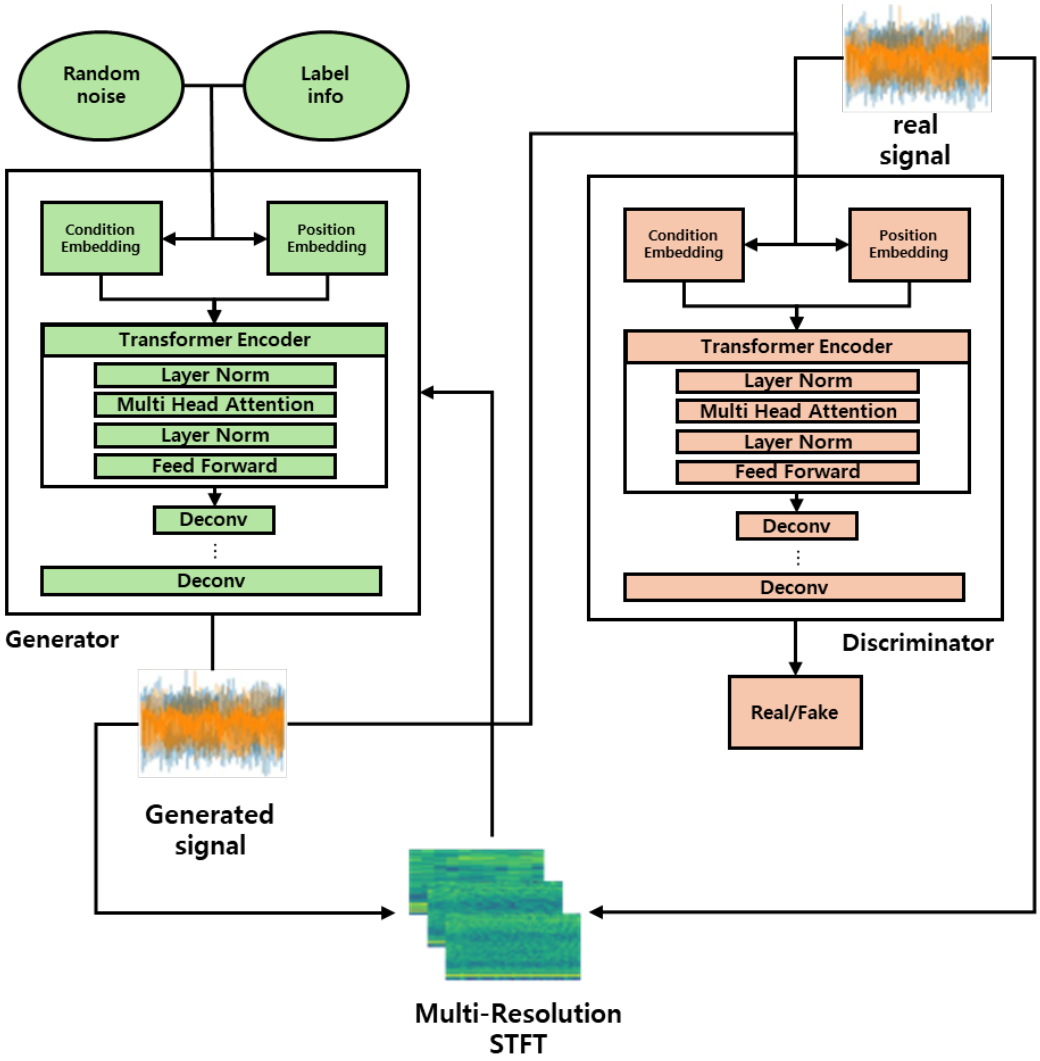


Figure 3. Architecture of proposed GAN with multi-resolution short-time Fourier transform (STFT) loss.

Table 1 shows the model architecture specifications and supplements Figure 3. The generator takes a latent vector z of dimension 100 as the input and transforms it through a linear layer into a tensor of sequence length 1024 and embedding dimension 10. Positional embedding is then added to allow the model to understand the sequence order. The transformed data passes through three Transformer encoder blocks, applying multi-head attention and feedforward networks in this process. Finally, the output is passed through a deconvolutional layer, converting it into generated signals of shape (batch_size,num_sensors,1,1024), where num_sensors represents the number of sensors.

Table 1. Model architecture specifications (supplementary to Figure 3).

