machine learning & knowledge extraction

MDPI

Article

A Combined Short Time Fourier Transform and Image Classification Transformer Model for Rolling Element Bearings Fault Diagnosis in Electric Motors

Christos T. Alexakos [1], Yannis L. Karnavas [1], Maria Drakaki [2,*], Ioannis A. Tziafettas [1]

[1] Electrical Machines Laboratory, Depterment of Electrical & Computer Engineering, Democritus University of Thrace, 671 00 Xanthi, Hellas, Greece; chrialex@ee.duth.gr (C.T.A.); karnavas@ee.duth.gr (Y.L.K.); ioantzia3@ee.duth.gr (I.A.T.)
[2] Department of Science and Technology, University Center of International Programmes of Studies, International Hellenic University, 570 01 Thermi, Hellas, Greece
* Correspondence: mdrakaki@ihu.gr; Tel.: +30-2310-807524

Abstract: The most frequent faults in rotating electrical machines occur in their rolling element bearings. Thus, an effective health diagnosis mechanism of rolling element bearings is necessary from operational and economical points of view. Recently, convolutional neural networks (CNNs) have been proposed for bearing fault detection and identification. However, two major drawbacks of these models are (a) their lack of ability to capture global information about the input vector and to derive knowledge about the statistical properties of the latter and (b) the high demand for computational resources. In this paper, short time Fourier transform (STFT) is proposed as a pre-processing step to acquire time-frequency representation vibration images from raw data in variable healthy or faulty conditions. To diagnose and classify the vibration images, the image classification transformer (ICT), inspired from the transformers used for natural language processing, has been suitably adapted to work as an image classifier trained in a supervised manner and is also proposed as an alternative method to CNNs. Simulation results on a famous and well-established rolling element bearing fault detection benchmark show the effectiveness of the proposed method, which achieved 98.3% accuracy (on the test dataset) while requiring substantially fewer computational resources to be trained compared to the CNN approach.

Keywords: bearing fault; convolutional neural network; electric motors; short time fourier transform; image classification transformer; fault diagnosis

Citation: Alexakos, C.T.; Karnavas, Y.L.; Drakaki, M.; Tziafettas, I.A. A Combined Short Time Fourier Transform and Image Classification Transformer Model for Rolling Element Bearings Fault Diagnosis in Electric Motors . *Mach. Learn. Knowl. Extr.* **2021**, *3*, 228–242. https://doi.org/10.3390/make3010011

Academic Editor: Andreas Holzinger
Received: 8 January 2021
Accepted: 11 February 2021
Published: 16 February 2021

1. Introduction

Electrical machines are commonly used in industrial and commercial applications, especially for electric motors. Due to their simplicity (e.g., induction motors) and efficiency, these rotating electrical machines are responsible for converting a great amount of electrical energy into mechanical energy worldwide [1]. Additionally, the rapidly evolving industries and the increasing demand for hybrid and electric vehicles indicate that there will be a further increase in this rate of usage. Electrical or mechanical faults will occur during lifetime of an electrical machine [2] and may lead to catastrophic failures. Fault detection and diagnosis is a crucial task to prevent and predict these undesirable failures that can lead to unscheduled and costly downtime. Rolling element bearings are extensively used in electrical machines to ensure their smooth (without much friction) operation. They are sturdy components with typically very long useful lives, yet bearing defects are responsible for the majority of the failures of electrical machines [3]. For example, bearing defects account for more than 51% of failures in induction motors alone [4]. The detection of incipient bearing defects is an important part of condition-based maintenance (CBM). Fault diagnosis can be accomplished with different methods, such as active/supervised diagnosis, which is

expensive and inefficient, and with data-driven diagnosis techniques. There are many ways to detect bearing faults using data-driven methods, such as fault diagnosis through stray flux analysis [5], Park's vector analysis method (PVA) [6], instantaneous power factor (IPF) monitoring [7] and so on. Among them, using the acceleration characteristics of the motor's housing to determine the existence of these faults is the most accurate method, which has become a very well-developed field in recent years [8]. Real-time bearing vibration signal fault detection methods, including traditional methods, machine learning and deep learning methods, have been researched to accurately diagnose bearing defects. Traditional methods require feature extraction of the bearing vibration signal, dimension reduction and classification, which lead to complex mathematical models. To automate the process, machine learning (ML) methods have been introduced, such as the k-nearest neighbors (KNNs) [9], the adaptive neuro-fuzzy inference system (ANFIS) [10], fuzzy cognitive networks (FCNs) [11], the multi-agent system (MAS) approach using intelligent classifiers [12] and the support vector machine (SVM) [13]. In recent years, deep learning methods and associated techniques have been achieving dramatically increased popularity among the research areas of neural networks and artificial intelligence. Their increased processing capabilities, the huge amount of data used for training and the recent advances in machine learning and signal processing are the main reasons for being held in high regard by researchers [14,15].

However, there are many major challenges for this kind of fault diagnosis. Some of them refer to (a) the large amount of data due to the large number of monitoring points of electromechanical equipment, (b) the high sampling frequency of the related sensors, (c) the need to be detected automatically, (d) the nature of the data types, which are diverse and e) the difficulty of extracting the features. Deep learning methods have the ability to overcome those challenges [16]. CNNs have the ability to detect faults by learning optimal filters and they can directly extract and learn the best features from the original signals. They have been applied to behavior recognition [17], classification of electrocardiogram signals [18], speech recognition [19,20] and many other fields. The CNN is superior to traditional methods not only in terms of accuracy, but also in terms of speed, and another key feature of the CNN is its adaptive design. Based on these advantages, researchers have tried to apply deep learning to bearing fault diagnosis [21,22]. While the CNNs are the leading models for image classification, their computational complexity is a drawback; therefore, there is a need for alternative models to be as highly accurate as CNNs but with reduced computational complexity, especially for their usage on embedded (e.g., microcontroller based) systems and other portable applications.

In this context, the paper proposes an effective methodology—which is presented for the first time in the literature—to diagnose electric motor rolling element bearings' faults based on a combination of the short time Fourier transform (STFT) and the image classification transformer (ICT). The main feature of this work is that the authors combined a method to convert time-series data, such as vibration signals, into images using STFT and in-turn apply an ICT to classify the STFT images for bearing fault diagnosis. The paper is organized as follows: In Section 2, a brief description of rolling element bearings is given. The developed methodology is analytically presented in Section 3; the relevant experimental results and discussion are unfolded in Section 4. Finally, Section 5 concludes the work and future directions are given.

