POLITECNICO DI TORINO
Repository ISTITUZIONALE

Zero-Shot Generative AI for Rotating Machinery Fault Diagnosis: Synthesizing Highly Realistic Training Data via Cycle-Consistent Adversarial Networks

Original
Zero-Shot Generative AI for Rotating Machinery Fault Diagnosis: Synthesizing Highly Realistic Training Data via Cycle-Consistent Adversarial Networks / DI MAGGIO, L.G., Brusa, E., Delprete, C.. - In: APPLIED SCIENCES. - ISSN 2076-3417. - ELETTRONICO. - 13:22(2023). [10.3390/app132212458]

Availability:
This version is available at: 11583/2984352 since: 2023-12-05T14:20:38Z

Publisher:
Multidisciplinary Digital Publishing Institute (MDPI)

Published
DOI:10.3390/app132212458

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

28 November 2026

applied sciences

MDPI

Article

Zero-Shot Generative AI for Rotating Machinery Fault Diagnosis: Synthesizing Highly Realistic Training Data via Cycle-Consistent Adversarial Networks

Luigi Gianpio Di Maggio *, Eugenio Brusa * and Cristiana Delprete

Dipartimento di Ingegneria Meccanica e Aerospaziale (DIMEAS), Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Torino, Italy; cristiana.delprete@polito.it
* Correspondence: luigi.dimaggio@polito.it (L.G.D.M.); eugenio.brusa@polito.it (E.B.)

Abstract: The Intelligent Fault Diagnosis of rotating machinery calls for a substantial amount of training data, posing challenges in acquiring such data for damaged industrial machinery. This paper presents a novel approach for generating synthetic data using a Generative Adversarial Network (GAN) with cycle consistency loss function known as cycleGAN. The proposed method aims to generate synthetic data that could effectively replace real experimental data. The generative model is trained to transform wavelet images of simulated vibrational signals into authentic data obtained from machinery with damaged bearings. The utilization of Maximum Mean Discrepancy (MMD) and Fréchet Inception Distance (FID) demonstrates a noteworthy resemblance between synthetic and real experimental data. Also, the generative model enables the synthesis of data that may have been entirely lacking from the experimental observation, indicating generative zero-shot learning capabilities. The efficacy of synthetic data in training diagnosis algorithms by means of Transfer Learning (TL) on Convolutional Neural Networks (CNNs) has been demonstrated to be comparable to that of real data. The study has been validated by means of the test rig for medium-sized industrial bearings accessible at the Politecnico di Torino.

Keywords: intelligent fault diagnosis; generative adversarial networks; cycleGANs; transfer learning; machine fault diagnosis; convolutional neural networks; maximum mean discrepancy; bearings; rotating machinery; condition monitoring

check for updates

Citation: Di Maggio, L.G.; Brusa, E.; Delprete, C. Zero-Shot Generative AI for Rotating Machinery Fault Diagnosis: Synthesizing Highly Realistic Training Data via Cycle-Consistent Adversarial Networks. *Appl. Sci.* **2023**, *13*, 12458. https://doi.org/10.3390/app132212458

Academic Editor: Qingbo He

Received: 30 October 2023
Revised: 12 November 2023
Accepted: 16 November 2023
Published: 17 November 2023

Copyright: © 2023 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

The impact of machine learning (ML) and deep learning (DL) approaches is growing in the domain of rotating machinery diagnosis [1–4]. The increasing interest in these methodologies can be attributed to the potential of these algorithms to acquire knowledge in a highly automated manner, enabling them to establish meaningful connections between input variables, frequently derived from condition monitoring signals [5], and diagnostic outputs. This correlation is achieved with a remarkable level of accuracy, making these techniques valuable for predictive maintenance objectives. In this context, the involvement and expertise of operators in maintenance decision making are diminished, since they are increasingly dependent on the knowledge acquired by intelligent algorithms through the analysis of training data. This procedure is frequently associated with the examination of rolling element bearings (REBs) [6–10], as they serve as mechanical components that can gather informative data pertaining to the entire machinery and its overall performance [11–14].

Intelligent Fault Diagnosis (IFD) [1] has emerged as a prominent field in the past decade, focusing on the utilization of Artificial Intelligence (AI) for diagnosing machine components. Numerous approaches have been employed in this context. In the field of ML, the extraction of diagnostic features is typically performed manually, such as in the application of algorithms like support vector machine (SVM) [15–17] or k-nearest

neighbors (kNN) [18–20]. Then, DL is distinguished by the adoption of an automated approach for extracting features [21,22], which might manifest through the application of deep networks like convolutional neural networks (CNNs) [23–28]. The latter frequently engage with multi-dimensional data such as images. In the scenario of diagnosing rotating machinery, these images are commonly depicted as time-frequency images, such as spectrograms or Continuous Wavelet Transform (CWT) representations [27,29]. However, the efficacy of ML and DL approaches relies on the availability of a sufficient amount of training data [3,30,31] that faithfully capture the potential operating conditions and machine failures.

The availability of data is often limited in industrial applications. This limitation is particularly relevant when attempting to capture a comprehensive and representative distribution of potential damages. The currently accessible public datasets available in the field of bearing fault diagnosis [32–37] have been derived from extensive experimental efforts. However, they do not provide the same magnitude of data as certain datasets found in other domains, such as image recognition [38], sound detection [39] and natural language processing [40,41] for which foundation models are available [42]. In those domains, AI has demonstrated significant potential, and the presence of labeled data is certainly a contributing factor to this outcome. Due to this rationale, a significant portion of the scholarly literature concerning IFD focuses on the implementation of approaches with the objective of diminishing the substantial amount of training data.

Transfer Learning (TL) [3,43–45] approaches are employed to transfer previously gained knowledge from ML models across different tasks and domains representing different diagnostic scenarios. This approach effectively reduces the requirement for extensive training data. In general, TL often entails the process of building a model on a large dataset and subsequently employing it as a pre-trained basis for addressing similar or related problems. This technique offers the advantage of requiring less data and training time compared to training models from scratch. In the literature concerning machine fault diagnosis, there have been reports of instances where diagnostic capabilities have been successfully transferred to new working conditions of the same machine [46,47], as well as to different machines [48]. Also, several techniques are available for the re-utilization of models pre-trained for image recognition [29] and sound recognition [49,50] in the field of rotating machinery diagnostics.

Generative Adversarial Networks (GANs) [51–53] could be intended in a broad sense as TL approaches. In the domain of IFD, GANs are employed to construct generative models capable of producing synthetic data resembling the patterns observed in operating machinery [54–61]. Artificially produced data serve the purpose of enhancing the training dataset for IFD models for improving the performances and the feature extraction capabilities of the latter. The concept of GAN leverages the underlying mechanism employed by stacked autoencoders and variational autoencoders, which have been employed in the past to produce high-level features for training machine learning classifiers [62] and to diagnose rotor-bearing systems [63,64]. In recent years, a growing body of literature has drawn attention to the potential of GANs in the field of rotating machinery fault diagnosis. The significance of these models is highlighted by their capacity to enhance the quality and amount of data, as well as to address the issue of data imbalance for rotating machinery fault diagnosis. Moreover, the existing research emphasizes challenges in this field. For instance, one notable challenge is model generalization. This refers to the issue where a model trained on a specific set of machinery may not exhibit satisfactory performance when applied to another set, owing to variations in operating conditions, machine types, and fault characteristics. Additionally, the utilization of generative models can be intricate and demand substantial computational resources for effective training. Shao et al. [65] synthetized one-dimensional signals for augmenting the training set of diagnosis models, Liu et al. [56] proposed latent optimized stable GAN (LOSGAN) for data augmentation, Liang et al. [66] combined wavelet transform, GANs and CNNs for varying working conditions, Zhao and Yuan [67] boosted the training process by using improved GAN for

imbalanced datasets, whereas Cao et al. [59] employed GANs by transforming time signals into 2D images. Wasserstein GANs were investigated by Pu et al. [68] and Zhao et al. [58], while Liu et al. [69] included self-attention modules. Cycle-consistent adversarial learning and cycleGANs [70] were investigated by Jiao et al. [71], while Luleci et al. [72] employed cycleGANs to translate an undamaged domain to a damaged domain for structural health monitoring. CycleGANs for bearing fault diagnosis were also investigated by Xie and Zhang [73]. The authors developed a cycleGAN model for converting data from healthy operating conditions to inner race damaged data for the well-known CWRU dataset [32]. In this instance, all the operating conditions are seen by the generative model, although only for undamaged data.

At its core, existing research on GANs for rotating machinery frequently necessitates the presence of data from all the operating conditions to train generative algorithms that effectively augment or balance the training datasets. Furthermore, due to the extensive utilization of the CWRU experiments, distinct working conditions are frequently interpreted as different loads since speeds slightly vary in the CWRU signals. However, it has been recently noted by Hendriks et al. [33] that the absence of wide-ranging RPMs could result in the lack of a meaningful change in working conditions.

