

Article

Gearbox Fault Identification Framework Based on Novel Localized Adaptive Denoising Technique, Wavelet-Based Vibration Imaging, and Deep Convolutional Neural Network

Cong Dai Nguyen , Zahoor Ahmad  and Jong-Myon Kim * 

Department of Electrical, Electronics, and Computer Engineering, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 44610, Korea; dainc@mail.ulsan.ac.kr (C.D.N.); zahooruou@mail.ulsan.ac.kr (Z.A.)

* Correspondence: jmkim07@ulsan.ac.kr; Tel.: +82-52-259-2217

Abstract: This paper proposes an accurate and stable gearbox fault diagnosis scheme that combines a localized adaptive denoising technique with a wavelet-based vibration imaging approach and a deep convolution neural network model. Vibration signatures of a gearbox contain important fault-related information. However, this useful fault-related information is often overwhelmed by random interference noises. Furthermore, the varying speed of gearboxes makes it difficult to distinguish the fault-related frequencies from the interference noises. To obtain a noise-free signal for extraction of fault-related information under variable speed conditions, first, a new localized adaptive denoising technique (LADT) is applied to the vibration signal. The new localized adaptive denoising technique results in optimized vibration sub-bands with negligible background noise. To obtain fault-related information, the wavelet-based vibration imaging approach (WVI) is applied to the denoised vibration signal. The wavelet-based vibration imaging approach decomposes the vibration signal into different time–frequency scales, these scales are reflected by a two-dimensional image called a scalogram. The scalograms obtained from the wavelet-based vibration imaging approach are provided as an input to the deep convolutional neural network architecture (DCNA) for extraction of discriminant features and classification of multi-degree tooth faults (MDTFs) in a gearbox under variable speed conditions. The proposed scheme outperforms the already existing state-of-the-art gearbox fault diagnosis methods with the highest classification accuracy of 100%.

Keywords: deep convolutional network; gearbox fault diagnosis; localized adaptive denoising



Citation: Nguyen, C.D.; Ahmad, Z.; Kim, J.-M. Gearbox Fault Identification Framework Based on Novel Localized Adaptive Denoising Technique, Wavelet-Based Vibration Imaging, and Deep Convolutional Neural Network. *Appl. Sci.* **2021**, *11*, 7575. <https://doi.org/10.3390/app11167575>

Academic Editor: Jordi Cusido

Received: 25 June 2021

Accepted: 14 August 2021

Published: 18 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Gearboxes play an important role in numerous industrial machines, vehicles, and wind turbines [1–3]. Due to the operation of gearboxes in harsh conditions, gear defects are found to be the most common defects in gearboxes [4]. A fault in the gearbox can result in catastrophic failures, economic losses, and danger to the operating staff. For this reason, early fault detection of the gearbox is of primary importance. The condition-based monitoring approach suggests maintenance action based on the data collected from the gearbox. This strategy allows the gearbox to function for a long time with minimal maintenance costs [5,6].

Gear fault signatures are sensed and acquired by two types of sensors: accelerometers and acoustic emission sensors [7]. Vibration signatures collected by the accelerometer from a gearbox carry enough fault-related information and can be used for efficient gear fault diagnosis [8]. Vibration signals obtained from a gearbox consist of meshing frequency harmonics, blended sideband frequencies, and other free oscillation frequencies. Therein, the meshing frequency harmonics and blended sideband frequencies are the fundamental defect-related frequencies that help in the process of identifying gear defects [9,10]. The vibration signals obtained from a gearbox under variable speeds are complex and non-stationary; furthermore, the gear fault-related elements are often overwhelmed by the

noise. To identify the fault symptoms in this complex vibration signal, the fault diagnosis technique tries to reduce the noise in the raw vibration signal [11]. In its raw form, the gearbox vibration signal contains various types of interference noises. The main sources for these interference noises are the interconnected systems, such as the electrical-electronic control and measuring systems, the mechanical systems (the influence of the mechanical resonances such as shaft, bearings, gears, etc.), and background noise [12,13]. The random behavior of these noises (i.e., random magnitudes, random appearances anywhere in the observed ranges of vibration signals) makes the noisy components dominant over the fault-related components in the vibration signal, and thus these noises overwhelm the fault-related components. To address this issue, a signal-processing technique, which can reduce the noise in the raw vibration signal, to highlight the fault-related meshing frequency harmonics and sidebands (fault-affiliated elements) for gearbox fault diagnosis in early stages is urgently needed.