2. Bearing Faults

The rolling element bearings act as an electromechanical interface between the stator and the rotor of the motor. In addition, they represent the holding element from the shaft of the machine to ensure a proper rotation of the rotor. The bearings are constituted by two races, the inner race and the outer race, a number of rolling balls and the cage which provides equidistance between the balls as shown in Figure 1. Bearing faults can occur due to a number of factors. The most common factors are misalignments of the rotor, improper lubrication, excessive load and mechanical fatigue [15]. Failures may affect the bearing

on both races and/or on the ball. Several studies have shown that the failure of each bearing element is manifested by a vibration frequency characterizing the fault type [23]. In particular, the following relationships apply:

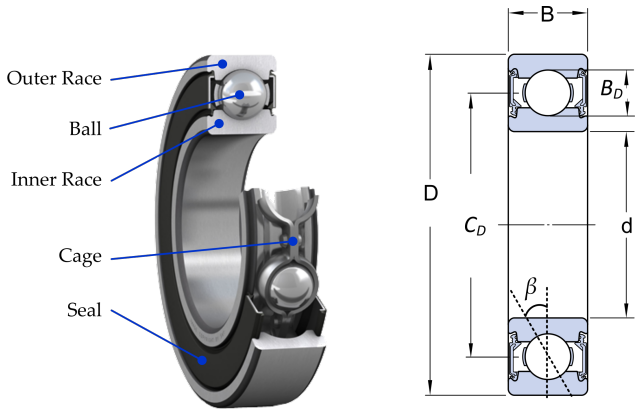

Figure 1. Geometry of a rolling element bearing.

- Characteristic frequency of the ball fault:

$$f_{bf} = \frac{C_D}{B_D} f_r \left(1 - \frac{B_D^2}{C_D^2} cos^2\beta\right) \tag{1}$$

- Characteristic frequency of the inner race fault:

$$f_{irf} = \frac{N_b}{2} f_r \left(1 + \frac{B_D}{C_D} cos\beta\right) \tag{2}$$

- Characteristic frequency of the outer race fault:

$$f_{orf} = \frac{N_b}{2} f_r \left(1 - \frac{B_D}{C_D} cos\beta\right) \tag{3}$$

where N_b is the number of bearing balls, B_D and C_D are the ball and the cage diameters respectively, β is the contact angle and f_r is the mechanical rotor frequency. For experimental purposes, in fault diagnosis studies, artificially generated faults are commonly used which are acquired by drilling or cutting the bearings. Such examples of the fault types can been seen in Figure 2.

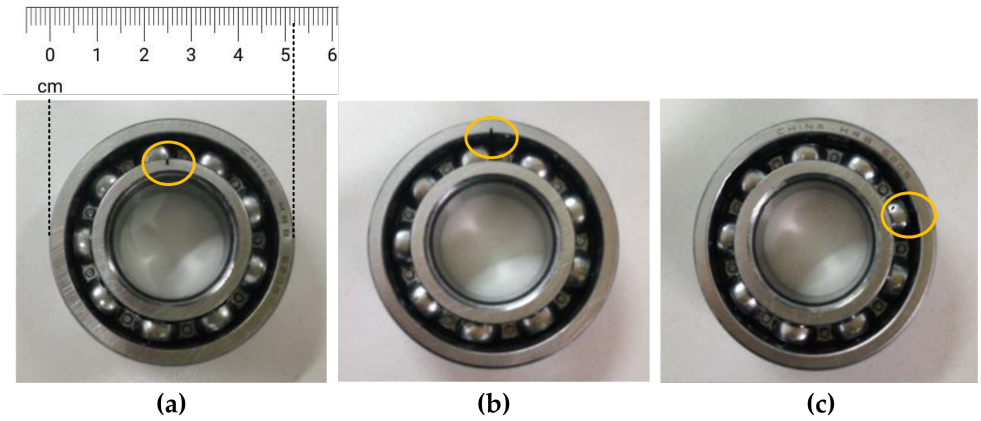

Figure 2. Typical examples of artificially generated bearing faults (i.e., dents and cracks using machining tools). In common scale: (**a**) outer race fault, (**b**) inner race fault, (**c**) ball fault.

3. Methodology and Materials

The proposed methodology mainly comprises three major tasks, i.e.,: (a) vibration raw data input; (b) transformation to time-frequency representation using short time Fourier transformation (STFT); (c) image classification of healthy status or faulty status using the image classification transformer (ICT). Figure 3 illustrates the overall proposed system which is described next.

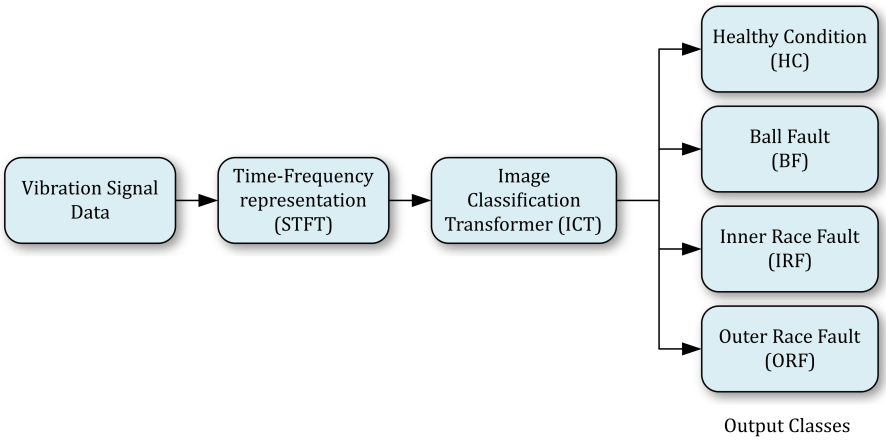

Figure 3. Block diagram of the proposed fault diagnosis system.