The primary objective of this study is to produce synthetic data that could replace the experimental damaged data required for training IFD algorithms in the field of industrial bearing diagnosis. Namely, this study suggests the generation of synthetic data also for operating conditions that have not been explicitly trained in the generative model. To achieve the intended objective, a GAN with cycle consistency loss function (cycleGAN) [70] is implemented and trained to transform CWT images of simulated vibrational signals into real signals, subsequently enabling the generation of synthetic data on the base of simulations. The simulated signals are produced by means of the model initially proposed by McFadden and Smith [74–76]. Recent studies by Sobie et al. [77] suggest that despite the fact that the use of advanced modeling approaches could provide additional understanding on the interaction between components of the bearing system, McFadden and Smith's model still accurately depicts the system's general behavior. The synthetic data are employed for the purpose of training diagnosis models by means of TL. The experimental activity is related to the test rig for medium-sized industrial bearings available at Politecnico di Torino [78,79]. It is shown that synthetic data can replace real experimental damaged data for training IFD models.

The novel contributions introduced in this study are reported in the following sections. Firstly, it enables the generation of data pertaining to machine working conditions that were previously unexplored in experimental settings. In this sense, the model stands as a generative zero-shot learning approach for machine fault diagnosis. Secondly, it introduces the analysis of the potential of GANs for the fault diagnosis of medium-sized industrial systems, which differ from the commonly encountered laboratory bearings. Lastly, it investigates the capabilities of generating data as machine speed varies rather than focusing solely on loads. This implies the capability to generate frequency distributions that significantly deviate from those encountered during the training of the generative model. In the context of zero-shot generative AI for rotating machinery, there is no work to date that modulates the rotating speed to produce operating conditions never seen by GANs.

2. Generative AI and Machine Fault Diagnosis

The application of generative AI in diagnosing rotating systems predominantly involves the use of GANs on a large scale. The concept of GAN was first introduced in 2014 by Goodfellow et al. [52] within the field of generative AI. GANs are composed of two convolutional architectures engaged in a zero-sum game [80,81] as shown in Figure 1.

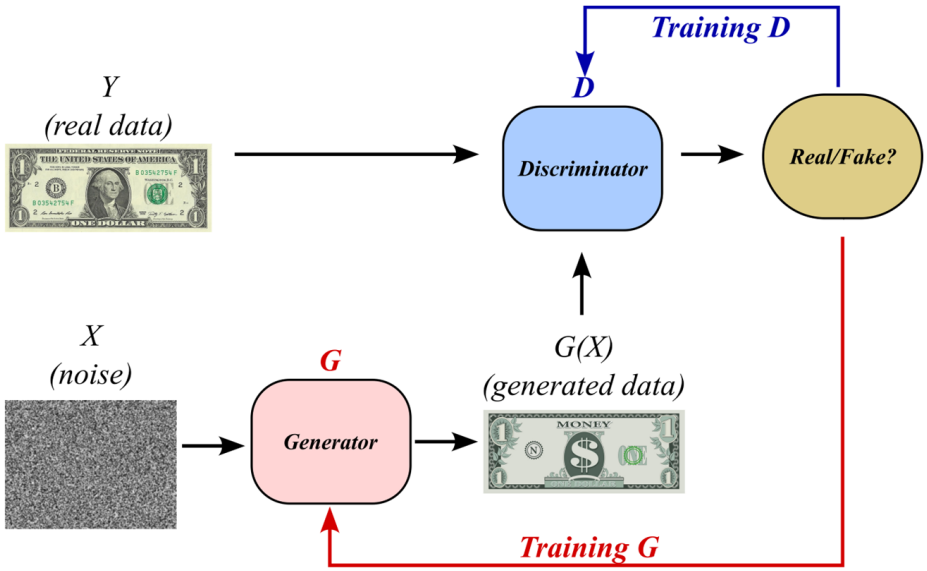


Figure 1. Generative Adversarial Network (GAN): general outline.

The objective of generator G is to generate images that closely resemble a specific distribution Y of images, using random input in the form of white noise X. Discriminator D consists of a CNN that functions as a binary classifier, discerning between authentic images and synthetic ones generated by the generator by transformation $X \rightarrow G(X)$. The mathematical representation of the competitive interaction between the two convolutional structures can be described by objective function $\mathcal{L}_{GAN}$, as shown in Equation (1). The expected value operator $\mathbb{E}[\cdot]$ is applied to generic y, which belongs to domain Y and follows probability distribution p_y. Similarly, generic x belongs to domain X and follows probability distribution p_x. The output of the generator is denoted as $G(\cdot)$, while the output of the discriminator is denoted as $D_Y(\cdot)$. Namely, this latter represents the probability of the input image being real. From the perspective of the discriminator, it is advantageous to maximize the $\mathcal{L}_{GAN}$ function, whereas for the generator, it is suitable to decrease it. The existence of such a dichotomy leads to the establishment of a zero-sum game. Then, the training of GAN involves the identification of a Nash equilibrium [80,81] between the generator and discriminator components, as represented by Equation (2).

$$\mathcal{L}_{GAN} = \mathbb{E}_{y \sim p_y}[\log D_Y(y)] + \mathbb{E}_{x \sim p_x}[\log(1 - D_Y(G(x)))], \tag{1}$$

$$\min_{G} \max_{D} \mathcal{L}_{GAN}. \tag{2}$$

A particular type of generative networks is represented by cycleGANs, which were originally introduced in 2017 by Zhu et al. [70]. The fundamental architecture of cycleGANs is shown in Figure 2 and consists of two generators, denoted as G and F, as well as two discriminators, referred to as D_Y and D_X. The two GANs are mathematically interconnected through cycle consistency loss function $\mathcal{L}_{cyc}$ defined in Equation (3). The global loss function is reported in Equation (4), where λ is the weight assigned to $\mathcal{L}_{cyc}$.

$$\mathcal{L}_{cyc} = \mathbb{E}_{x \sim p_x}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_y}[\|G(F(y)) - y\|_1], \tag{3}$$

$$\mathcal{L}_{tot} = \mathcal{L}_{GAN(x \rightarrow y)} + \mathcal{L}_{GAN(y \rightarrow x)} + \lambda \mathcal{L}_{cyc}. \tag{4}$$

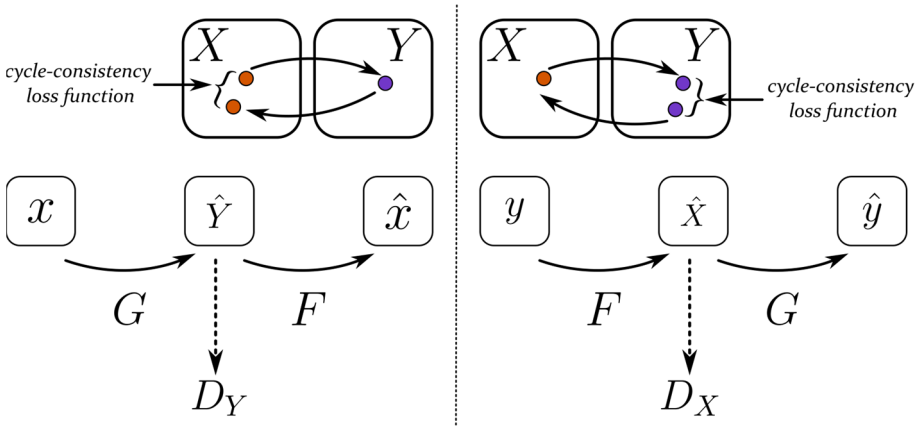


Figure 2. CycleGAN working diagram and role of the cycle-consistency loss function in the coupling of two generators and two discriminators.

Function $\mathcal{L}_{cyc}$ enables the establishment of a bidirectional correspondence between two image domains X and Y, even in cases where training data pairings $\{x_i, y_i\}$ lack an exact equivalent in the coupled domain (i.e., unpaired image-to-image translation). The $\mathcal{L}_{cyc}$ function in Equation (3) imposes a constraint on the model built on two GANs, ensuring that it produces an image that is virtually identical to input image x after undergoing transformation $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. This method obviates the need for a perfect coupling between training images x_i and y_i. For instance, in the original paper, the methodology is employed to convert the style of images, thereby transforming them from paintings to photos. This transformation could be achieved without prior training by providing both the painting and its associated photo as input. This is because the model is cycle consistent and it learns the specific mapping from the x to the y domain for which $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ rather than a random one.

The methodology adopted in this study employs cycleGAN networks to generate synthetic data by transforming CWT transform images of simulated vibrational signals into their corresponding experimentally observable counterparts. Therefore, cycleGAN translates the output of a simulation model into its corresponding real-world representation, closing the gap between them. This study also postulates that the cyclic consistency property can be leveraged to generate data for rotating machinery operating situations that have never been explicitly trained in the cycleGAN conversion model.

3. Generating Synthetic Bearing Fault Data via cycleGAN

The methodology described in this section involves the implementation of an experimental campaign with the objective of extracting vibrational signals from medium-sized bearings under both normal and fault states. Then, analytical simulations are conducted to reproduce vibrational behavior. The signals derived from each of these sources undergo pre-processing for the purpose of training image-based generative AI systems. The latter aims to transform the outcomes derived from a simulation model into data that closely resemble those seen in a real experimental scenario. The generative model trained in this manner is able to provide data that can serve as alternatives to data obtained from the actual experiment, hence becoming valuable for the training of diagnostic models.