Component	Generator	Discriminator
Input	Latent vector z (dimension: 100)	Generated or real signal
Output	Generated signal $(batch_size, num_sensors, 1, 1024)$	Binary classification result $(batch_size, 1)$
Key components	$nn.Linear$, $Transformer\ Encoder$, $Conv2d$	$PatchEmbedding$, $Transformer\ Encoder$, $ClassificationHead$
Dropout	0.5 (attention, feedforward)	0.5 (attention, feedforward)

The discriminator takes either the generated signal or the real signal as input, and the signal is divided into small patches through the PatchEmbedding_Linear. These patches are then mapped to the embedding space. After adding class tokens and positional encoding, the patch sequence is processed through three Transformer encoder blocks. Finally, the ClassificationHead reduces the sequence dimension and outputs binary classification results through a linear layer to distinguish whether the signal is real or generated.

Moreover, a cGAN architecture is used to generate the vibrations that occur in industrial facilities under various operating conditions. This cGAN architecture injects the class information of the desired vibration into the inputs of both networks. In the case of the generator, each vibration class label and Gaussian noise (z) are input as the conditions. These inputs are then passed through an embedding layer and a linear layer to synthesize them into a single feature vector. The generator uses this feature vector to generate a vibration signal. The discriminator takes the vibration sensor data and the condition as its inputs and determines whether the signal is real or artificially generated.

Many existing approaches to generating vibrations typically either learn and generate data solely from time-domain 1D signals or use STFT-transformed two-dimensional (2D) image data [39]. However, when only 1D signals are used, the generated data often exhibit inappropriate characteristics in the frequency domain because the model focuses primarily on learning the time-domain characteristics. In this study, 1D signals were generated, and an additional step was introduced. After the signal generation, the generated signal was subjected to STFT transformation by using various parameters, and a multi-STFT loss term was introduced to compare the STFT-transformed generated signal to the STFT-transformed original signal. This allowed the generator to learn not only the characteristics of the 1D signals but also those of the frequency domain. Notably, the STFT trades off between the frequency-domain resolution and time-domain resolution. Increasing both resolutions simultaneously can be challenging. In this study, a combination of various parameters was employed to facilitate the use of high-frequency-resolution STFT and high-time-resolution STFT to enhance the learning process.

3.2. Multi-Resolution Short-Time Fourier Transform Loss

$$STFTx(t)(f,\tau) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j2\pi ft}dt \tag{2}$$

In the case of an STFT, the magnitude varies across frequency bands depending on the size of the window used to divide a signal into segments. Accordingly, by adjusting the parameter called window size, the frequency characteristics represented by an STFT can be altered.

$$Multi-STFTx(t)(f,\tau,w) = \int_{-\infty}^{\infty} x(t)w_w(t-\tau)e^{-j2\pi ft}dt \tag{3}$$

From this perspective, the multi-STFT transformation is introduced to capture various frequency characteristics across different segments [40]. This transformation is an extension of the STFT transformation for multiple resolutions, and it facilitates the capture of frequency characteristics by using multiple window sizes, even when precise periodicity information about the input sequence is unavailable. This enhances the sophistication of the signal analysis.

The following Figure 4 (left) shows the results of an STFT transformation applied to some samples of the data used in the experiment with window sizes of 32, 128, and 512. Depending on the window size, the resolution changed, which allowed us to visually observe how the frequency characteristics varied. With smaller window sizes, detailed frequency features could be captured, while larger window sizes could capture overall frequency characteristics that may not be discernible with smaller window sizes. Therefore, we aimed to reflect the learning process of the model that proposes multi-window STFT transformations.

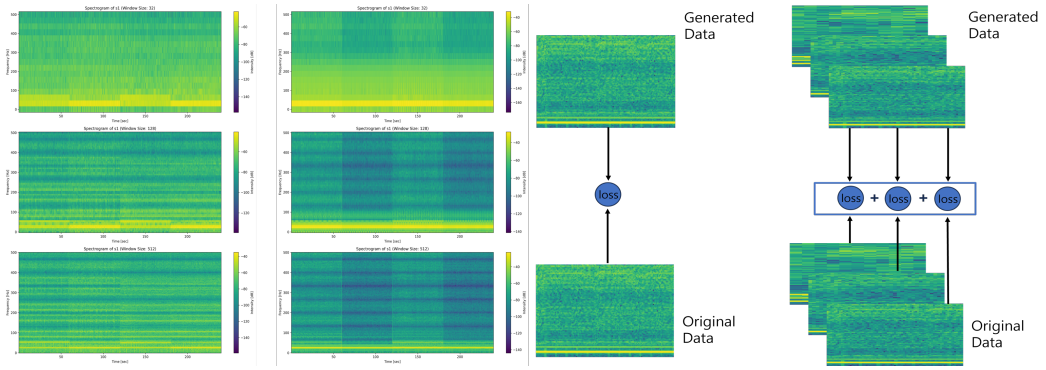


Figure 4. Comparison that STFT each resolution (**Left**), Single STFT loss & Multi STFT loss (**Right**).