3.1. Vibration/Acceleration Data and Characteristics

To evaluate the method, the publicly available seeded fault bearing dataset by Case Western Reserve University (CWRU) Bearing Data Center was used. The data were acquired by using a 2-horsepower (Hp) reliance electric motor with a torque transducer and a dynamometer for applying different loads, ranging from 0 to 3 Hp (Figure 4). Rotating speeds of the motor also varied from 1730 to 1797 rpm. Data for normal bearings were recorded using the deep groove ball bearing SKF6205-2RS JEM type at the drive end (with dimensions: inner race d = 25 mm, outer race D = 52 mm, width B = 15 mm). The drive end bearings were seeded with defects on the inner raceways, outer raceways and rolling elements with the assistance of an electro-discharge machine. The faulty bearings were reinstalled into the test-bed and vibration data were recorded for the same motor loads. The final overall dataset consists of ratings of healthy condition (HC), inner raceway fault (IRF), ball fault (BF) and outer raceway fault (ORF) signals under the considered operating conditions. Especially for the ORF, this has three variants: ORF at the center, ORF at the orthogonal and ORF at the opposite position. In this study, the variable length vibration acceleration signals were recorded at 12,000 samples/s (Hz) for the drive-end bearings.

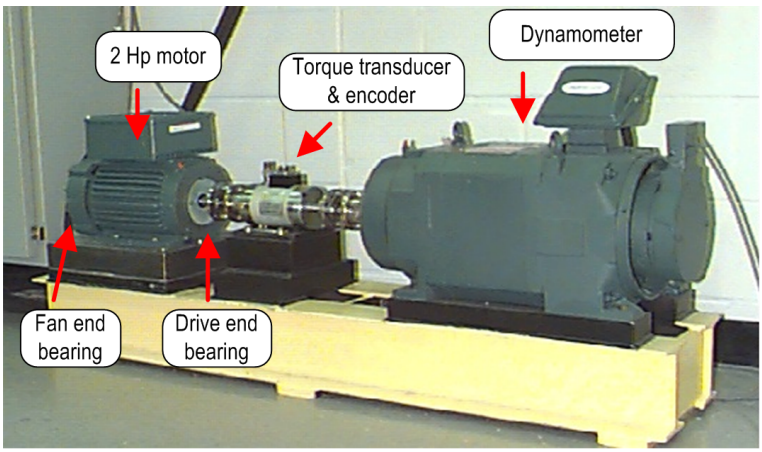

Figure 4. Experimental data acquisition system workbench by CWRU.

For every different health condition of the bearings, as shown in Table 1, 50 samples were extracted from the original data signal duration; i.e., each sample corresponds to the duration of one revolution of the rotor. Thus, 3200 samples were extracted in total and 70% of them were used for training; the remaining 30% used for testing the ICT model.

Table 1. Analytical data description of the CWRU dataset.

Bearing Condition Status		Bearing Defect Diameter (Inches)	Load (Hp)			
			0	1	2	3
			Speed (rpm)			
			1797	1772	1750	1730
HC		-	√	√	√	√
BF		0.007	√	√	√	√
		0.014	√	√	√	√
		0.021	√	√	√	√
		0.028	√	√	√	√
IRF		0.007	√	√	√	√
		0.014	√	√	√	√
		0.021	√	√	√	√
		0.028	√	√	√	√
ORF	Centered @6	0.007	√	√	√	√
		0.014	√	√	√	√
		0.021	√	√	√	√
		0.028	-	-	-	-
	Orthogonal @3	0.007	√	√	√	√
		0.014	-	-	-	-
		0.021	√	√	√	√
		0.028	-	-	-	-
	Opposite @12	0.007	√	√	√	√
		0.014	-	-	-	-
		0.021	√	√	√	√
		0.028	-	-	-	-

3.2. Pre-Processing Using STFT

Raw time domain data are usually used directly as input for deep learning methods for forecasting time-series because of faster training times and reduced computational complexity. For fault diagnosis in electrical machines in the process of acquiring the vibration signals, noise interference is common due to the sensor's noise input or other environmental factors. Additionally, the vibration signals are usually non-stationary due to time-varied rotational speeds. In both cases, time domain signal processing and frequency domain signal processing are not viable, and in order to overcome these limitations, it is necessary to convert the raw time domain data into a two-dimensional function of both time and frequency [24]. Therefore, using a time-frequency analysis technique, such as the short-time Fourier transform (STFT), on the input for deep learning classification methods for bearing fault diagnosis is improving the robustness of the overall methodology and lessening the need for large datasets in order to satisfactory train the models. The short-time Fourier transform (STFT) was the first time-frequency method, which was applied by Gabor [25] in 1946 to speech communication. The STFT may be considered as a method that

breaks down the non-stationary signal into many small segments, which can be assumed to be locally stationary, and applies the conventional FFT to these segments. The STFT of a signal $s_i(\tau)$ is achieved by multiplying the signal by a window function, $h(\tau)$ to produce a modified signal. Since the modified signal emphasizes the signal around time τ, the Fourier transformation will reflect the distribution of frequency around that time; i.e.,

$$S_i(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega t} s(\tau) h(\tau - t) d\tau \quad (4)$$

We may consider $S(\omega)$ as the sum of the Fourier base functions, but the base functions are modulated versions of the window function. Resolutions in time and frequency will be determined by the length of window $h(\tau)$ (Figure 5). A large window length is chosen when we need greater accuracy in frequency and a small window length when we want to have greater accuracy in time. However, the STFT depends greatly on the length of the window, and by varying the window used, one can exchange accuracy in time for accuracy in frequency.

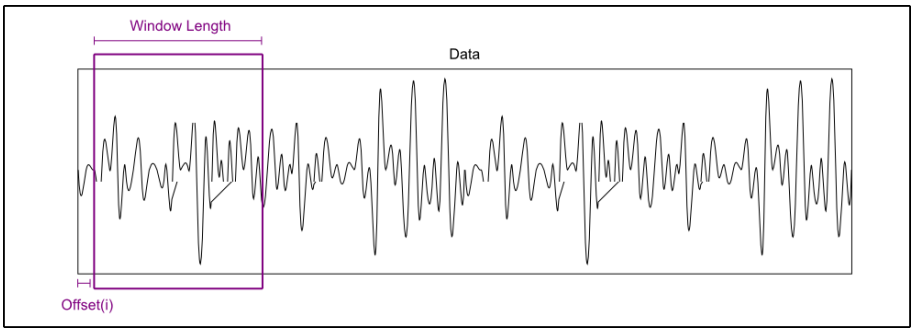

Figure 5. A typical example of a window function concept.

For our proposed methodology, vibration samples' duration varied from 33.4 to 34.6 ms due to load variations. A relevant large time window length $h(\tau)$ was chosen to be one third of the sample duration from 11.1 to 11.5 ms, in order to have greater accuracy in frequency, but also for the signal during that time window length to be considered stationary. After STFT had been applied on raw vibration samples, we were able to create the input images for the classification model in both time and frequency representations. Representative examples of the different fault conditions after using STFT on raw vibration data are shown in Figure 6.