3.1. Experimental Activity

The experimental study relates to the test rig designed for medium-sized industrial bearings; the test rig is located in the laboratories of the Politecnico di Torino [78]. The experimental campaign associated with this test rig represents a pioneering effort to investigate vibration data for large industrial bearings with localized damages [50]. The test rig shown in Figure 3a can accommodate a maximum of four bearings, each having an outer diameter ranging from 280 mm to 420 mm. The main shaft is driven by a three-phase

motor with a power output of 30 kW. A PERIFLEX® (Unna, Germany) elastic coupling connects the shaft to the electric motor. Hydraulic actuators are employed to apply radial and axial loads independently, with a maximum magnitude of 200 kN, by means of adapters (Figure 3b) that enclose the bearings being tested. The hydraulic actuators are fed by air–oil conversion pumps that are coupled to the pneumatic system present within the laboratory. The bearings being examined are lubricated by an external recirculation system, which involves the injection of ISO VG 150 oil at a flow rate of 2.5 L/min under a pressure of 6 bar. The architectural design of the test rig is characterized by the so-called "self-contained box" enclosure. The present architecture ensures that loads are balanced by the elastic deformation of the box, thereby dissipating the load circuit within the box. One notable benefit of such a design is in its capability to effectively manage substantial loads without necessitating the use of oversized main bearings supporting the whole system. Each adapter contains a SKF CMS 2200T sensor (Gothenburg, Sweden) for measuring acceleration and temperature. The sensors are coupled to a Scadas III LMS acquisition system. Further details regarding the experimental setup are provided in references [50,78].

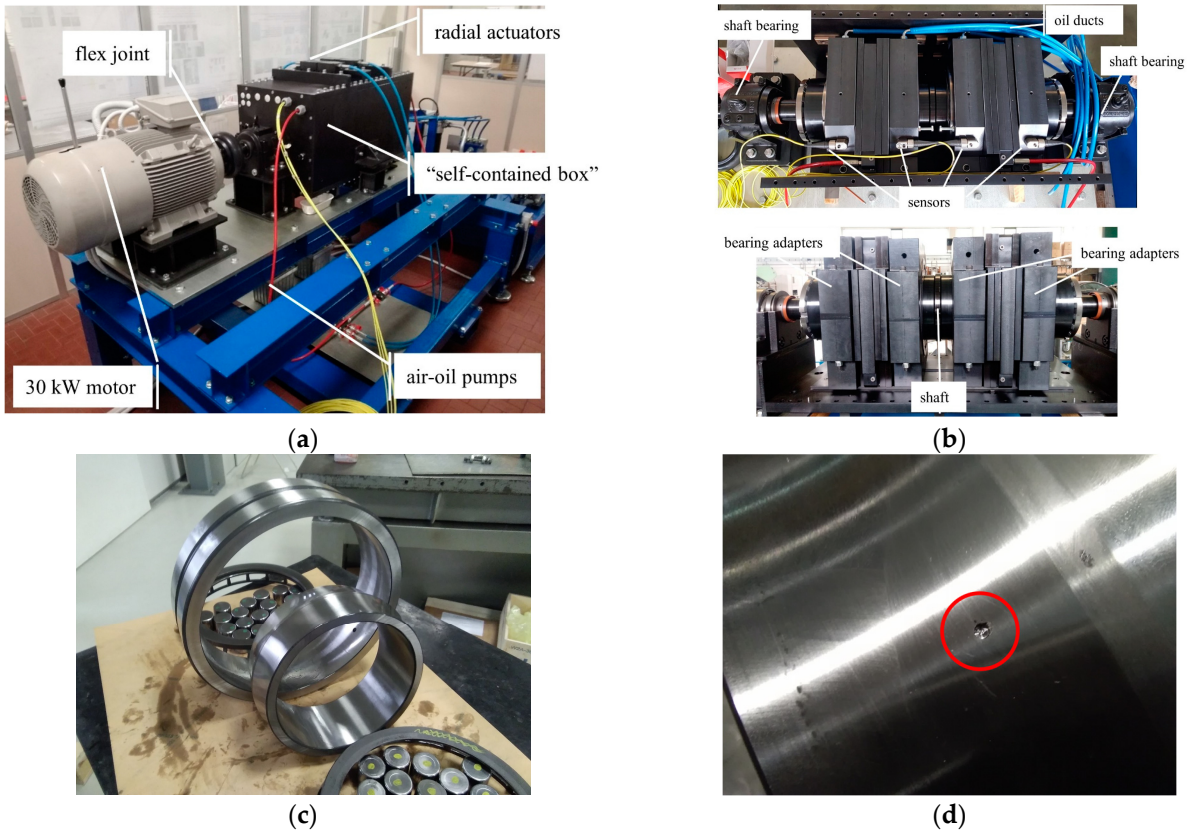


Figure 3. Experimental activity: (**a**) test rig for industrial bearings [78]; (**b**) interior of the "self-contained box" [78]; (**c**) SKF 22240 CCK/W33 bearing disassembly; (**d**) inner race damage (IR) [50,79].

The experimental activity involved the analysis of vibration samples extracted from the test rig being investigated. Namely, the spherical roller bearing SKF 22240 CCK/W33 underwent testing. The bearing has an internal diameter of 200 mm, with a taper ratio of 1:12, and an external diameter of 360 mm. The disassembled bearing is shown in Figure 3c. The bearing was tested under both normal operating conditions and a damaged condition specifically focused on the inner race (IR). The localized defect was induced by machining, and it has a diameter of 2 mm with a depth of 0.5 mm, as shown in Figure 3d. The test settings and parameters of accelerometer signal extraction are presented in Table 1. The generative model was trained using exclusively data related to the nominal speed of 877 rpm. By converse, the data obtained from the working conditions at 607 rpm

and 997 rpm were excluded from the training process of the cycleGAN model. Thus, the generative model was unaware of the existence of images representing frequency distributions at speeds other than the nominal 877 rpm.

Table 1. Test conditions and signal extraction for the bearing SKF 22240 CCK/W33.

Radial load (kN)	124.8
Speed (rpm)	607, 877, 997
Duration (s)	30
Sampling frequency f_s (Hz)	20,480
Overlap	0.7
Chunk duration (s)	0.8
Chunks per signal	250

3.2. Simulating Bearing Vibration Signal

The vibrational behavior of rolling bearings with localized damages was simulated using the analytical model originally proposed by McFadden and Smith [74–76]. The model has gained significant recognition in scholarly works, despite being among the initial attempts to simulate the behavior of faulty bearings. According to recent research [77], incorporating more sophisticated modeling techniques such as three-dimensional finite element analysis and considering contact mechanics phenomena offers supplementary insights into the interactions among the constituents of the bearing system. Nevertheless, Sobie et al. [77] emphasize that that McFadden and Smith's model continues to effectively capture the overall behavior of the system.

Also, the simple model of McFadden and Smith was utilized in this study because it leaves out some details about the mechanical system being examined. For instance, the model does not consider the inertial interactions that occur between the rolling elements, lubricant, and races, nor does it account for the elasto-hydrodynamic behavior of the lubricant. In the context of this research, these limitations can be viewed as advantageous as they enable the generative AI model based on the cycleGAN architecture to assess its capacity to integrate all the absent information into the simulated model, particularly in the scenario where the complexity of the simulated model is significantly reduced. Consequently, if the generative model can generate synthetic data using minimal initial information, it is reasonable to assume that it can effectively evolve to more intricate simulation models with a higher level of detail.