In the past, numerous signal processing techniques, such as Fourier transform (FT), envelope spectral analysis, Hilbert transform (HT), spectrogram or spectral analysis of a fixed timing-window Fourier transform (STHT), empirical mode decomposition (EMD), and wavelet-based spectral analysis (WA), have been developed for the processing of stationary and non-stationary complex signals [14–19]. To enhance the performance of the basic signal processing techniques, hybrid signal processing techniques such as EMD and HT and EMD and WA have also been introduced [20,21]. The vibration signatures obtained from a gearbox under faulty conditions are non-stationary. To obtain fault-related information from the non-stationary vibration signal, time–frequency domain techniques are applied. These techniques commonly use window-based filtering, digital filtering, threshold estimation, decomposition modes in the form of intrinsic mode functions, and wavelet-based transformation. Their fault identification efficiencies have been confirmed by classifying fault states (e.g., a fault-free state and a defective state) and denoising in some cases. In the case of a gearbox, the fault-related information is distorted by the hefty noise present in the raw signal. Therefore, noise reduction techniques before applying the time–frequency domain signal processing technique will be helpful for the identification of MDTF in gearboxes. Nguyen et al. [22,23] proposed an adaptive noise reduction model, which effectively reduced the noise in the impaired signal. The resultant impaired signal is then used for the classification of gearbox multi-level tooth cut faults under variable speed conditions. The effectiveness of the adaptive noise reduction model lies in adaptively adjusting optimal parameters of the Gaussian function, which are connected to the optimal weights of the adaptive filter, along the whole frequency range of a vibration signal. Nevertheless, the frequency spectrum of a vibration signal obtained from a gearbox is composed of meshing frequency harmonics, sideband components, and random noises, with different probability distributions. It should be noted that the influence of random noises and the change in stiffness of the gear under defect makes the vibration signal non-stationary and complex. Therefore, a single optimal parameter set of the Gaussian reference signal along the entire frequency range is less effective for noise reduction. To address this issue, a localized adaptive denoising technique (LADT) is proposed in this paper. The proposed LADT is a modified version of the adaptive noise reduction model proposed in [22]. The LADT adaptively transforms the raw vibration signals to the optimized subbands, which accounts for the majority of the defect-related information. The proposed method can reduce noise more effectively than the previous adaptive denoising models, while maintaining original fault-related information. The resultant impaired signal from LADT is then used for feature engineering and fault classification in the proposed scheme of gearbox fault diagnosis.

After signal preprocessing, feature preprocessing and fault classification are the most important steps in the fault identification system. Conventional methods for the fault diagnosis of gearboxes used handcrafted features. After extracting a limited number of features from the signal in the conventional methods, domain knowledge was used for discriminant feature selection. These discriminant features were then classified using machine

learning algorithms, such as support vector machines (SVMs), k-nearest neighbors (KNNs), decision tree algorithms (DTAs), and artificial neural networks (ANNs) [24–29]. However, the handcrafted features need domain knowledge and expertise for the identification of discriminant features. Furthermore, feature engineering techniques, such as dimensionality reduction for discriminant feature selection, result in fault-related information loss. Thus, the conventional methods might not be appropriate for the classification and identification of MDTF defects in the gearbox. In addition, classification algorithms, such as KNNs, SVMs, DTAs, and ANNs, is strongly dependent on the quality of the provided features. To address the above-discussed problems, this paper proposes a scheme of self-generating feature space. The proposed scheme first transforms a low-noise vibration signal into a two-dimensional (2D) image using wavelet transform and obtains WVI's. The WVI's reflect the 2D distributed power spectra of the optimized vibration sub-bands. To obtain fault-related information from the WVI and classify them into their representative classes, the proposed method used DCNA. Deep learning models (DLMs) have been used widely in the areas of finance, natural language processing, and image processing [30–33]. For condition monitoring of a rotating machine, there exist a variety of DLMs based fault diagnosis frameworks, such as stacked denoising autoencoder [34], recurrent neural network [35], long short term memory (LSTM) networks [36], gated recurrent unit network [37], and convolutional neural network (CNN) [37,38]. One of the deep learning models, CNN, is a famous model because of its visual understanding [39]. Deep convolutional network architecture (DCNA) has been created for image processing and recognition, and then developed for fault diagnosis of rotation types of machinery by self-regulation and deep exploration of the latent fault-reflected features of vibration signals [40–43].