In this study, a multi-STFT loss function was introduced to increase the stability of the adversarial learning process. Figure 4 (right) shows the difference between the single-STFT and multi-STFT losses. This multi-STFT function consists of the sum of several losses with different analysis parameters (e.g., fast Fourier transform size, window size, and frame shift). This approach helps to optimize the balance between the time and frequency resolutions. For example, increasing the window size reduces the time resolution but improves the frequency resolution. The proposed multi-STFT loss function prevents the generator from overfitting the fixed STFT representation, leading to more effective learning results.

4. Experiment

4.1. Evaluation Metrics

Jensen–Shannon (JS) divergence is one of the methods for measuring the similarity between two probability distributions P and Q [41]. The equation for two divergences is as follows:

$$D_{JS}(P||Q) = \frac{1}{2}D(P||\frac{1}{2}(P+Q)) + \frac{1}{2}D_{KL}(Q||\frac{1}{2}(P+Q)) \quad (4)$$

$$D_{KL}(P||Q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) \quad (5)$$

The fact that the Kullback–Leibler (KL) divergence cannot be used as a distance metric is an important issue affecting its utilization as an evaluation metric. This is because the resulting value can vary depending on the selected reference probability distribution. Consequently, KL divergence may not be an ideal performance measure for evaluating generative models. By contrast, JS divergence, which is based on KL divergence, is symmetric and invariably has a finite value, which makes it quantifiable. In this study, we numerically evaluated the

similarity between the distributions of generated data and real data by using JS divergence, which has these features.

4.2. Dataset

We utilized a dataset that represented various types of failures in rotating machinery, which was the 'Rotating Machinery Fault Type AI Dataset' (Ministry of SMEs and Startups, Korea AI Manufacturing Platform (KAMP), KAIST, 23 December 2022). The machinery structures used in power plants operate mostly under normal conditions, and the proportion of normal-type data in this dataset exceeded 90 percent. Therefore, this dataset was considered suitable for analyzing the characteristics of mechanical structure failure types and for gathering data corresponding to various failure types by using a purpose-built rotor testbed, Figure 5.

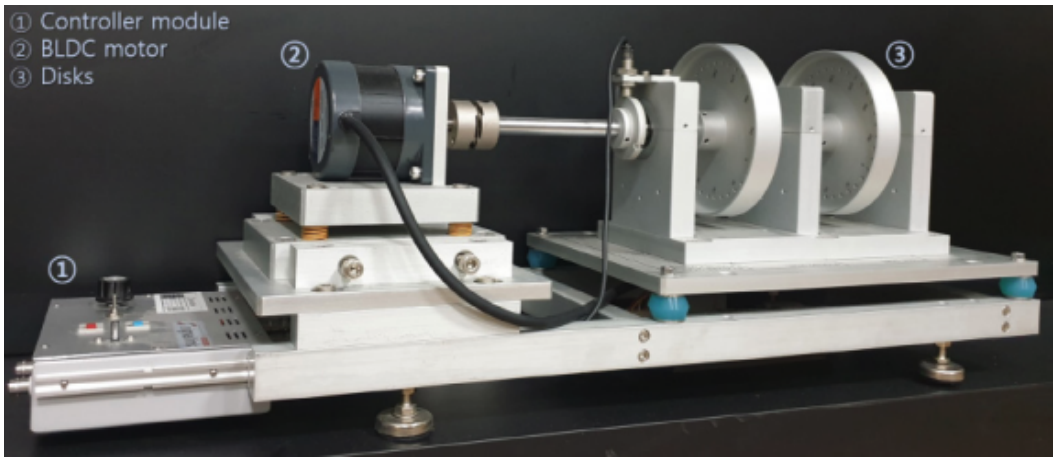


Figure 5. Components of Rotor testbed.

As followed Table 2, the speed of the rotor testbed could be adjusted from 0 to 3000 rpm. In this study, data were acquired at a rotor speed of approximately 1500 rpm. In the analysis, we utilized four sensors, and over the course of 140 s, we collected 3,772,385 data points. The dataset comprised data corresponding to four types of conditions: normal, mass imbalance, mechanical looseness, and a combination of mass imbalance and mechanical looseness. Mass imbalance was considered to occur when the centers of mass of the rotor and motor were misaligned, while mechanical looseness was considered to occur when the rotor was not properly secured or was tilted.

Table 2. Rotor testbed specifications.