The model is shown in Equations (5)–(8), whereby the impulse response $h(t)$ is derived from decay parameter $\beta = 500$ Hz and structural resonance $f_{struct} = 1700$ Hz. Function $d(t)$ represents the Dirac comb function. Each pulse of the Dirac comb is separated by a distance equal to the reciprocal of defect characteristic frequency f_{defect}. According to manufacturer SKF® the IR characteristic frequency related to the bearing under analysis is $f_{defect} = 10.824 \cdot f_r$, where f_r represents the rotation frequency of the machine. The load operating on the rolling element at the angular coordinate ψ, denoted as Q_ψ, is determined using the well-known formulation by Harris [82]. The load distribution factor is represented by ϵ. The convolution operator is denoted by symbol $*$, and the characteristic frequency of the asynchronous motor, f_{motor}, is defined as $f_{motor} = 6f_{stator}$, where f_{stator} represents the frequency of the stator power supply. Parameters β, f_{struct}, A_1 and A_2 were selected based on first attempt values. It was therefore intended for the generative algorithm to adjust the influence of these parameters by acquiring knowledge on how to modify the simulated data to align them with experimental observations.

$$h(t) = e^{-\beta t}\sin(2\pi f_{struct}t), \tag{5}$$

$$d(t) = \sum_k \delta(t - kT), \quad T = 1/f_{defect}, \tag{6}$$

$$Q_\psi = Q_{max}\left[1 - \frac{1-\cos(\psi_i)}{2\epsilon}\right]^{\frac{10}{9}}, -\psi_m < \psi < \psi_m, \tag{7}$$

$$s(t) = \left[Q_\psi d(t)\right] * h(t) + A_1 \sin(2\pi f_r t) + A_2 \sin(2\pi f_{motor} t). \tag{8}$$

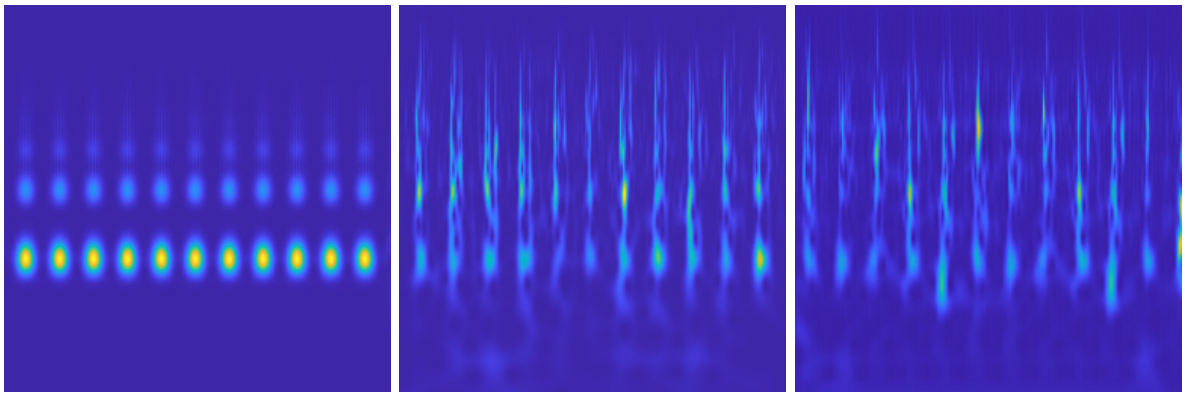
3.3. Data Pre-Processing and CycleGAN Training

The simulated and experimental signals were pre-processed to emphasize the existence of bearing damages within the harmonic content. In particular, the 250 sample pairs obtained from the simulated and experimental signals (Table 1) were subjected to normalization based on the mean value and variance of each sample. The implementation of the normalizing procedure aims to ensure that both simulated and actual signals can be effectively characterized using comparable scales. Subsequently, the signals were filtered within the frequency range of 1400 Hz to 2800 Hz; then, envelope [10,12,13] $env(t)$ was extracted. The CWT transform shown in Equation (9) was subsequently utilized. In this equation, operator $\psi^*(\cdot)$ denotes the complex conjugate of the Morse wavelet, while a and b are the scaling and translation factors of the wavelet transform, respectively. The selection of the Morse wavelet was based on its widespread application in signal analysis for this particular type of signals, owing to its advantageous properties of temporal and frequency localization. These characteristics render it well-suited for the purpose of identifying sudden pulses due to bearing defects. The Morse wavelet was employed with a symmetry parameter $\gamma = 3$ and a time-bandwidth product $P^2 = 60$; in addition, the CWT had 24 voices per octave. The CWT images are produced by operation $W^2(a;b)$, which enhances the trace of the rolling element's transition from the defect inside the CWT spectrum.

$$W(a;b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} env(t)\psi^*\left(\frac{t-b}{a}\right) dt. \tag{9}$$

Figure 4a illustrates an instance of the CWT spectrum associated with a simulated signal operating at a rotational speed of 877 rpm. Conversely, Figure 4b shows the corresponding experimental signal. The cycleGAN model developed in a Matlab® environment was trained using a dataset consisting of 250 CWT image pairs. The training lasted 9 h and 26 min, utilizing an NVIDIA® T4 GPU that was accessible on the High-Performance Computing (HPC) infrastructure provided by the commercial cloud environment Amazon® AWS. The training hyperparameters are reported in Table 2. Given the evident computational cost complexities involved in implementing hyperparameter optimization techniques on such large and complex architectures, the authors use as a starting point the hyperparameters shown in Table 2, which are easily found in the literature inherent to cycleGANs [70]. Appendix A provides the description of the architectures employed in cycleGAN, specifically pertaining to the generators and discriminators. Tables A1 and A2 present detailed information regarding these frameworks. After the completion of the training process, the model was utilized to generate images based on simulated data. The signal generated under the condition of 877 rpm is illustrated in Figure 4c. Striking resemblances to the real signal are observed (Figure 4b). The process of generating the data can be summarized as follows:

- simulating the accelerometer signal;
- pre-processing the signal through normalization, filtering and envelope extraction;
- applying the continuous wavelet transform and generating 256 × 256 images by squaring the CWT coefficients;
- utilizing the images as input for the cycleGAN model previously trained to transform images of simulated signals into their corresponding real counterparts. The resulting output of the cycleGAN model is an image that is a surrogate for a real image from experimental activity.



(**a**) (**b**) (**c**)

Figure 4. Images of CWT spectra @ 877 rpm: (**a**) simulated signal; (**b**) experimental signal; (**c**) image generated by cycleGAN.

Table 2. cycleGAN training hyperparameters.

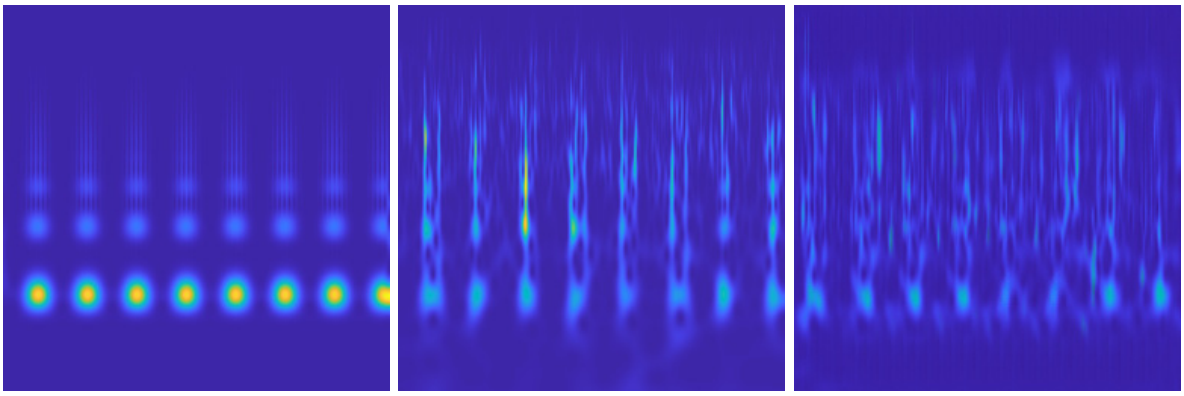
Epochs	200
Mini-batch size	1
Optimizer	Adam
Learning rate	0.0002
Gradient decay factor	0.5
Squared gradient decay factor	0.999
Adversarial loss weight (λ)	10

Figure 4 demonstrates the capabilities of the generative model to produce synthetic data that retrace the operational conditions on which the algorithm was trained. The acquisition of this ability can be of great value in augmenting the existing data on operational conditions if damage data are already accessible, hence enabling the adjustment of any class imbalance within the datasets. Indeed, fault data are much rarer and more difficult to find in practical industrial settings. However, certain operating conditions, namely those pertaining to various rotational speeds, may not be available from experimental activities. Hence, it is desirable for the generative model to have the capability to produce also some of these operating conditions. Such capabilities belong to a zero-shot generative learning framework.

Taking into consideration the aforementioned scenario, the cycleGAN model was utilized for the purpose of generating the continuous wavelet transform (CWT) data corresponding to the rotation speeds of 607 rpm and 997 rpm. The decision to manipulate the operating conditions by altering the rotational speeds instead of loads was undertaken to test the ability of the generative AI algorithm to produce frequency content that had not been experienced before. On the other hand, it should be noted that altering the load while maintaining a constant rotational speed of 877 rpm would not have significant impact on the characteristic frequencies associated with the fault. These frequencies are primarily influenced by kinematic parameters and remain unchanged regardless of load variations. The rates of 607 rpm and 997 rpm were chosen in close proximity to the training speed of 877 rpm to ensure that the demodulation band ranging from 1400 Hz to 2800 Hz remained pertinent in detecting damage within the demodulated signal. In these instances, it is evident from Figures 5 and 6 that there were notable resemblances between the real experimental signals and the signals generated by the cycleGAN model.