The contributions of this study are briefly explained as follows:

- (1) A new signal preprocessing approach LADT is developed. The LADT is an adaptive algorithm that considers each principal frequency segment along the frequency spectrum of a vibration signal to fetch the optimized Gaussian parameters, called localized optimal parameters. The outputs of the LADT, which are optimized vibration sub-bands, contain fault-related information with very low interference noise.
- (2) To discriminate and highlight the fault-related information in the vibration signals of MDTF defect types in the time–frequency domain, the WVI technique is applied.
- (3) Potential features are extracted from the WVI's and classified using DCNA. The latent features of DCNA contain discriminant fault-related features. To classify the fault-related features into their respective classes, DCNA then uses the fine-tuning process based on the backpropagation algorithm.

The remaining sections of this work are arranged as follows: the vibration characteristics of the gearbox are explained in Section 2; Section 3 describes the technical background. The experimental setup and proposed diagnosis scheme are explained in Section 4. Section 5 presents the discussion and evaluation of the experimental results obtained from the proposed scheme, and finally, the conclusion of this study is presented in Section 6.

2. The Specification of a Gearbox Vibration Signal

A fault in the gear results in a change in the stiffness. This stiffness can be observed in the vibration spectrum at specific characteristic frequencies. These characteristic frequencies represent the tooth meshing stiffness. The meshing frequency in the vibration spectrum of the gearbox represents the symptoms of a defect in the gearbox, as the meshing frequency changes whenever an MDTF occurs in the gearbox [44]. Considering a gearbox operating under normal conditions, the vibration signature obtained from the gearbox is a stationary signal with tooth meshing frequency; this signal can be formulated as follows [45]:

$$x_g(t) = \sum_{p=1}^P X_p \cos(2\pi p f_m t + \varepsilon_p), \quad (1)$$

where $x_g(t)$ represents the vibration signal of a gear operating under normal conditions, X_p and ε_p stand for amplitude and phase of p -th harmonic of a meshing frequency, P denotes harmonics of the meshing frequency, and f_m denotes meshing frequency, which can be computed using the parameter of pinion wheel ($f_m = \text{number of pinion teeth} \times \text{rotational frequency of a pinion wheel}$) or using a gear wheel ($f_m = \text{number of gear teeth} \times \text{rotational frequency of a gear wheel}$). The meshing frequency and its harmonics are considered useful components for the fault diagnosis process. Figure 1a shows an example of the frequency spectrum of a vibration signal in the perfect condition.