Specification	Details
Size	673 mm (W) × 280 mm (D) × 281 mm (H)
Weight	25 kg
Material	Aluminum
Bearing	6202ZZ × 2EA
Motor	DC 12 V × 0.25 HP (0.2 W), 0~3000 RPM
Main power	220 VAC

The sensor used for data acquisition on the rotor testbed was a smart vibration sensor manufactured by Signallink Co., Ltd. (Suwon-City, Republic of Korea), as shown in Figure 6. Detailed specifications can be found in Table 3. The sensor collected data from the three bars that supported the rotor disk.

The collected sensor data followed several preprocessing steps to ensure the consistency and accuracy of the dataset. First, linear interpolation was applied to synchronize the measurement times of sensors with different sampling periods. This allowed for the

unification of the time axis, enabling a comparison between the sensor data. Next, moving average filtering was utilized to remove noise from the data while preserving the dynamic characteristics of the signal. This method effectively reduced transient fluctuations while maintaining the key features of the signal. Finally, the sensor data were transformed using min–max normalization to suit the requirements of machine learning algorithms. This technique scaled the data values between 0 and 1, which prevented distortions in the learning performance due to differences in the data magnitudes.



Figure 6. Sensor for data acquisition.

Table 3. Sensor specifications.

Specification	Details
Size	44 mm (W) × 44 mm (D) × 25 mm (H)
Weight	About 40 gf
Material	PC
Fix	M4 Bolt or magnetic
Frequency span	672 Hz
Frequency resolution	1.3125 Hz

4.3. Implementation

The proposed model was implemented using the PyTorch framework, and experiments were conducted on a computer with the following environment. The generator and discriminator were built based on the encoder part of the transformer structure, and a 1D convolutional neural network (CNN) structure was used to generate data or extract features. Only the last layer of the discriminator included a sigmoid layer for the binary classification of real and fake data.

The training was performed in batches with a batch size of 32, and an additional loss function, multi-STFT, was used in the generator. Consequently, the learning rate of the generator was slower than that of the discriminator, which introduced the risk of model collapse.

Ablation studies were conducted to compare the performance of the proposed model with those of other models based on the GAN architecture. In total, six experiments were set up, and these experiments differed in terms of the model and use of the proposed multi-STFT loss function. In each of these experiments, training and validation were performed for 50 epochs under the same conditions.

4.4. Model Training Convergence

In this section, we analyze the convergence of the proposed model during the training process. Figure 7 illustrates the training results over 150 epochs, showing the variations in the loss functions for both the discriminator and the generator. The discriminator's loss (loss_D) is represented in the left graph of Figure 7. Initially, the loss value experienced a rapid decrease, followed by a gradual decline, indicating a continuous improvement in the

discriminator's performance. Notably, after epoch 20, the loss stabilized, suggesting that the model had reached a convergent state.

Conversely, the generator's loss (loss_G) is depicted in the right graph of Figure 7. Similar to the discriminator, the generator's loss showed a significant reduction in the initial epochs, followed by a gradual tapering off. As the epochs progressed, the loss tended to stabilize, indicating that the generator progressively produced more realistic data.

Overall, both loss functions exhibited a decreasing trend, demonstrating that the proposed model effectively converged during the training process. These findings can be interpreted as positive indicators of the model's future performance. Through this analysis, we could clearly describe and evaluate the convergence and stability of the proposed model during its training.

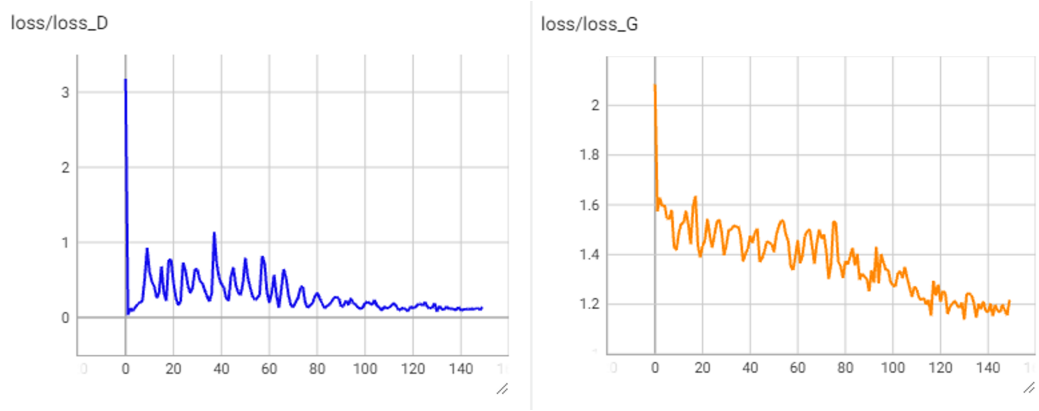


Figure 7. Proposed model training loss.