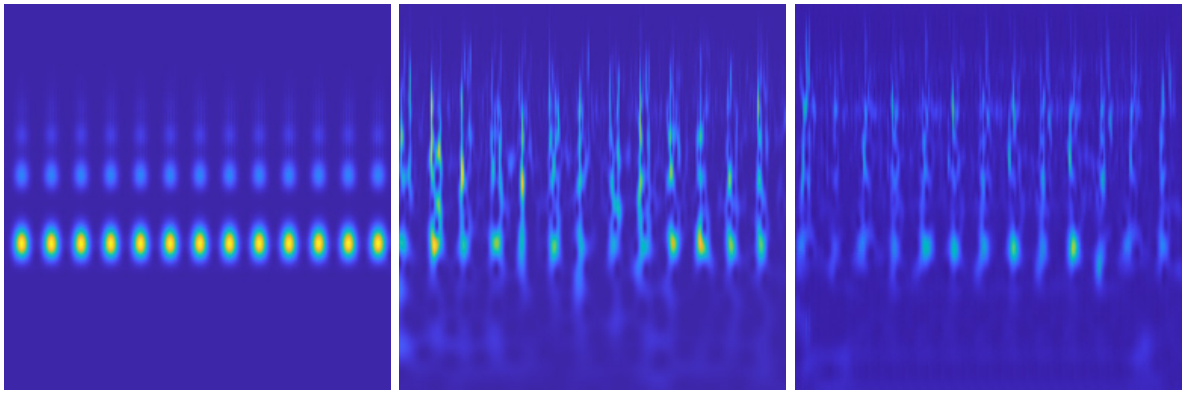
3.4. Validation of the CycleGAN Generative Model

The validation of the synthetic data generation methodology presented in this study involved the utilization of the MMD metric [83], the Fréchet Inception Distance (FID) [84,85] and a specifically designed TL methodology.



(**a**) (**b**) (**c**)

Figure 5. Images of CWT spectra @ 607 rpm: (**a**) simulated signal; (**b**) experimental signal; (**c**) image generated by cycleGAN.



(**a**) (**b**) (**c**)

Figure 6. Images of CWT spectra @ 997 rpm: (**a**) simulated signal; (**b**) experimental signal; (**c**) image generated by cycleGAN.

The MMD is a non-parametric statistical metric used to quantify the dissimilarity of two probability distributions. The empirical estimate of the MMD is calculated using Equation (10), where $\phi(\cdot)$ represents the characteristic kernel of the nonlinear mapping function, and $X = \{x_i\}_{i=1}^{N}$ and $Y = \{y_j\}_{j=1}^{M}$ are two given datasets. In this study, a radial basis function (RBF) kernel [86] is employed. Similarly, the FID metric is widely employed in GANs to measure how distant the images produced by generative AI are from reality. Specifically, FID employs the components of the latent space formed by the Inception Net-V3 [87] for the purpose of computing the statistic specified in Equation (11), where μ_r is the mean value of the real image, μ_g is the mean value of the generated image, Σ_r is the covariance matrix of the real image, Σ_g is the covariance matrix of the generated image and Tr represents the trace. Within the framework of this research, the MMD and FID metrics were employed to assess the discrepancy between probability distributions pertaining to simulated data, real data, and synthetic data. This enables the measurement of the degree of similarity that arises in a qualitative manner (as shown in Figures 4–6) between the real and synthetic data. Consequently, decreased MMD and FID values signify a higher degree of similarity between the distributions, thereby implying that the synthetic data could potentially serve as valid alternatives to the experimental data.

Table 3 presents the computed values of the MMD and FID for different machine conditions. The data generated by cycleGAN exhibit a higher degree of resemblance to the authentic data in comparison to the simulated data. This statement is true for both

the explicit training condition of the generative algorithm (877 rpm) and the two distinct operational conditions (607 rpm and 997 rpm).

$$MMD^2(X, Y) = \left\| \frac{1}{N}\sum_{i=1}^{N}\phi(x_i) - \frac{1}{M}\sum_{j=1}^{M}\phi(y_i) \right\|^2, \quad (10)$$

$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}}). \quad (11)$$

Table 3. MMD and FID measuring the distance from real machine data for different operating conditions.

Speeds (rpm)	Data Distributions	MMD	FID
607	Real—Simulated	0.187	265.717
	Real—Generated	0.106	78.270
877	Real—Simulated	0.218	317.714
	Real—Generated	0.069	47.888
997	Real—Simulated	0.202	278.152
	Real—Generated	0.073	48.607

In order to evaluate the effectiveness of synthetic data as substitutes for real data, a series of training experiments were performed on three distinct diagnosis models based on the TL approach [49,50]. Namely, the CNNs AlexNet [88], VGG16 [89], and ResNet18 [90], which were initially pre-trained for image recognition on ImageNet [38], were trained through the fine-tuning process. The initial and intermediate layers, which are responsible for extracting image features that can differentiate between different classes, were kept the same. However, the final layer was replaced, and its weights were adjusted through retraining to the specific diagnostic task. This methodology enables the utilization of the knowledge existing in models trained on image recognition, particularly on a vast dataset like ImageNet [38]. Then, a general comprehensive knowledge can be applied into a specific domain, such as machinery diagnosis, through the implementation of time-frequency image recognition techniques.

Each model was tested exclusively with real data from the test rig. However, the models were trained using three distinct configurations: in the first scenario, real damage data were utilized for training; in the second scenario, simulated damage data were employed for training; and lastly, in the final scenario, the training fault data were generated through the cycleGAN generative model. The diagnosis models were trained using a total of 900 training samples. These samples were divided into 150 samples for each speed and two health conditions (i.e., healthy and IR), amounting to 60% of the dataset. Also, 300 validation samples were used, with 50 samples for each speed and two health conditions, representing 20% of the total dataset. Finally, 300 test samples were employed, consisting of 50 samples for each speed and two health conditions. The fine-tuning hyperparameters employed for the fault diagnosis models are reported in Table 4.

Table 4. Fine-tuning hyperparameters for CNNs-based fault diagnosis.

Epochs	4
Mini-batch size	32
Optimizer	Momentum
Initial learning rate	10^{-4}
L2 regularization	10^{-4}
Momentum factor	0.9

4. Results and Discussion

The analyses were performed accounting the metrics of accuracy and recall for diagnostic testing. Accuracy is reported in Equation (12), whereas Equation (13) formalizes the

concept of recall. In the equations, the variable TP represents the number of true positives, which identifies the accurate identification of damages. Variable TN represents the number of true negatives, which refers to accurately identified healthy data. FP represents the number of false positives, which refers to misclassified normal data as damaged. FN represents the count of false negatives, which indicates misclassified damaged data as normal. Hence, the recall rate serves as a metric to quantify the proportion of the fault samples that are correctly classified as such by the diagnosis model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{12}$$

$$Recall = \frac{TP}{TP + FN}. \tag{13}$$

The test results for the three diagnostic models were produced using real experimental data and are presented in Figure 7. The figure presents the performance results in the case of training using real, simulated, and synthetic damaged data. The performance was measured in terms of mean and standard deviation, which were derived from five separate tests. It is evident that all models exhibit high accuracy when trained using real damaged data, but they fail in identifying damage (as shown in Figure 7b) when trained using simulated data. One notable finding is that two of the utilized models, specifically AlexNet and VGG16, reveal noteworthy diagnosis metrics when trained with synthetic data. The metrics derived from the use of synthetic IR data for training purposes exhibit a level of similarity to the high metrics produced from the utilization of real data. This implies that synthetic data have the potential to effectively replace actual data in the training of diagnostic algorithms. The training times are reported in Figure 7c. The AlexNet model is noted for its superior performance, whereas there are no notable distinctions identified among the real, simulated, and synthetic datasets in terms of training times.

The findings indicate that the cycleGAN generative model successfully produces data that closely resemble the real data obtained from the test rig. In the specific scenario under investigation, this assertion was corroborated for a particular operational state (877 rpm) that underwent explicit training in the generative model. The aforementioned outcome can be observed in diverse manifestations within the literature. It is frequently employed to enhance the quality of training data by rectifying imbalances in datasets that may be, for instance, disadvantaged with respect to fault data. The findings reported in this study demonstrate that by the implementation of a suitable pre-processing approach and the integration of a cycleGAN generative model, it becomes feasible to generate data for operational scenarios that may not have been included in the experimental procedures, setting up as a zero-shot generative learning approach. Namely, it is observed that there are two distinct rotation speeds (607 rpm and 997 rpm) where data could be effectively generated. These conditions indicate a notable alteration in the frequency distribution of the accelerometer data when compared to the initial condition of 877 rpm. This capability is characteristic of generative zero-shot learning and can be attributed to several factors. The utilization of an envelope-based pre-processing technique enables the identification and emphasis of fault features within spectra. The fault features exhibit a somewhat consistent shape across the various rotation speeds that are being studied. Hence, if the generative algorithm acquires the ability to manipulate those fault characteristics, it is conceivable that it can do so for varying rotation speeds as well, provided that the fault features in question have resemblance to those it has already acquired. Hence, it is postulated that the efficacy of this approach is constrained by its capacity to generate CWT spectra exhibiting fault features of comparable morphology when the operational parameters of the machinery undergo changes. Nevertheless, it is important to emphasize that the cycle consistency function plays a crucial role in facilitating a direct mapping between simulated and real signals. The $\mathcal{L}_{cyc}$ function was specifically designed to facilitate the training of image translation models in the presence of unpaired images within the training

dataset. In this scenario, it is observed that training images are paired. However, it is important to note that the algorithm does not explicitly learn the process of pairing images. Instead, the $\mathcal{L}_{cyc}$ functions make the model focus on mapping an image to a different domain and subsequently reconstruct the original image, as presented in Figure 2 and Equation (3). This approach has the potential to significantly enhance the capacity to generate data under diverse operating conditions, exhibiting a generalized knowledge that can be extrapolated to novel and unfamiliar scenarios. Nevertheless, it is crucial to highlight that while validating the methods on speeds other than the training speed, it is important to consider the potential issues that may arise from moving away from the established 877 rpm, as it is uncertain how such deviations could impact the system's capabilities. The presence of uncertainty can be attributed to potential limitations within the cycleGAN model, as well as the choice of a demodulation band that may no longer adequately emphasize the transition from defects in CWT spectra.