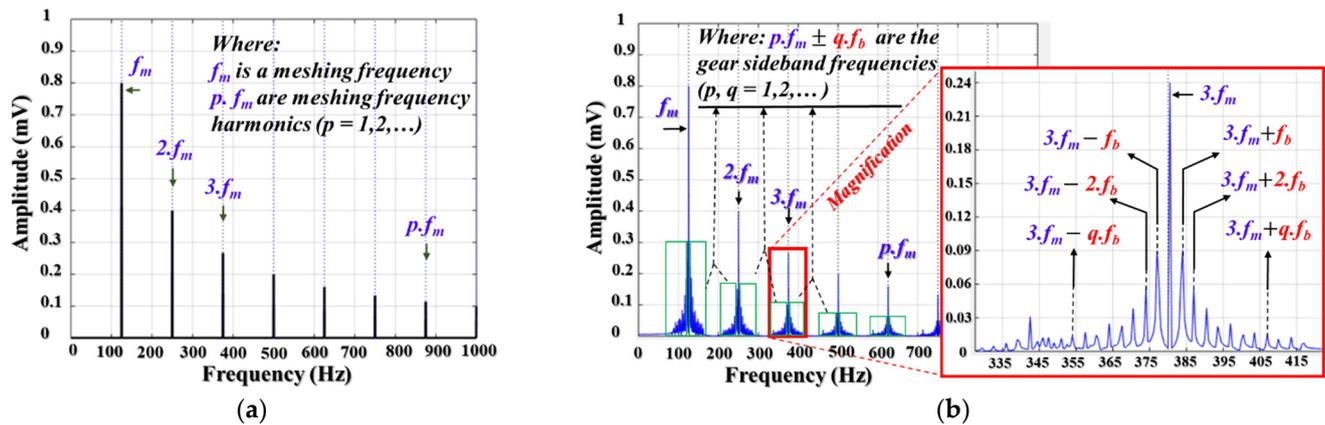


Figure 1. The frequency spectrum of a gearbox (a) under normal conditions and (b) under defective conditions.

A fault in the gearbox makes the vibration signal non-stationary, resulting in a complex frequency spectrum. During the gearbox operation, transmission occurs between the motion source (e.g., three-phase motor and a drive shaft) and a load (a non-drive-shaft and a load) through a pair of gears (pinion wheel and gear wheel). The non-stationary impulses start appearing in the vibration signal when there is an impulsive change in the angular velocity. The angular velocity changes impulsively when the two wheels rotate across a faulty tooth (e.g., missing tooth, cracked tooth, chipped tooth, or worn tooth) [46]. Therefore, the vibration signals obtained from a faulty gearbox exhibit non-stationary behavior, for which the frequency spectrum contains harmonics of tooth meshing frequency, sidebands (the frequency tones are distributed in the two sides of harmonics of a meshing frequency), and other oscillation components. The vibration signal can be presented as a combination of phase and amplitude modulation signal [47], as follows:

$$x_m(t) = \sum_{p=0}^P X_p (1 + \beta_p(t)) \cos(2\pi p f_m t + \varphi_p + \phi_p(t)). \tag{2}$$

Here, $\beta_p(t) = \sum_{q=0}^Q B_{pq} \cos(2\pi q f_b t + \gamma_{pq})$ and $\phi_p(t) = \sum_{q=0}^Q \Phi_{pq} \cos(2\pi q f_b t + \varepsilon_{pq})$ represent the amplitude and phase modulation functions of the defective vibration signal. f_b is the sideband frequency, Q stands for the total number of sideband tones around the p -th harmonics, B_{pq} , Φ_{pq} represents the amplitudes, and γ_{pq} and ε_{pq} denote phases of the q -th sideband in the amplitude and phase modulation functions, respectively. Figure 1b shows the frequency spectrum of the vibration signal obtained from the gearbox under defective conditions; the fault signatures or fault-related components are the harmonics of meshing frequency and sideband frequencies.

3. The Preliminaries

This section provides insight into the methods used in the proposed gearbox fault diagnosis scheme.

3.1. The Proposed Localized Adaptive Denoising Technique

Generally, a vibration signal obtained from a gearbox contains fault-related vibration signatures and noise. Denoising of the signal is required for the extraction of fault-related vibration signatures. Suppose the observed signal is s and the informative signal is x ; then $s = x + \partial$, where ∂ represents the noise. The denoising technique tries to filter out noise for obtaining the estimation signal \hat{x} in a tendency to approximate the useful signal x as much as possible. The adaptive denoising technique uses the concept of destructive interference for denoising of an impaired signal. This technique utilizes the noise-simulated reference signal to access frequency segments in a frequency domain of the observed impaired signal in order to remove noise. The adaptive noise-reducer-based Gaussian reference signal (ANR-GRS), which has been proposed and verified in [22,23], has achieved great performance in reducing noise and avoiding distortion of the fault-related ingredients. In this method, the noise ∂ in a gearbox vibration signal is analyzed and divided into two types of noise: white noise (α) and band noise (β), $\partial = \alpha + \beta$. Then, the reference signal is created by combining two noise-simulated signals, which are analogous with two existing sources of noise in the observed signal, a white noise signal and a Gaussian signal. Moreover, the parameters of the reference signal are adjustable by adaptive algorithm regarding the varying input values of rotation shaft speeds.