3.3. Image Classification Transformer

In order to classify the output vibration images after applying STFT, we are proposing here the image classification transformer (ICT). ICT is an adaptation of the initial architecture of the transformer designed for natural language processing (NLP) tasks, which was introduced recently by Vaswani et al. [26].

To use the vibration images as input for our transformer, every image $x \in \mathbb{R}^{H \times W \times C}$ has been reshaped into a sequence of flattened 2D patches $x_p \in \mathbb{R}^{N \times (P^2 C)}$, where H, W is the resolution of the initial image, C is the number of channels of the image and P is the resolution of each flattened image patch. Additionally, $N = HW/P^2$ is then the effective sequence length for the transformer. The transformer uses constant widths through all of its layers, so a trainable linear projection maps each vectorized patch to the model dimension D. Moreover, position embeddings are added to the flattened patches to retain positional information of each patch. In this work, 1D position embeddings are used, with $E_{pos} \in \mathbb{R}^{(N+1) \times D}$. Finally, in order for the transformer to work as an image classifier, an extra learnable class embedding has been added to the input sequence. The joint embeddings serves as the input sequence to the encoder, as described by Equation (5).

The overall block diagram of the ICT model is shown in Figure 7 and its topology is described hereafter.

$$z_0 = [x_{class}; x_p^1; x_p^2; ...; x_p^N;] + E_{pos} \quad (5)$$

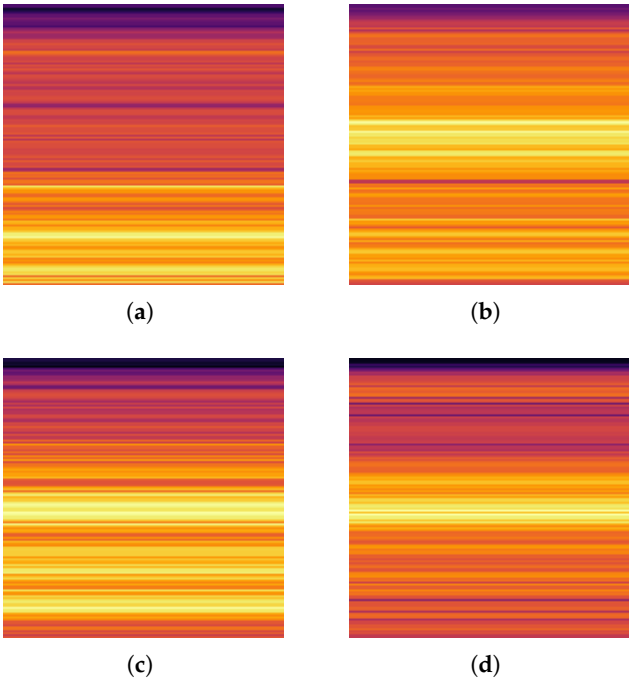

Figure 6. Example output vibration images obtained after applying STFT on raw vibration signals: (**a**) healthy condition (HC), (**b**) ball fault (BF), (**c**) inner race fault (IRF), (**d**) outer race fault (IRF).

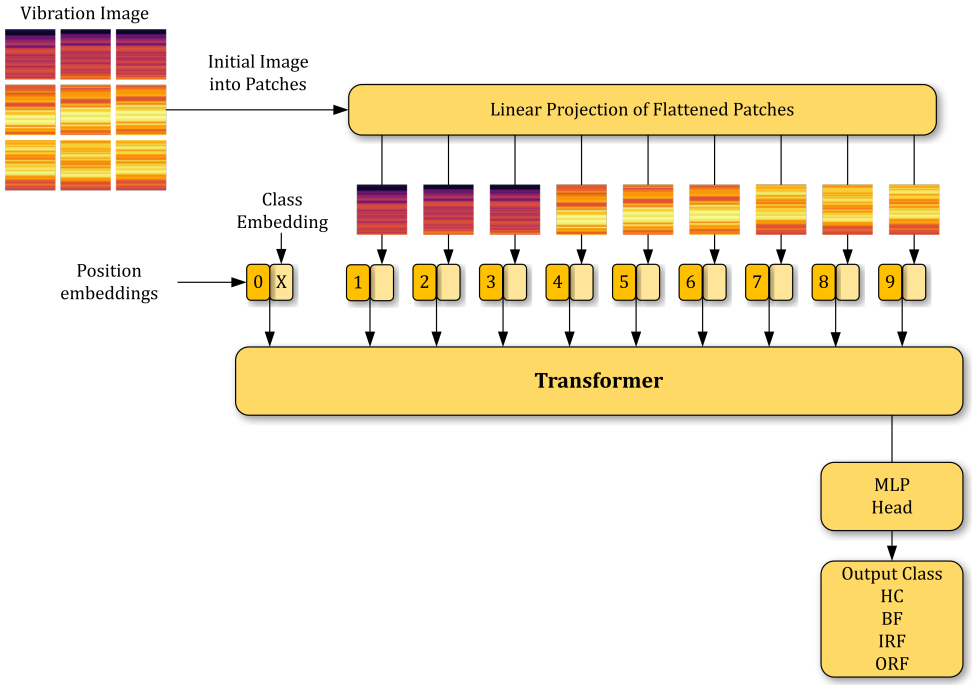

Figure 7. Block diagram of the overall model.

In this work, the transformer is composed by $L = 6$ identical transformer encoders; each one of them contains two layers. The first layer is a "multi-head" self-attention mechanism (MSA) as per the following relationship:

$$z'_l = MSA(LN(z_{l-1})) + z_{l-1} \tag{6}$$

and the second is a multilayer perceptron (MLP) described by

$$z_l = MLP(LN(z'_l)) + z'_l \tag{7}$$

Before every layer, Layernorm (LN) is applied [27] along with residual connections after every layer [28]. The block diagram of the tranformer encoder shown in Figure 8.

The MSA is an extension of self-attention (SA) mechanism, in which k self-attention operations, called "heads," run in parallel and project their concatenated outputs. To keep computing and the number of parameters constant when changing k, the D_h is typically set to D/k. The latter is described by

$$MSA(z) = [SA_1(z); SA_2(z); ...; SA_k(z);]U_{msa} \tag{8}$$

where $U_{msa} \in \mathbb{R}^{kD_h \times D}$ and

$$SA(z) = Softmax(\frac{qk^T}{\sqrt{D_h}}) \tag{9}$$

Regarding the "self attention" operation, for each element in an input sequence $z \in \mathbb{R}^{N \times D}$, a weighted sum over all values v in the sequence is computed. The attention weights $A_{i,j}$ are based on the pairwise similarity between two elements of the sequence and also their respective query q^i and key k^j representations (Figure 9).