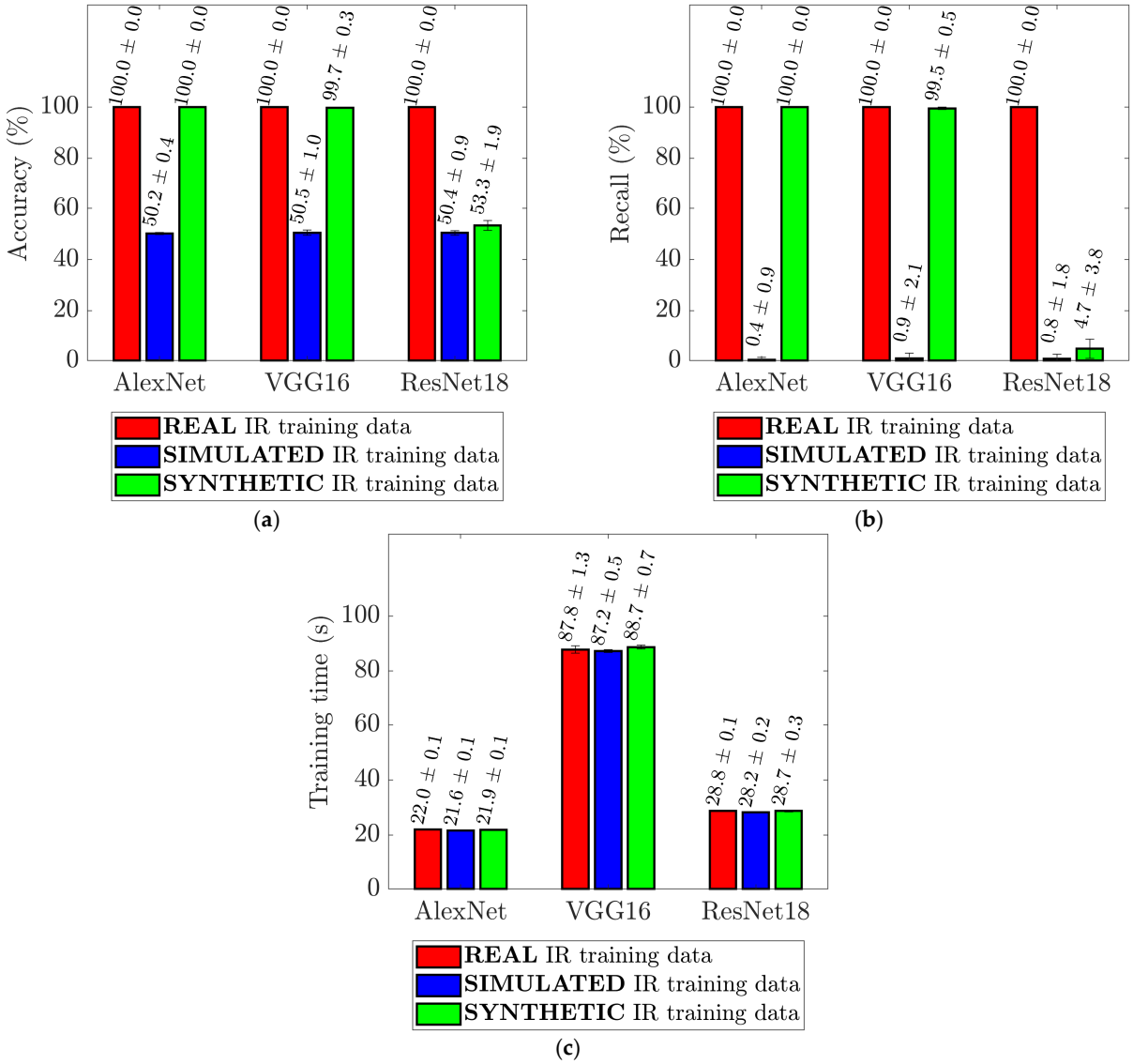


Figure 7. Results for real experimental test data for different diagnosis models trained on real, simulated and synthetic IR data: (**a**) accuracy; (**b**) recall; (**c**) training times.

In conclusion, it is important to acknowledge that the diagnostic performances achieved by ResNet18 are poor when compared to those achieved by AlexNet and VGG16. This element is of significant interest and merits additional examination. The authors

highlight that Resnet18 distinguishes itself from AlexNet and VGG16 by incorporating residual connections. These connections are also found in the cycleGAN architecture and may potentially contribute to Resnet18′s ability to discern differences in synthetic data with respect to real ones. However, it is important to note that the aforementioned speculation is currently conjectural in nature and warrants additional research.

5. Conclusions

The objective of this work was to provide a generative AI-based methodology for synthesizing data that can serve as substitute for real-world data on damage in industrial rotating machinery. These synthetic data are used to train diagnostic algorithms based on deep learning and transfer learning. In order to achieve this objective, a cycleGAN generative neural network was employed to transform wavelet images associated with simulated signals into their experimental counterparts. The AI-based generative process was further validated by employing the MMD and FID metric. The examination of these indicators makes it possible to quantitatively determine the effectiveness of the generative network on the entire set of images investigated. Subsequently, three CNNs-based diagnosis models were trained utilizing the transfer learning technique, employing real, simulated, or synthetic data pertaining to damaged machinery. The diagnosis models were tested exclusively on authentic data acquired from a dedicated experiment carried out on a test platform featuring industrial bearings of medium size. Based on the evidence and analysis presented, it can be inferred that:

- cycleGANs were found to be very efficient architectures for producing synthetic data in the form of CWT images, which may be utilized to augment or substitute real bearing fault data;
- the synthetic data generated by cycleGANs exhibiedt a higher degree of resemblance to real data when compared to the outcomes obtained from a simulation model. This assertion holds true for both the specific machine operating conditions that the generative algorithm was explicitly trained on, as well as for different conditions;
- the proposed methodology was able to generate data for working conditions that could be entirely lacking in the experimental activity. It is claimed that generative zero-shot capabilities arise also as a result of the cycle consistency inherent in the generative model;
- the synthetic data generated by the model were found to be effective for training diagnostic algorithms, which exhibited very high accuracy when evaluated using real data collected from the rotating machinery.

Potential future advancements encompass the investigation of diverse radial and axial loading conditions, alongside the examination of distinct forms of bearing defect. Speeds more distant from the training speed will also be tested in the future. Furthermore, the objective is to perform a comprehensive examination by employing increasingly complex simulation models. Further research is essential to gain a comprehensive understanding of the underlying factors contributing to the ineffectiveness of the ResNet18 model when trained using synthetic data. Also, future developments will require the adoption of generative approaches available in the literature that will be measured under the conditions stated in this work to deliver a genuine and legitimate comparison to the scientific community. Ultimately, there are intentions to explore the boundaries in which data can be generated for operating conditions for which the generative model has not been explicitly trained as well as testing on different types of bearing.

Author Contributions: Conceptualization, L.G.D.M.; methodology, L.G.D.M.; software, L.G.D.M.; resources, E.B. and C.D.; writing—original draft preparation, L.G.D.M.; writing—review and editing, E.B. and C.D.; supervision, E.B. and C.D.; project administration, E.B. and C.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix accounts for the architecture of the cycleGAN model. Table A1 shows the modules that comprise the generators of the two GANs within the cycleGAN framework, whereas Table A2 illustrates the modules that constitute the discriminators.

Table A1. Architecture of cycleGAN generators.

Structure	Block	Layer	Filter Size	Stride/ Padding	Channels
Image input layer 256 × 256 × 3					
Encoder	1	Convolution	4 × 4	2/1	64
		Batch norm	-	-	-
		ReLU	-	-	-
	2	Convolution	4 × 4	2/1	128
		Batch norm	-	-	-
		ReLU	-	-	-
	3	Convolution	4 × 4	2/1	256
		Batch norm	-	-	-
		ReLU	-	-	-
Residuals	1–6	Convolution	3 × 3	1/1	256
		Batch norm	-	-	-
		ReLU	-	-	-
Decoder	1–2	Transposed convolution	4 × 4	2/	128
		Barch norm	-	-	-
		ReLU	-	-	-
	3	Transposed convolution	4 × 4	2/	3
		Tanh	-	-	-

Table A2. Architecture of cycleGAN discriminators.