The Gaussian signal is responsible for building the simulated noise reference signal. The parameters of the Gaussian signal (a mean value and a standard deviation value) are adaptively adjusted so as to reduce the noise between two consecutive sideband frequencies (the sideband frequency is the gear frequency in this study). The process for generating a reference signal is depicted in Figure 2, and the Gaussian signal is formulated as follows:

$$G_{ns}(k) = \sum_{k=1}^K e^{-\frac{(k-F_m)^2}{2\sigma^2}} \quad (3)$$

where K is the number of sideband segments, and the mean value F_m and the standard deviation value σ are the function of the shaft rotation frequency. Those parameters are adjusted by an optimization process to select the optimized vibration sub-band as an output of the ANR-GRS module [22].

From each parameter set (F_m, σ), which is randomly selected from the specific required range defined in [22], a noise-simulated signal is generated. This reference signal is provided as an input to the adaptive filter along with the impaired observed signal. The adaptive noise filter contains a digital filter, which employs an L-tap FIR type digital filter and weight vector as $w(n) \equiv [w_0, w_1, \dots, w_{L-1}]^T$, and a least mean square adaptive algorithm. The adaptive filter works as follows: The noise-simulated signal is provided as an input to the digital filter, then the filtered output signal is summed with the vibration signal (impaired observed signal) to compute the error signal. This output error signal is provided as a feedback input to the adaptive algorithm to measure its mean square value. Next, the adaptive algorithm tunes the weights of the digital filter according to the converging criterion of least mean square (LMS) error to obtain the optimal weight vector (w_0) and then expose the optimal vibration sub-band corresponding to the particular parameter set. The schematic diagram of the ANR-GRS is provided in Figure 2.

From Figure 2, it can be observed that the ANR-GRS method tries to look for the general optimal parameters of the Gaussian reference signal applied to the whole frequency range of input vibration signals (0–10 kHz).

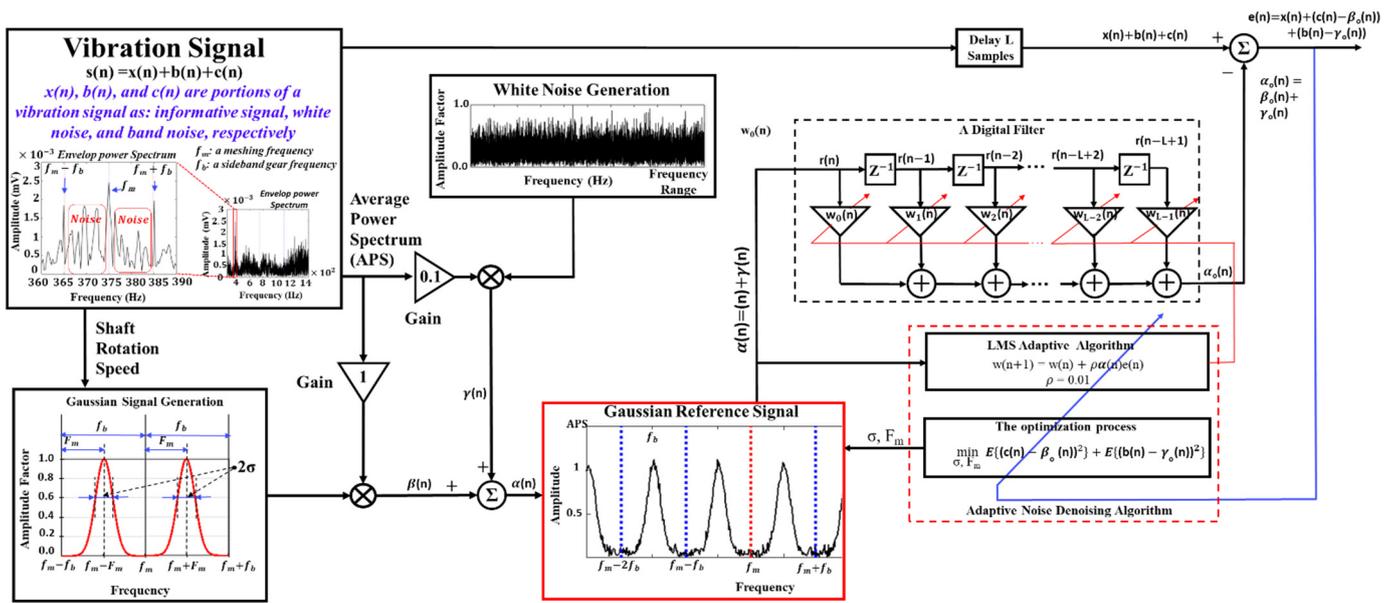


Figure 2. Schematic diagram of the ANR-GRS module.