Finally, the multilayer perceptron (MLP) is a fully connected feedfoward neural network and contains two layers with GELU activation functions. The GELU activation function is $x\phi(x)$, where $\phi(x)$ is the standard Gaussian cumulative distribution function. The GELU nonlinearity weights inputs by their percentiles, rather than by their signs, as in ReLUs. Consequently the GELU can be thought of as a smoother ReLU. The GELU performs slightly better than ReLU [29] and is commonly preferred over other activation functions for transformer architecture by Vaswani et al. [26] and for state-of-art transformers, such as Google's BERT [30] and OpenAI's GPT-2/3.

$$GELU(x) = xP(X \leq x) = x\phi(x) = x\frac{1}{2}[1 + erf(\frac{x}{\sqrt{2}})] \tag{10}$$

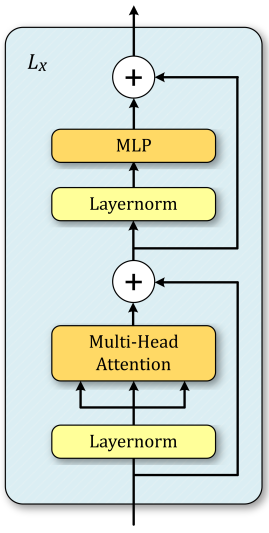

Figure 8. Block diagram of the transformer encoder.

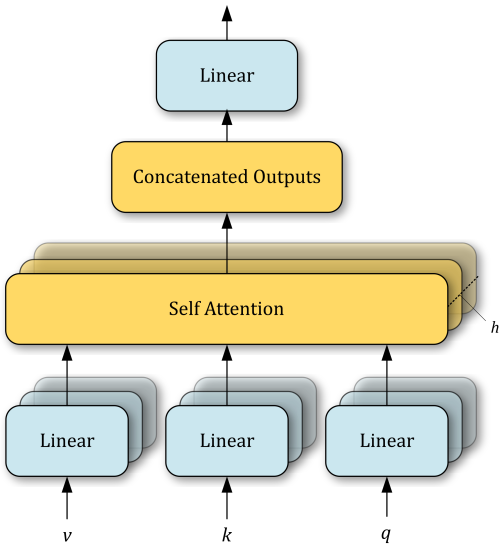

Figure 9. Block diagram of the multi-head self attention mechanism.

4. Experimental Results and Discussion

The data pre-processing and the image classification transformer (ICT) algorithms were developed and written by using *Python* 3.7 and the *pytorch* framework. The training and the evaluation of the ICT has been accomplished using the *Google Colab* cloud service with runtime type *cuda* cores (GPU). Due to the different number of variations of each class mentioned in Table 1 and because they were sampled equally to form our four major classes, the dataset was imbalanced. After the examination of the results, we concluded that each given class is handled perfectly from the model, and therefore there is no necessity to balance the dataset by resampling approaches. As aforementioned, the input data was split as 70% for training and 30% for testing out the overall CWRU dataset. Analytically, the data used for each bearing fault type and for the different algorithmic phases are shown in Table 2.

Table 2. Data split of overall data samples.

Bearing Condition	Training Data	Testing Data	Overall Data
Healthy Condition	140	60	200
Ball Fault	544	256	800
Inner Race Fault	562	238	800
Outer Race fault	966	434	1400
Overall Data	**2212**	**988**	**3200**

The ICT model was trained over 60 epochs in order to learn the robust features for each type of bearing health condition. Additionally, the ICT model was trained to extract and learn the features with a batch size set to 32 and a learning rate of 0.001 through the *Adam* optimizer. *Adam* is a recently introduced optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iteratively based in training data. It was also proven that the *Adam* optimizer performs better compared to other optimization methods [31].

In order to decide which are the best fitting parameters for the ICT model, the training dataset was split randomly into 85% for the training and 15% for validation. We initiated multiple runs of the same training process and we observed that the best fitting parameters occur red between 25 and 50 epochs. The algorithm saved the parameters of each epoch and was able to determine in which epoch the best fitting existed, discarding all the others. As can been seen in Figure 10, at the 48th epoch, the training loss was the lowest, and the

validation loss was minimum; and in the next epochs the training loss kept reducing, but the validation loss was slightly increasing. That means that the model reached the best fitting parameters at epoch 48, and after that was overfitting on the training data. The training accuracy surpassed 90% after the first five epochs and reached 100% at around 48. At the same time, the validation accuracy reached the highest accuracy at epoch 48, and then slightly decreased, as can been seen in Figure 11. Moreover, we trained the model with the Kfolds cross-validation approach with K set to five, and the validation accuracy of each fold along with the average validation accuracy can been seen in Figure 12. Thus, it can be said that, over the training period, the ICT model was able to learn the robust and generalized features of the STFT vibration images in order to diagnose the bearing faults and classify them into healthy or faulty classes accordingly.

The overall computational complexity of the ICT model was calculated to be only 0.05 GMac (billions of multiply and accumulate operations) and the number of parameters was 745,220. It is clarified here that "parameters" were the coefficients of weights and biases. The model during training optimizes these coefficients according to a given optimization algorithm and returns an array which minimize the loss function. Therefore, the number of those parameters is a crucial factor in terms of computational complexity of the model. The storage space used by the trained ICT was found to be only 3 MB, which is important for embedded and portable applications. Moreover, to evaluate the performance of the trained ICT model, 988 samples of the testing dataset were used (Table 2). With an average accuracy of 98.3% on the testing dataset, as clearly described in the classification report shown in Table 3, it can be said that the performance of the trained ICT model is very satisfactory.