4.5. Ablation Study

The proposed model was constructed as a GAN based on the encoder structure of the existing transformer architecture, and additional loss functions were introduced to improve the model performance. The additional function was used to compare using multi-STFT to capture the frequency characteristics of the model.

To evaluate the performance of the proposed model in terms of these two aspects, the transformer-based model and additional loss, an ablation study was conducted to verify the degree of improvement of the proposed model relative to other models with GAN-based structures. First, wGAN (Wasserstein GAN) and the Transformer-based GAN were used to investigate the effects of changes in the model structure on the data generation performance [42]. Additionally, to analyze the changes in performance that resulted from the use of the STFT loss function, three conditions were set: not using any loss function, using the single-STFT loss function, and using the multi-STFT loss function. A total of six experiments were conducted, and the results of all experiments were analyzed quantitatively in terms of the JS divergence. The results of the experiments are summarized in Table 4.

Table 4. Comparison of data distribution differences in terms of JS divergence.

Model Name	STFT Loss	JS-Div
WGAN	-	0.29
WGAN	Single-STFT	0.244
WGAN	Multi-STFT	0.229
Transformer-based GAN	-	0.156
Transformer-based GAN	Single-STFT	0.135
Transformer-based GAN	Multi-STFT	0.096

As presented in Table 4, the average JS divergence in the experiments conducted using the wGAN model was 0.254, whereas the average JS divergence in the experiments

conducted using the proposed model, that is, the transformer-based GAN, was 0.129. Therefore, on average, the proposed model yielded a performance improvement of approximately 50.78%. Specifically, in the performance evaluation based on the distribution difference of each model when not using any STFT loss function, the transformer-based GAN model yielded a 53.79% improvement over the wGAN model, indicating that the transformer-based GAN model was more suitable for data generation.

In addition, the average JS divergence when using the multi-STFT loss function was 0.162, that when the function was not applied was 0.223, and that when using the single-STFT loss function was 0.189. These values confirmed that the performance improved by approximately 72.64% when using the multi-STFT loss function compared with not using it. Moreover, these experiments demonstrated that the two methodologies proposed in this study improved the model performance. Therefore, the use of domain-specific transformation techniques in data-driven deep learning models can lead to a higher performance compared with that when using raw data alone.

4.6. Validation of Generated Data Through Training

To check whether the vibration data generated using the proposed model were effective as input data for actual model training, we selected a representative deep learning model and conducted an experiment. The selected deep learning model was a 1D-CNN model that consisted of four 1D convolution layers and one fully connected layer [43]. In the first experimental case, only the real collected data were used without generating any synthetic data. In the second and third experimental cases, ablation studies were conducted with the synthetic data generated using the wGAN-based model and transformer-based GAN model, respectively.

These data were grouped into four classes: normal, imbalance, mechanical looseness, and a combination of imbalance and mechanical looseness. For the quantitative performance evaluation, we used the average accuracy. The results are summarized in Table 5.

Table 5. Comparison of classification scores obtained using a 1D convolutional neural network classifier.

Model Name	STFT Loss	Acc (%)
No generation	-	75.65
wGAN	-	76.39
wGAN	Single-STFT	77.41
wGAN	Multi-STFT	86.22
Transformer-based GAN	-	76.52
Transformer-based GAN	Single-STFT	79.13
Transformer-based GAN	Multi-STFT	85.79

According to the results, an accuracy of 75.65% was achieved when using the trained classifier only on the real data. By contrast, when using the transformer-based GAN with the multi-STFT loss function to generate synthetic data, which was the best-performing case, an accuracy of 85.79% was achieved, representing a performance improvement of more than 10%. These results indicate that in scenarios with limited data availability, it is more efficient to generate data by using the proposed approach and use the generated data for classification training.

4.7. Application to Real Power Plant Vibration Data

The future challenge of this study is to contribute to intelligent fault diagnosis by augmenting the vibration fault data of the internal machinery used in a power plant rather than using the data of general rotating machines. Therefore, we collected vibration data of the internal rotating machinery of an operational power plant and used them to train the proposed model and demonstrate its robustness to real data. The accelerometer sensor data were configured to collect data from six channels at a high sample rate of 50,000 Hz, as

shown in Table 6. Data were stored as soon as a trigger signal was received, with a pre-time of 1000 ms to capture the relevant context before the trigger event.

Table 6. Sensor data collection specifications.