According to the vibration characteristic of the fault signal presented in Section 2, the frequency domain of the phase–amplitude modulation signals is visualized as a set of many similar frequency segments, each of which contains a meshing frequency harmonic as a center frequency and the sideband gear frequency tones are distributed around the center frequency in the ideal condition. The principal frequency segment (PFS) is defined as a frequency segment with a meshing frequency harmonic as a center frequency and frequency wide equally to a meshing frequency (i.e., the frequency range of PFS is from $(p - 0.5) * f_m$ to $(p + 0.5) * f_m$ with $p * f_m$, a p -th harmonic of a meshing frequency, is a center value). However, in the real world, the amplitudes of frequency tones in each PFS (PFS power distribution) of the gearbox vibration signals are uncorrelated to each other because of the influence of random noise (white noise and band noise) on the non-linear and phase–amplitude modulation signal [48].

Due to the differences of power distributions of PFSs, the general optimal parameter set of Gaussian reference signals cannot be used. Therefore, this paper proposes a new denoising technique called the localized adaptive denoising technique (LADT). The localized adaptive denoising technique adopts the ANR-GRS module from [22]. To improve the denoising capability of ANR-GRS, the LADT applies ANR-GRS to each PFS. By localized adaptive optimization, the new denoising methodology tries to find the localized optimal parameter set of a noise-simulated reference signal, which is appropriate to each specific PFS. The function block diagram of LADT is demonstrated in Figure 3. To implement the ANR-GRS method on each PFS, the band-pass Chebyshev Type-I IIR filter of order 30 [49] is used to segment the frequency spectrums of a vibration signal to M sub-signals whose frequency spectrum is as a PFS. The band-pass filter had a bandwidth similar to meshing frequency, where M is computed as the quotient of the division of the frequency range and the meshing frequency. The localized optimizing process of LADT improves the noise-reducing capability in comparison with that of the ANR-GRS method; therefore, in this study it is used for denoising the vibration signal before the feature engineering process.