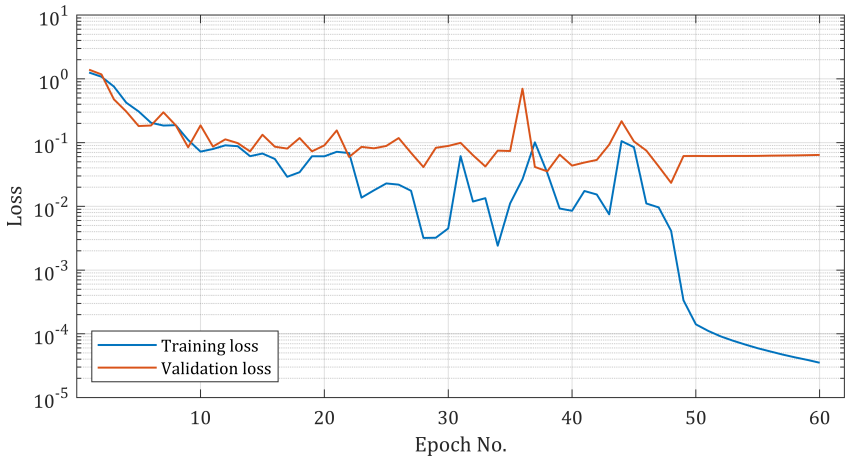

Figure 10. Training and validation loss during training process.

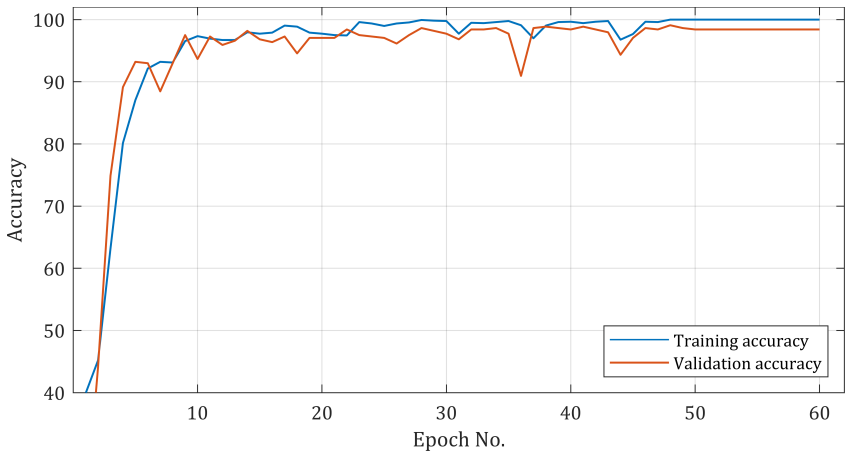

Figure 11. Accuracy of the training dataset.

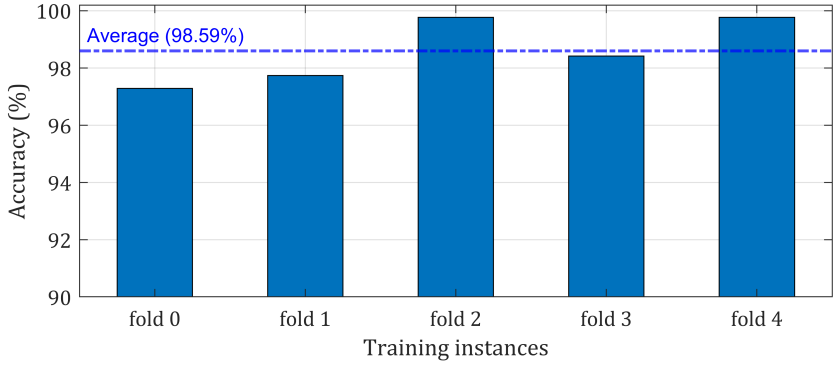

Figure 12. Cross-validation accuracy on validation folds.

However, using overall accuracy as the only metric to evaluate the model may lead to false assumptions with an under-representative, imbalanced dataset. We tried to determine whether any of the classes were under-represented in the dataset metrics, such as precision and recall, for each class used. Given the overall performance of the model during the training/validation process; the high values for precision and recall metrics for each class [32], as can be seen in Table 3; and the confusion matrix proving that the overall accuracy is the real accuracy, the dataset is representative for each class and the model can handle each class correctly. Therefore, there was no necessity to proceed in resampling approaches in order to balance the dataset.

The ICT model is fully capable of extracting the features from the testing dataset and classifying the features for the respective healthy or faulty rolling element bearing conditions. Furthermore, as seen in [33], the accuracy results of an imbalanced dataset are lower than any other resambling approach in order to balance the dataset; therefore, this indicates the proposed methodology would have achieved even higher overall accuracy with resambling approaches. Finally, the confusion matrix shown in Table 4 explains the classification results of the testing dataset. The testing samples classified with satisfactory accuracy, with only a few misclassifications as false positives and false negatives.

Table 3. Classification report of the testing dataset.

Classification Report					
Class	**Accuracy (*x*100%)**	**Precision**	**Recall**	**F1 Score**	**Support**
HC	1.00	1.00	1.00	1.00	60
BF	0.98	0.99	0.98	0.98	238
IRF	0.98	0.97	0.98	0.97	256
ORF	0.97	0.98	0.98	0.98	434
Total	0.98	0.98	0.98	0.98	988

Table 4. The confusion matrix of the testing dataset.

		Predicted Class				
		HC	BF	IRF	ORF	**Total**
Actual Class	HC	**60**	0	0	0	60
	BF	0	**234**	1	3	238
	IRF	0	1	**250**	5	256
	ORF	0	1	5	**428**	434
	Total	60	236	256	436	**988**

4.1. Comparison with Other Models and Methods

For comparison purposes regarding our proposed ICT model and state-of-art leading CNN models in image classification (and commonly used recently as components in electric motors related fault diagnosis systems), we also developed two CNN models with architectures such those described in the recent works of Lee et al. [34], Tra et al. [35] and Hsueh et al. [36]. The two CNN models were trained and tested with the same dataset as the one used in our proposed ICT model.

The first CNN model was composed by two convolutional layers followed by three fully connected layers. It was found that the trained CNN model used a huge amount of storage space due to the large number of parameters stored. Moreover, the overall computational complexity was also very high, with an average accuracy of 97.8% on the testing dataset (the computational complexity and number of parameters can been seen at Table 5). The second CNN model which was trained and tested had a similar composition as the first one, i.e., by two convolutional layers, but a max-pooling layers had been added after each convolutional layer followed by three fully connected layers. Due to the max-pooling layers, the number of parameters, storage space and overall computational complexity were reduced compared to the first CNN model, but were still high compared to those of our proposed ICT model, as can be seen in Table 5. Additionally, a loss in average accuracy of 0.5% was also observed in this case, due to the possible useful information loss after every max-pooling layer. Thus, the average accuracy of the second CNN model was found to be 97.2%.