Parameter	Value
Number of channels	6
Sample rate (Hz)	50,000
Store type	Fast on trigger
Pre-time (ms)	1000

Table 7 lists the results of the experiments conducted using the proposed model with and without the STFT loss function.

Table 7. JS divergence of real nuclear power plant data.

Model Name	STFT Loss	JS-Div
Transformer-based GAN	-	0.255
Transformer-based GAN	Single-STFT	0.238
Transformer-based GAN	Multi-STFT	0.229

Although the overall score was lower than that obtained using the data employed in the ablation study, the JS divergence was the highest at 0.229 when using the model with the proposed multi-STFT loss function.

In addition, we used kernel density estimation to visualize the data distributions obtained using the two model variants (with and without the proposed loss function) as continuous curves [44]. This step facilitated a straightforward comparison of the differences between the distributions.

To construct this plot, we used a Gaussian kernel to estimate and visualize the density function of each dataset. As illustrated in Figure 8, the first peak of the data generated using the model with the single-STFT loss function was smoothed out compared with that of the real data. However, when using the model with the proposed multi-STFT loss function, the data distribution was similar to that of the original data. Therefore, the proposed multi-STFT loss function performed well even when real power plant data were used for the model training.

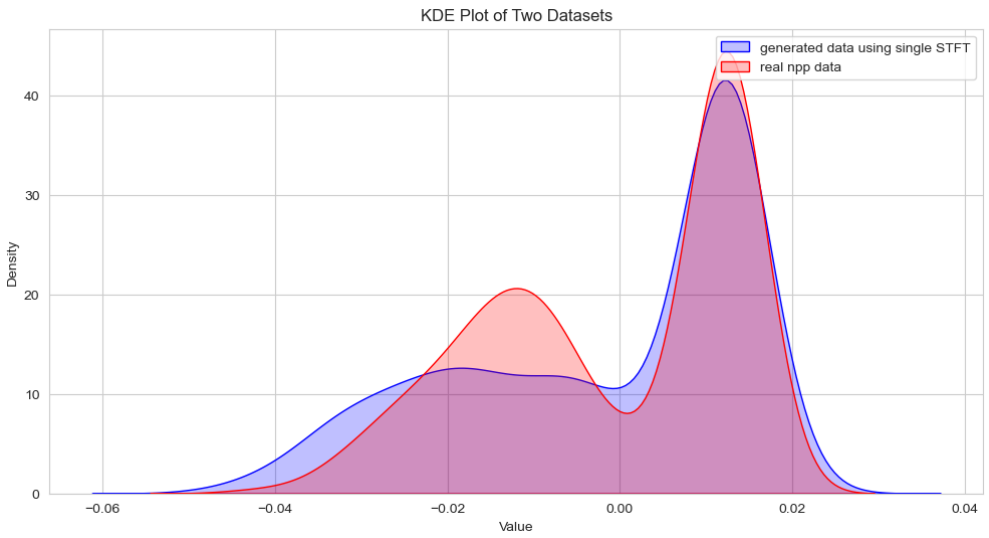


Figure 8. *Cont.*

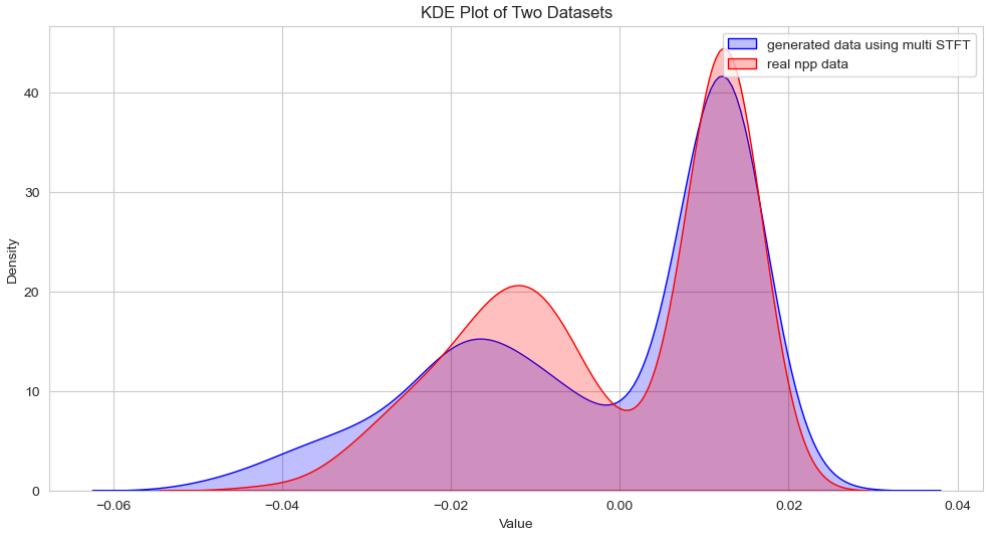


Figure 8. Distributions obtained through kernel density estimations.