Layer	Filter Size	Stride/ Padding	Channels
Image input layer 256 × 256 × 3			
Convolution	4 × 4	2/1	80
ReLU	-	-	-
Convolution	4 × 4	2/1	160
Batch norm	-	-	-
ReLU	-	-	-
Convolution	4 × 4	2/1	320
Batch norm	-	-	-
ReLU	-	-	-
Convolution	1 × 1	1/0	1

References

1. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
2. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery Health Prognostics: A Systematic Review from Data Acquisition to RUL Prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [CrossRef]
3. Li, C.; Zhang, S.; Qin, Y.; Estupinan, E. A Systematic Review of Deep Transfer Learning for Machinery Fault Diagnosis. *Neurocomputing* **2020**, *407*, 121–135. [CrossRef]
4. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
5. Mohanty, A.R. *Machinery Condition Monitoring: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2014; ISBN 978-1-4665-9305-3.
6. Alabsi, M.; Liao, Y.; Nabulsi, A.A. Bearing Fault Diagnosis Using Deep Learning Techniques Coupled with Handcrafted Feature Extraction: A Comparative Study. *JVC J. Vib. Control.* **2021**, *27*, 404–414. [CrossRef]
7. Abbasion, S.; Rafsanjani, A.; Farshidianfar, A.; Irani, N. Rolling Element Bearings Multi-Fault Classification Based on the Wavelet Denoising and Support Vector Machine. *Mech. Syst. Signal Process.* **2007**, *21*, 2933–2945. [CrossRef]
8. Brusa, E.; Delprete, C.; Di Maggio, L.G. Eigen-Spectrograms: An Interpretable Feature Space for Bearing Fault Diagnosis Based on Artificial Intelligence and Image Processing. *Mech. Adv. Mater. Struct.* **2022**, *30*, 4639–4651. [CrossRef]
9. Delprete, C.; Brusa, E.; Rosso, C.; Bruzzone, F. Bearing Health Monitoring Based on the Orthogonal Empirical Mode Decomposition. *Shock. Vib.* **2020**, *2020*, 8761278. [CrossRef]
10. Brusa, E.; Bruzzone, F.; Delprete, C.; Di Maggio, L.G.; Rosso, C. Health Indicators Construction for Damage Level Assessment in Bearing Diagnostics: A Proposal of an Energetic Approach Based on Envelope Analysis. *Appl. Sci.* **2020**, *10*, 8131. [CrossRef]
11. Genta, G. *Dynamics of Rotating Systems*; Springer Science & Business Media: Berlin, Germany, 2007; ISBN 978-0-387-28687-7.
12. Randall, R.B. *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2011; ISBN 978-0-470-74785-8.
13. Randall, R.B.; Antoni, J. Rolling Element Bearing Diagnostics—A Tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [CrossRef]
14. Brusa, E. Design of a Kinematic Vibration Energy Harvester for a Smart Bearing with Piezoelectric/Magnetic Coupling. *Mech. Adv. Mater. Struct.* **2020**, *27*, 1322–1330. [CrossRef]
15. Baccarini, L.M.R.; Rocha e Silva, V.V.; de Menezes, B.R.; Caminhas, W.M. SVM Practical Industrial Application for Mechanical Faults Diagnostic. *Expert Syst. Appl.* **2011**, *38*, 6980–6984. [CrossRef]
16. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
17. Widodo, A.; Yang, B.-S. Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560–2574. [CrossRef]
18. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
19. Lei, Y.; Zuo, M.J. Gear Crack Level Identification Based on Weighted K Nearest Neighbor Classification Algorithm. *Mech. Syst. Signal Process.* **2009**, *23*, 1535–1547. [CrossRef]
20. Moosavian, A.; Ahmadi, H.; Tabatabaeefar, A.; Khazaee, M. Comparison of Two Classifiers; K-Nearest Neighbor and Artificial Neural Network, for Fault Diagnosis on a Main Engine Journal-Bearing. *Shock. Vib.* **2013**, *20*, 263–272. [CrossRef]
21. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep Learning and Its Applications to Machine Health Monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
22. Duan, L.; Xie, M.; Wang, J.; Bai, T. Deep Learning Enabled Intelligent Fault Diagnosis: Overview and Applications. *JIFS* **2018**, *35*, 5771–5784. [CrossRef]
23. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2018**, *65*, 5990–5998. [CrossRef]
24. Grezmak, J.; Zhang, J.; Wang, P.; Loparo, K.A.; Gao, R.X. Interpretable Convolutional Neural Network Through Layer-Wise Relevance Propagation for Machine Fault Diagnosis. *IEEE Sens. J.* **2020**, *20*, 3172–3181. [CrossRef]
25. Duan, A.; Guo, L.; Gao, H.; Wu, X.; Dong, X. Deep Focus Parallel Convolutional Neural Network for Imbalanced Classification of Machinery Fault Diagnostics. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8680–8689. [CrossRef]
26. Li, X.; Li, J.; Qu, Y.; He, D. Gear Pitting Fault Diagnosis Using Integrated CNN and GRU Network with Both Vibration and Acoustic Emission Signals. *Appl. Sci.* **2019**, *9*, 768. [CrossRef]
27. Xiao, Q.; Li, S.; Zhou, L.; Shi, W. Improved Variational Mode Decomposition and CNN for Intelligent Rotating Machinery Fault Diagnosis. *Entropy* **2022**, *24*, 908. [CrossRef] [PubMed]
28. Zheng, X.; Wu, J.; Ye, Z. An End-To-End CNN-BiLSTM Attention Model for Gearbox Fault Diagnosis. In Proceedings of the 2020 IEEE International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, 18 December 2020; IEEE: Piscataway, NJ, USA, 2020.
29. Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2446–2455. [CrossRef]
30. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **2021**, *8*, 53. [CrossRef]

31. Brunton, S.L.; Kutz, J.N. *Data Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2019.
32. CWRU Bearing Data Center. Available online: https://engineering.case.edu/bearingdatacenter (accessed on 3 August 2020).
33. Hendriks, J.; Dumond, P.; Knox, D.A. Towards Better Benchmarking Using the CWRU Bearing Fault Dataset. *Mech. Syst. Signal Process.* **2022**, *169*, 108732. [CrossRef]
34. Smith, W.A.; Randall, R.B. Rolling Element Bearing Diagnostics Using the Case Western Reserve University Data: A Benchmark Study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–131. [CrossRef]
35. Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Chebel-morello, B.; Zerhouni, N.; Varnier, C. PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests. In Proceedings of the IEEE International Conference on Prognostics and Health Management, Denver, CO, USA, 18–21 June 2012.
36. Lee, J.; Qiu, H.; Yu, G.; Lin, J. *Rexnord Technical Services: Bearing Data Set*; IMS, University of Cincinnati, NASA Ames Prognostics Data Repository: Moffett Field, CA, USA, 2007.
37. Daga, A.P.; Fasana, A.; Marchesiello, S.; Garibaldi, L. The Politecnico Di Torino Rolling Bearing Test Rig: Description and Analysis of Open Access Data. *Mech. Syst. Signal Process.* **2019**, *120*, 252–273. [CrossRef]
38. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]
39. Gemmeke, J.F.; Ellis, D.P.W.; Freedman, D.; Jansen, A.; Lawrence, W.; Moore, R.C.; Plakal, M.; Ritter, M. Audio Set: An Ontology and Human-Labeled Dataset for Audio Events. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 776–780. [CrossRef]
40. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805. [CrossRef]
41. Gozalo-Brizuela, R.; Garrido-Merchan, E.C. ChatGPT Is Not All You Need. A State of the Art Review of Large Generative AI Models. *arXiv* **2023**, arXiv:2301.04655. [CrossRef]
42. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the Opportunities and Risks of Foundation Models. *arXiv* **2021**, arXiv:2108.07258.
43. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
44. Hasan, M.J.; Sohaib, M.; Kim, J.-M. A Multitask-Aided Transfer Learning-Based Diagnostic Framework for Bearings under Inconsistent Working Conditions. *Sensors* **2020**, *20*, 7205. [CrossRef] [PubMed]
45. Wang, Q.; Michau, G.; Fink, O. Domain Adaptive Transfer Learning for Fault Diagnosis. In Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Paris), Paris, France, 2–5 May 2019; IEEE: Piscataway, NJ, USA, 2019.
46. Cao, N.; Jiang, Z.; Gao, J.; Cui, B. Bearing State Recognition Method Based on Transfer Learning under Different Working Conditions. *Sensors* **2019**, *20*, 234. [CrossRef]
47. Pacheco, F.; Drimus, A.; Duggen, L.; Cerrada, M.; Cabrera, D.; Sanchez, R.-V. Deep Ensemble-Based Classifier for Transfer Learning in Rotating Machinery Fault Diagnosis. *IEEE Access* **2022**, *10*, 29778–29787. [CrossRef]
48. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7316–7325. [CrossRef]
49. Brusa, E.; Delprete, C.; Di Maggio, L.G. Deep Transfer Learning for Machine Diagnosis: From Sound and Music Recognition to Bearing Fault Detection. *Appl. Sci.* **2021**, *11*, 11663. [CrossRef]
50. Di Maggio, L.G. Intelligent Fault Diagnosis of Industrial Bearings Using Transfer Learning and CNNs Pre-Trained for Audio Classification. *Sensors* **2022**, *23*, 211. [CrossRef]
51. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative Adversarial Networks: An Overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
52. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Cambridge, MA, USA, 8–13 December 2014; MIT Press: Cambridge, MA, USA, 2014; Volume 2, pp. 2672–2680.
53. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; Adaptive Computation and Machine Learning; The MIT Press: Cambridge, MA, USA, 2016; ISBN 978-0-262-03561-3.
54. Guo, Q.; Li, Y.; Liu, Y.; Gao, S.; Song, Y. Data Augmentation for Intelligent Mechanical Fault Diagnosis Based on Local Shared Multiple-Generator GAN. *IEEE Sens. J.* **2022**, *22*, 9598–9609. [CrossRef]
55. He, W.; Chen, J.; Zhou, Y.; Liu, X.; Chen, B.; Guo, B. An Intelligent Machinery Fault Diagnosis Method Based on GAN and Transfer Learning under Variable Working Conditions. *Sensors* **2022**, *22*, 9175. [CrossRef] [PubMed]
56. Liu, S.; Chen, J.; Qu, C.; Hou, R.; Lv, H.; Pan, T. LOSGAN: Latent Optimized Stable GAN for Intelligent Fault Diagnosis with Limited Data in Rotating Machinery. *Meas. Sci. Technol.* **2021**, *32*, 045101. [CrossRef]
57. Liu, J.; Zhang, C.; Jiang, X. Imbalanced Fault Diagnosis of Rolling Bearing Using Improved MsR-GAN and Feature Enhancement-Driven CapsNet. *Mech. Syst. Signal Process.* **2022**, *168*, 108664. [CrossRef]
58. Zhao, C.; Zhang, L.; Zhong, M. An Improved WGAN-Based Fault Diagnosis of Rolling Bearings. In Proceedings of the 2022 IEEE International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), Chongqing, China, 5–7 August 2022; IEEE: Piscataway, NJ, USA, 2022.