Table 5. Analytical comparison with state-of-art CNN architectures in terms of accuracy, computational complexity and storage requirements.

Model	Accuracy (Avg.)	Parameters	Computational Complexity	Storage
CNN	97.8%	355.27 M	0.41 GMac	1 GB
CNN with max-pooling	97.2%	1.79 M	0.16 GMac	7 MB
ICT (proposed)	**98.3%**	**745.22 K**	**0.05 GMac**	**3 MB**

Finally, we also compared the results of our proposed methodology with other traditional machine learning and deep learning methods, such as the dense neural network (DNN) [37], the support vector machine (SVM) [38], the deep belief network (DBN) [39], a time-domain 1D-CNN [40] and a time-domain 2D-CNN [41]. Table 6 clearly demonstrates that the obtained results are by far superior compared to those of the aforementioned methods, and therefore, in conjunction with the previous comparison, the authors believe that the proposed methodology can be used as a highly accurate and "lightweight" system for rolling element bearing fault diagnosis in relevant electrical motor drives.

Table 6. Comparison results with other machine learning competitive methods.

Methods	Accuracy
Dense Neural Network (**DNN**) [37]	80.0%
Support Vector Machine (**SVM**) [38]	89.5%
Deep Belief Network (**DBN**) [39]	92.2%
1-D (time domain) CNN (**1DtdCNN**) [40]	93.3%
2-D (time domain) CNN (**2DtdCNN**) [41]	96.0%
Image Classification Transformer (ICT) (proposed)	**98.3%**

5. Conclusions and Future Work

A modified transformer scheme, the image classification transformer (ICT), was successfully combined with STFT and proposed in this work as an effective methodology to diagnose the faults of rolling elements bearings in electric motors. The novel idea behind this work is the combination of a method to convert initially 1D time-series data, such as vibration signals, into 2D information samples (images) using STFT, and in turn, applying a well-designed ICT to classify the STFT images for bearing fault diagnosis. The examined methodology performed better than other traditional and deep learning methods. Future work will be focused on collecting more diverse vibration data samples in order to make the methodology even more robust when collecting different types of electric motor fault data, such as rotor broken bars, broken end rings, winding faults and different types of data such as current signals and acoustic signals. to examine whether the same methodology is as or maybe more effective for different types of data and faults. Additionally, effort will be made to set up the proposed scheme experimentally in an embedded microcontroller or system-on-chip (SoC) system for real time applications.

Author Contributions: Conceptualization, C.T.A. and Y.L.K.; Formal analysis, C.T.A. and Y.L.K.; Investigation, Y.L.K., M.D. and I.A.T.; Methodology, C.T.A., Y.L.K. and I.A.T.; Software, C.T.A. and Maria Drakaki; Supervision, Y.L.K.; Validation, M.D.; Visualization, Y.L.K.; Writing—original draft, C.T.A.; Writing—review & editing, Y.L.K. and I.A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not Applicable, the study does not report any data.

Acknowledgments: The authors are thankful to Case Western Reserve University for providing them access to their ball bearing datasets.

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

Abbreviations

The following abbreviations are used in this manuscript:

ANFIS	Adaptive Neuro-Fuzzy Inference System
IRF	Inner Race Fault
BF	Ball Fault
KNN	k-Nearest Neighbor
CBM	Condition-Based Maintenance
MAS	Multi Agent System
CNN	Convolutional Neural Network
ML	Machine Learning
CWRU	Case Western Reserve University
MLP	Multi-Layer Perceptron
DBN	Deep Belief Networks
MSA	Multi-head Self-Attention
DNN	Dense Neural Networks
NLP	Natural Language Processing
FCN	Fuzzy Cognitive Networks
ORF	Outer Race Fault
GMAC	billions of Mac (multiply+sum operations)
PVA	Park's Vector Analysis
GPU	Graphich Processing Unit
SA	Self-Attention
HC	Healthy Condition
SoC	System-on-Chip
ICT	Image Classification Transformer
STFT	Short Time Fourier Transform
IPF	Instantaneous Power Factor
SVM	Support Vector Machine