5. Conclusions

In this study, we proposed an approach to generate data for rotating machinery by using a transformer-based GAN model. The proposed model leverages the frequency-domain characteristics of the data to ensure that the deep learning model can adequately capture frequency features. As a result, the model demonstrated a JS divergence of over 50.78% compared with the baseline model, wGAN, which allowed it to generate all types of anomaly and normal data for rotating machinery with a high degree of similarity to real data.

We applied multi-STFT loss functions in our experiments and observed that the proposed model improved the classifier accuracy by 10%, indicating that the data generated by our model contributed to an enhanced classifier performance. Furthermore, we validated the robustness and performance of the proposed model by training it with real-world data from a power plant. The similarity of the data distributions was visually represented using a KDE plot, which recorded a JS divergence of 0.229.

In the future, we aim to acquire even small amounts of real anomaly data from power plants, generate synthetic anomaly data, and derive anomaly diagnosis performances based on this data.

Author Contributions: Conceptualization, S.L. and H.J.; methodology, S.L. and J.K.; software, H.J.; validation, S.L. and H.J.; formal analysis, S.L.; investigation, S.L.; resources, J.K.; data curation, S.L.; writing—original draft preparation, S.L. and H.J.; writing—review and editing, H.J.; visualization, S.L.; supervision, J.K.; project administration, S.L.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MOTIE) (20224B10100060, Development of Artificial Intelligence Vibration Monitoring System for Rotating Machinery).

Data Availability Statement: We utilized a dataset representing various types of failures in rotating machinery, which was the 'Rotating Machinery Fault Type AI Dataset' (Ministry of SMEs and Startups, Korea AI Manufacturing Platform (KAMP), KAIST), accessed on 23 December 2022.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Coble, J.B.; Ramuhalli, P.; Bond, L.J.; Hines, W.; Upadhyaya, B. Prognostics and Health Management in Nuclear Power Plants: A Review of Technologies and Applications. 2012. Available online: https://www.osti.gov/biblio/1047416 (accessed on 17 July 2012).

2. Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement* **2013**, *46*, 1551–1564. [CrossRef]
3. Hocken, R.J.; Pereira, P.H. (Eds.) *Coordinate Measuring Machines and Systems*; CRC Press: Boca Raton, FL, USA, 2012; Volume 6.
4. Zhao, X.; Kim, J.; Warns, K.; Wang, X.; Ramuhalli, P.; Cetiner, S.; Kang, H.G.; Golay, M. Prognostics and health management in nuclear power plants: An updated method-centric review with special focus on data-driven methods. *Front. Energy Res.* **2021**, *9*, 696785. [CrossRef]
5. Qiao, W.; Lu, D. A survey on wind turbine condition monitoring and fault diagnosis—Part II: Signals and signal processing methods. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6546–6557. [CrossRef]
6. Soualhi, A.; Medjaher, K.; Zerhouni, N. Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE Trans. Instrum. Meas.* **2014**, *64*, 52–62. [CrossRef]
7. Samanta, B. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mech. Syst. Signal Process.* **2004**, *18*, 625–644. [CrossRef]
8. Li, Q.; Li, H.; Hu, W.; Sun, S.; Qin, Z.; Chu, F. Transparent Operator Network: A Fully Interpretable Network Incorporating Learnable Wavelet Operator for Intelligent Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2024**, *20*, 8628–8638. [CrossRef]
9. Hamadache, M.; Jung, J.H.; Park, J.; Youn, B.D. A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: Shallow and deep learning. *JMST Adv.* **2019**, *1*, 125–151. [CrossRef]
10. Jalayer, M.; Orsenigo, C.; Vercellis, C. Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, Fast Fourier and continuous wavelet transforms. *Comput. Ind.* **2021**, *125*, 103378. [CrossRef]
11. Ha, J.M.; Fink, O. Domain knowledge-informed synthetic fault sample generation with health data map for cross-domain planetary gearbox fault diagnosis. *Mech. Syst. Signal Process.* **2023**, *202*, 110680. [CrossRef]
12. Zerveas, G.; Jayaraman, S.; Patel, D.; Bhamidipaty, A.; Eickhoff, C. A transformer-based framework for multivariate time series representation learning. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; pp. 2114–2124.
13. Hong, K.; Jin, M.; Huang, H. Transformer winding fault diagnosis using vibration image and deep learning. *IEEE Trans. Power Deliv.* **2020**, *36*, 676–685. [CrossRef]
14. Jin, C.c.; Chen, X. An end-to-end framework combining time–frequency expert knowledge and modified transformer networks for vibration signal classification. *Expert Syst. Appl.* **2021**, *171*, 114570. [CrossRef]
15. Xie, J.; Zhang, J.; Sun, J.; Ma, Z.; Qin, L.; Li, G.; Zhou, H.; Zhan, Y. A transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2126–2136. [CrossRef] [PubMed]
16. Zollanvari, A.; Kunanbayev, K.; Bitaghsir, S.A.; Bagheri, M. Transformer fault prognosis using deep recurrent neural network over vibration signals. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 2502011. [CrossRef]
17. Chen, X.W.; Lin, X. Big data deep learning: Challenges and perspectives. *IEEE Access* **2014**, *2*, 514–525. [CrossRef]
18. Japkowicz, N.; Stephen, S. The class imbalance problem: A systematic study. *Intell. Data Anal.* **2002**, *6*, 429–449. [CrossRef]
19. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [CrossRef]
20. Douzas, G.; Bacao, F. Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Syst. Appl.* **2018**, *91*, 464–471. [CrossRef]
21. Ferreira, J.; Ferro, M.; Fernandes, B.; Valenca, M.; Bastos-Filho, C.; Barros, P. Extreme learning machine autoencoder for data augmentation. In Proceedings of the 2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Arequipa, Peru, 8–10 November 2017; pp. 1–6.
22. Liu, C.; Antypenko, R.; Sushko, I.; Zakharchenko, O. Intrusion detection system after data augmentation schemes based on the VAE and CVAE. *IEEE Trans. Reliab.* **2022**, *71*, 1000–1010. [CrossRef]
23. Bouallegue, G.; Djemal, R. EEG data augmentation using Wasserstein GAN. In Proceedings of the 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Sfax, Tunisia, 20–22 December 2020; pp. 40–45.
24. Frid-Adar, M.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. Synthetic data augmentation using GAN for improved liver lesion classification. In Proceedings of the 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 289–293.
25. Fu, Q.; Wang, H. A novel deep learning system with data augmentation for machine fault diagnosis from vibration signals. *Appl. Sci.* **2020**, *10*, 5765. [CrossRef]
26. Jeong, H.; Jeung, S.; Lee, H.; Kwon, J. BiVi-GAN: Bivariate Vibration GAN. *Sensors* **2024**, *24*, 1765. [CrossRef]
27. Liu, J.S.; Wu, Y.N. Parameter expansion for data augmentation. *J. Am. Stat. Assoc.* **1999**, *94*, 1264–1274. [CrossRef]
28. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
29. Wei, G.C.; Tanner, M.A. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Stat. Assoc.* **1990**, *85*, 699–704. [CrossRef]
30. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.

31. Gidaris, S.; Komodakis, N. Generating classification weights with gnn denoising autoencoders for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 21–30.
32. Pol, A.A.; Berger, V.; Germain, C.; Cerminara, G.; Pierini, M. Anomaly detection with conditional variational autoencoders. In Proceedings of the 2019 18th IEEE international conference on machine learning and applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1651–1657.
33. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
34. Bowles, C.; Chen, L.; Guerrero, R.; Bentley, P.; Gunn, R.; Hammers, A.; Dickie, D.A.; Hernández, M.V.; Wardlaw, J.; Rueckert, D. Gan augmentation: Augmenting training data using generative adversarial networks. *arXiv* **2018**, arXiv:1810.10863.
35. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **2018**, *321*, 321–331. [CrossRef]
36. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
37. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
38. Li, X.; Metsis, V.; Wang, H.; Ngu, A.H.H. Tts-gan: A transformer-based time-series generative adversarial network. In *International Conference on Artificial Intelligence in Medicine*; Springer: Cham, Switzerland, 2022; pp. 133–143.
39. Huang, J.; Chen, B.; Yao, B.; He, W. ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access* **2019**, *7*, 92871–92880. [CrossRef]
40. Yamamoto, R.; Song, E.; Kim, J.M. Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 6199–6203.
41. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [CrossRef]
42. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 214–223.
43. Khodabandehlou, H.; Pekcan, G.; Fadali, M.S. Vibration-based structural condition assessment using convolution neural networks. *Struct. Control Health Monit.* **2019**, *26*, e2308. [CrossRef]
44. Węglarczyk, S. Kernel density estimation and its application. In *ITM Web of Conferences*; EDP Sciences: Les Ulis, France, 2018; Volume 23, p. 00037.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.