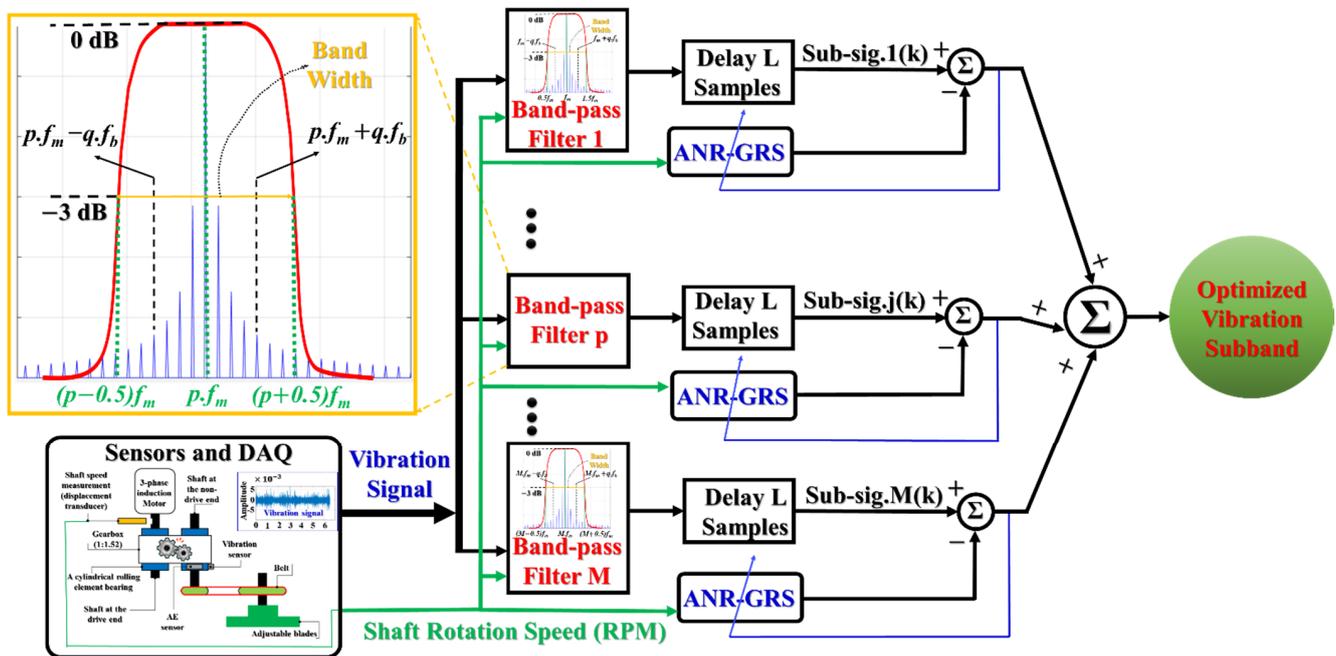


Figure 3. Block diagram of the LADT.

3.2. Wavelet-Based Vibration Imaging (WVI)

To obtain discriminant features from the preprocessed vibration signal, intrinsic information of the vibration signal should be utilized, such that it can provide enough information about MDTF types of defects. For this reason, a proper method that can highlight the key representative elements of MDTF-type defects in gearbox vibration signal is needed. Accordingly, the optimized output sub-bands from the LADT, which contains condensed defect-related useful information, are converted into two-dimensional time-frequency representation images by employing the CWT method; these images are called WVIs. These WVIs, which carry enough fault-related information, are referred to as the enriched feature pool in this paper. The enriched feature pool of the WVIs can be utilized for identifying each defect type of MDTF states (i.e., PC, DT1, DT2, DT3, DT4, DT5, DT6) of the gearbox under variable speed conditions. The process of WVI formation can be explained in detail as follows:

To overcome the limitation of Fourier transform in processing non-linear and non-stationary signals, and the limitation of STFT with fixed timing-window transforming observation, a wavelet approach has been developed. The wavelet transform uses a mother wavelet for decomposing a signal into the spatiotemporal domain. The mother wavelet can be adjusted by expanding or compressing during the transforming process [30]. We denote the wavelet function as $\varphi(t)$, with $\phi(\omega)$ as Fourier transform. To apply the wavelet approach in terms of reversible transform, the admissibility condition must be satisfied:

$$C_\phi = \int_{-\infty}^{\infty} \frac{|\phi(\omega)|^2}{|\omega|} d\omega < \infty, \tag{4}$$

where C_ϕ is the admissibility constant. This (inequality 4) approximates $\phi(\omega) = 0$, which can be presented as:

$$\int_{-\infty}^{\infty} \varphi(t) dt = 0, \tag{5}$$

and this requirement also makes clear that the mother function is a band-pass filter. The term “wavelet” implies a small oscillation wave with the finite length of the window function, and “mother function” can be understood as a prototype function, such as Morlet

wavelet or Daubechies wavelet, whose variants are the wavelet window functions. The actual wavelets are generated from a mother wavelet by the following equation:

$$\varphi_{s,\tau}(t) = |s|^{-\frac{1}{2}} \varphi\left(\frac{t-\tau}{s}\right), \tag{6}$$

where τ is the translation parameter and s represents dilation in Equation (6). The translation parameter represents time in the wavelet domain. The dilation is the inversion of frequency. This scale of wavelet technique is analogous to the scale of map architecture. A large scale in mapping indicates the globalized scenery, and a smaller scale indicates more detail. Similar principles can be applied to the wavelet approach; the high scale (i.e., $s \gg 1$, low frequency) is used for observing the global features of a signal because the wavelets are expanded for extracting the low-frequency components, such as the large time window of STFT. In contrast, the low scale (i.e., high frequencies, $s \ll 1$) is used for observing more details of a signal, called local features. Consider the vibration sub-band $x(t)$ and the given wavelet family $\varphi_{s,\tau}(t)$, the continuous wavelet transform of $x(t) \in L^2(\mathbb{R})$ is calculated [31] by following inner products equation:

$$CWT_x^\varphi(s, \tau) = \langle x, \varphi_{s,\tau} \rangle = |s|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \varphi^*\left(\frac{t-\tau}{s}\right) dt. \tag{7}$$