References

1. Dineva, A.; Mosavi, A.; Ardabili, S.F.; Vajda, I.; Shamshirband, S.; Rabczuk, T.; Chau, K.W. Review of Soft Computing Models in Design and Control of Rotating Electrical Machines. *Energies* **2019**, *12*, 1049. [CrossRef]
2. Bazine, S.; Trigeassou, J.C. Faults in Electrical Machines and their Diagnosis. In *Electrical Machines Diagnosis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013; pp. 1–22. [CrossRef]
3. Frosini, L.; Harlisca, C.; Szabo, L. Induction Machine Bearing Fault Detection by Means of Statistical Processing of the Stray Flux Measurement. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1846–1854. [CrossRef]
4. Thorsen, O.; Dalva, M. Failure Identification and Analysis for High-Voltage Induction Motors in the Petrochemical Industry. *IEEE Trans. Ind. Appl.* **1999**, *35*, 810–818. [CrossRef]
5. Henao, H.; Capolino, G.A.; Fernandez-Cabanas, M.; Filippetti, F.; Bruzzese, C.; Strangas, E.; Pusca, R.; Estima, J.; Riera-Guasp, M.; Hedayati-Kia, S. Trends in Fault Diagnosis for Electrical Machines: A Review of Diagnostic Techniques. *IEEE Ind. Electron. Mag.* **2014**, *8*, 31–42. [CrossRef]
6. Neelam, M.; Ratna, D. Detection of Bearing Faults of Induction Motor Using Park's Vector Approach. *Int. J. Eng. Technol.* **2010**, *2*, 263–266.
7. Ibrahim, A.; Badaoui, M.E.; Guillet, F.; Bonnardot, F. A New Bearing Fault Detection Method in Induction Machines Based on Instantaneous Power Factor. *IEEE Trans. Ind. Electron.* **2008**, *55*, 4252–4259. [CrossRef]
8. 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*, 100–131. [CrossRef]
9. Toma, R.N.; Prosvirin, A.E.; Kim, J.M. Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers. *Sensors* **2020**, *20*, 1884. [CrossRef] [PubMed]
10. Karnavas, Y.L.; Chasiotis, I.D.; Vrangas, A. Fault Diagnosis of Squirrel-Cage Induction Motor Broken Bars based on a Model Identification Method with Subtractive Clustering. In Proceedings of the 2017 IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), Tinos, Greece, 29 August–1 September 2017. [CrossRef]
11. Karatzinis, G.; Boutalis, Y.S.; Karnavas, Y.L. Motor Fault Detection and Diagnosis Using Fuzzy Cognitive Networks with Functional Weights. In Proceedings of the 2018 26th Mediterranean Conference on Control and Automation (MED), Zadar, Croatia, 19–22 June 2018. [CrossRef]
12. Drakaki, M.; Karnavas, Y.L.; Karlis, A.D.; Chasiotis, I.D.; Tzionas, P. Study on Fault Diagnosis of Broken Rotor Bars in Squirrel Cage Induction Motors: A Multi-Agent System Approach using Intelligent Classifiers. *IET Electr. Power Appl.* **2020**, *14*, 245–255. [CrossRef]
13. Konar, P.; Chattopadhyay, P. Bearing Fault Detection of Induction Motor using Wavelet and Support Vector Machines (SVMs). *Appl. Soft Comput.* **2011**, *11*, 4203–4211. [CrossRef]
14. Deng, L. Deep Learning: Methods and Applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [CrossRef]
15. Karnavas, Y.L.; Plakias, S.; Chasiotis, I.D. Extracting Spatially Global and Local Attentive Features for Rolling Bearing Fault Diagnosis in Electrical Machines using Attention Stream Networks. *IET Electr. Power Appl.* **2021**. [CrossRef]
16. Lei, Y. A Deep Learning-based Method for Machinery Health Monitoring with Big Data. *J. Mech. Eng.* **2015**, *51*, 49. [CrossRef]
17. Carmona, J.M.; Climent, J. Human Action Recognition by Means of Subtensor Projections and Dense Trajectories. *Pattern Recogn.* **2018**, *81*, 443–455. [CrossRef]
18. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 664–675. [CrossRef]
19. Hu, X.; Lu, X.; Hori, C. Mandarin Speech Recognition using Convolution Neural Network with Augmented Tone Features. In Proceedings of the The 9th International Symposium on Chinese Spoken Language Processing, Singapore, 12–14 September 2014; pp. 15–18.
20. Turaga, S.C.; Murray, J.F.; Jain, V.; Roth, F.; Helmstaedter, M.; Briggman, K.; Denk, W.; Seung, H.S. Convolutional Networks Can Learn to Generate Affinity Graphs for Image Segmentation. *Neural Comput.* **2010**, *22*, 511–538. [CrossRef]
21. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent Advances in Convolutional Neural Networks. *Pattern Recogn.* **2018**, *77*, 354–377. [CrossRef]
22. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep Learning and its Applications to Machine Health Monitoring. *Mech. Syst.d Signal Process.* **2019**, *115*, 213–237. [CrossRef]
23. Harlişca, C.; Szabo, L. Bearing Faults Condition Monitoring—A Literature Survey. *J. Comput. Sci. Control Syst.* **2012**, *5*, 19–22.
24. Nguyen, T.P.K.; Khlaief, A.; Medjaher, K.; Picot, A.; Maussion, P.; Tobon, D.; Chauchat, B.; Cheron, R. Analysis and Comparison of Multiple Features for Fault Detection and Prognostic in Ball Bearings. In *Proceedings of the 4th European Conference of the Prognostics and Health Management Society*; CCSD: Utrecht, The Netherlands, 2018; pp. 1–9.
25. Gabor, D. Theory of Communication. Part 1: The Analysis of Information. *J. Inst. Electr. Eng.-Part III Radio Commun. Eng.* **1946**, *93*, 429–441. [CrossRef]
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. *arXiv* **2017**, arXiv:cs.CL/1706.03762.
27. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer Normalization. *arXiv* **2016**, arXiv:stat.ML/1607.06450.
28. 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. [CrossRef]

29. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs). *arXiv* **2020**, arXiv:cs.LG/1606.08415.
30. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:cs.CL/1810.04805.
31. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:cs.CL/1412.6980.
32. Santos, P.; Maudes, J.; Bustillo, A. Identifying Maximum Imbalance in Datasets for Fault Diagnosis of Gearboxes. *J. Intell. Manuf.* **2015**, *29*, 333–351. [CrossRef]
33. Zhang, W.; Li, X.; Jia, X.D.; Ma, H.; Luo, Z.; Li, X. Machinery Fault Diagnosis with Imbalanced Data using Deep Generative Adversarial Networks. *Measurement* **2020**, *152*, 107377. [CrossRef]
34. Lee, J.H.; Pack, J.H.; Lee, I.S. Fault Diagnosis of Induction Motor Using Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 2950. [CrossRef]
35. Tra, V.; Kim, J.; Khan, S.A.; Kim, J.M. Bearing Fault Diagnosis under Variable Speed Using Convolutional Neural Networks and the Stochastic Diagonal Levenberg-Marquardt Algorithm. *Sensors* **2017**, *17*, 2834. [CrossRef]
36. Hsueh, Y.M.; Ittangihal, V.R.; Wu, W.B.; Chang, H.C.; Kuo, C.C. Fault Diagnosis System for Induction Motors by CNN Using Empirical Wavelet Transform. *Symmetry* **2019**, *11*, 1212. [CrossRef]
37. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep Neural Networks: A Promising Tool for Fault Characteristic Mining and Intelligent Diagnosis of Rotating Machinery with Massive Data. *Mech. Syst. Signal Process.* **2016**, *72–73*, 303–315. [CrossRef]
38. Deng, W.; Yao, R.; Zhao, H.; Yang, X.; Li, G. A Novel Intelligent Diagnosis Method using Optimal LS-SVM with Improved PSO Algorithm. *Soft Comput.* **2017**, *23*, 2445–2462. [CrossRef]
39. Shao, H.; Jiang, H.; Zhang, X.; Niu, M. Rolling Bearing Fault Diagnosis using an Optimization Deep Belief Network. *Meas. Sci. Technol.* **2015**, *26*, 115002. [CrossRef]
40. Eren, L.; Ince, T.; Kiranyaz, S. A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier. *J. Signal Process. Syst.* **2018**, *91*, 179–189. [CrossRef]
41. Li, M.; Wei, Q.; Wang, H.; Zhang, X. Research on Fault Diagnosis of Time-Domain Vibration Signal based on Convolutional Neural Networks. *Syst. Sci. Control Eng.* **2019**, *7*, 73–81. [CrossRef]