59. Cao, S.; Wen, L.; Li, X.; Gao, L. Application of Generative Adversarial Networks for Intelligent Fault Diagnosis. In Proceedings of the 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE), Munich, Germany, 20–24 August 2018; IEEE: Piscataway, NJ, USA, 2018.
60. Ding, Y.; Ma, L.; Ma, J.; Wang, C.; Lu, C. A Generative Adversarial Network-Based Intelligent Fault Diagnosis Method for Rotating Machinery Under Small Sample Size Conditions. *IEEE Access* **2019**, *7*, 149736–149749. [CrossRef]
61. Li, X.; Zhang, W.; Ding, Q. Cross-Domain Fault Diagnosis of Rolling Element Bearings Using Deep Generative Neural Networks. *IEEE Trans. Ind. Electron.* **2018**, *66*, 5525–5534. [CrossRef]
62. Thirukovalluru, R.; Dixit, S.; Sevakula, R.K.; Verma, N.K.; Salour, A. Generating Feature Sets for Fault Diagnosis Using Denoising Stacked Auto-Encoder. In Proceedings of the 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), Ottawa, ON, Canada, 20–22 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–7.
63. Yan, X.; She, D.; Xu, Y.; Jia, M. Deep Regularized Variational Autoencoder for Intelligent Fault Diagnosis of Rotor–Bearing System within Entire Life-Cycle Process. *Knowl. Based Syst.* **2021**, *226*, 107142. [CrossRef]
64. Yan, X.; She, D.; Xu, Y. Deep Order-Wavelet Convolutional Variational Autoencoder for Fault Identification of Rolling Bearing under Fluctuating Speed Conditions. *Expert Syst. Appl.* **2023**, *216*, 119479. [CrossRef]
65. Shao, S.; Wang, P.; Yan, R. Generative Adversarial Networks for Data Augmentation in Machine Fault Diagnosis. *Comput. Ind.* **2019**, *106*, 85–93. [CrossRef]
66. Liang, P.; Deng, C.; Wu, J.; Yang, Z. Intelligent Fault Diagnosis of Rotating Machinery via Wavelet Transform, Generative Adversarial Nets and Convolutional Neural Network. *Measurement* **2020**, *159*, 107768. [CrossRef]
67. Zhao, B.; Yuan, Q. Improved Generative Adversarial Network for Vibration-Based Fault Diagnosis with Imbalanced Data. *Measurement* **2021**, *169*, 108522. [CrossRef]
68. Pu, Z.; Cabrera, D.; Li, C.; de Oliveira, J.V. VGAN: Generalizing MSE GAN and WGAN-GP for Robot Fault Diagnosis. *IEEE Intell. Syst.* **2022**, *37*, 65–75. [CrossRef]
69. Liu, S.; Jiang, H.; Wu, Z.; Li, X. Data Synthesis Using Deep Feature Enhanced Generative Adversarial Networks for Rolling Bearing Imbalanced Fault Diagnosis. *Mech. Syst. Signal Process.* **2022**, *163*, 108139. [CrossRef]
70. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
71. Jiao, J.; Lin, J.; Zhao, M.; Liang, K.; Ding, C. Cycle-Consistent Adversarial Adaptation Network and Its Application to Machine Fault Diagnosis. *Neural Netw.* **2022**, *145*, 331–341. [CrossRef] [PubMed]
72. Luleci, F.; Necati Catbas, F.; Avci, O. CycleGAN for Undamaged-to-Damaged Domain Translation for Structural Health Monitoring and Damage Detection. *Mech. Syst. Signal Process.* **2023**, *197*, 110370. [CrossRef]
73. Xie, Y.; Zhang, T. A Transfer Learning Strategy for Rotation Machinery Fault Diagnosis Based on Cycle-Consistent Generative Adversarial Networks. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1309–1313.
74. McFadden, P.D.; Smith, J.D. Model for the Vibration Produced by a Single Point Defect in a Rolling Element Bearing. *J. Sound Vib.* **1984**, *96*, 69–82. [CrossRef]
75. Wang, Y.; Liang, M. An Adaptive SK Technique and Its Application for Fault Detection of Rolling Element Bearings. *Mech. Syst. Signal Process.* **2011**, *25*, 1750–1764. [CrossRef]
76. Ericsson, S.; Grip, N.; Johansson, E.; Persson, L.E.; Sjöberg, R.; Strömberg, J.O. Towards Automatic Detection of Local Bearing Defects in Rotating Machines. *Mech. Syst. Signal Process.* **2005**, *19*, 509–535. [CrossRef]
77. Sobie, C.; Freitas, C.; Nicolai, M. Simulation-Driven Machine Learning: Bearing Fault Classification. *Mech. Syst. Signal Process.* **2018**, *99*, 403–419. [CrossRef]
78. Brusa, E.; Delprete, C.; Giorio, L.; Di Maggio, L.G.; Zanella, V. Design of an Innovative Test Rig for Industrial Bearing Monitoring with Self-Balancing Layout. *Machines* **2022**, *10*, 54. [CrossRef]
79. Brusa, E.; Cibrario, L.; Delprete, C.; Di Maggio, L.G. Explainable AI for Machine Fault Diagnosis: Understanding Features' Contribution in Machine Learning Models for Industrial Condition Monitoring. *Appl. Sci.* **2023**, *13*, 2038. [CrossRef]
80. Nash, J.F. Equilibrium Points in *n*-Person Games. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 48–49. [CrossRef] [PubMed]
81. Goodfellow, I. NIPS 2016 Tutorial: Generative Adversarial Networks. *arXiv* **2016**, arXiv:1701.00160.
82. Harris, T.A. *Rolling Bearing Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2001; ISBN 0-471-35457-0.
83. Garreau, D.; Jitkrittum, W.; Kanagawa, M. Large Sample Analysis of the Median Heuristic. *arXiv* **2017**, arXiv:1707.07269.
84. Liu, H.; Zhang, D.; Liu, Z.; Liang, N.; Tao, Y.; He, W. A Method of Vibration Signal Data Enhancement and Fault Diagnosis of Generator Bearings Based on Deep Learning Model. In Proceedings of the 2022 IEEE International Conference on High Voltage Engineering and Applications (ICHVE), Chongqing, China, 25–29 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–4.
85. Meng, Z.; He, H.; Cao, W.; Li, J.; Cao, L.; Fan, J.; Zhu, M.; Fan, F. A Novel Generation Network Using Feature Fusion and Guided Adversarial Learning for Fault Diagnosis of Rotating Machinery. *Expert Syst. Appl.* **2023**, *234*, 121058. [CrossRef]
86. Fukumizu, K.; Gretton, A.; Sun, X.; Schölkopf, B. Kernel Measures of Conditional Dependence. *Adv. Neural Inf. Process. Syst.* **2007**, *20*.
87. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2014**, arXiv:1409.4842.
88. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]

89. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]
90. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.

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.