Equation (7) represents the coefficients of CWT. CWT coefficients are the combination of translation series (time series) and scale (1/frequency) series, which can be utilized for constructing the vibration imaging feature spaces (scalograms). Through the use of the effective denoising technique from the previous process, the vibration image feature pools are filled by condensed fault-related information that qualifies for the next identification step. The combination of the novel denoising technique and the CWT scalogram for the VWI are demonstrated in Figure 4 as the steps involved in the formation of VWI's.

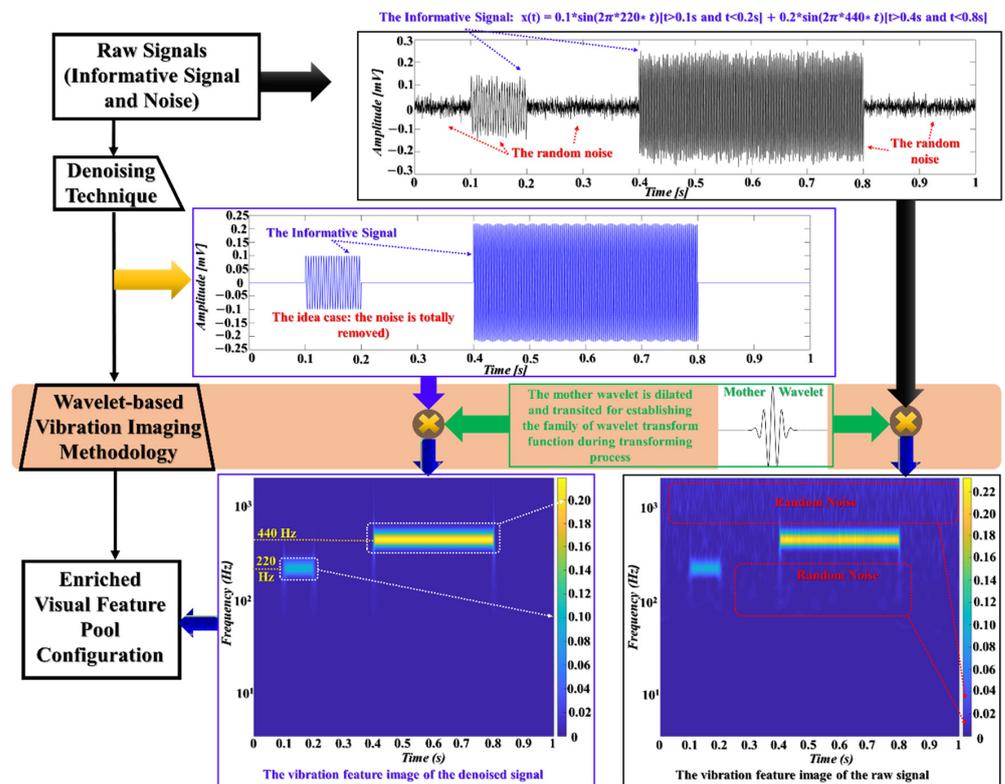


Figure 4. Steps involved in the construction of wavelet-based vibration imaging.

3.3. The Deep Convolutional Neural Network Architecture

DCNA comprises hidden layers (called convolutional layers), pooling layers, and fully connected layers [40,41]. The convolutional layer performs feature extraction from the input image data through a kernel-filter-based convolutional process; then, the pooling layer implements the down-sampling process. The pooling layer helps to reduce computational complexity and to recognize the learned extracted features. In addition, a variety of constraint-optimizing layers, such as rectified linear units (ReLU), dropout, and normalization, are integrated into the DCNA for classification improvement [50]. Afterward, the fully connected layer uses weighted-base wiring to connect the output of the final convolutional or pooling layer for transferring information to the classification layer, which outputs the likelihood decision for classifying the fault types, normally using a SoftMax function [51]. Figure 5 demonstrates the general structure of the DCNA.

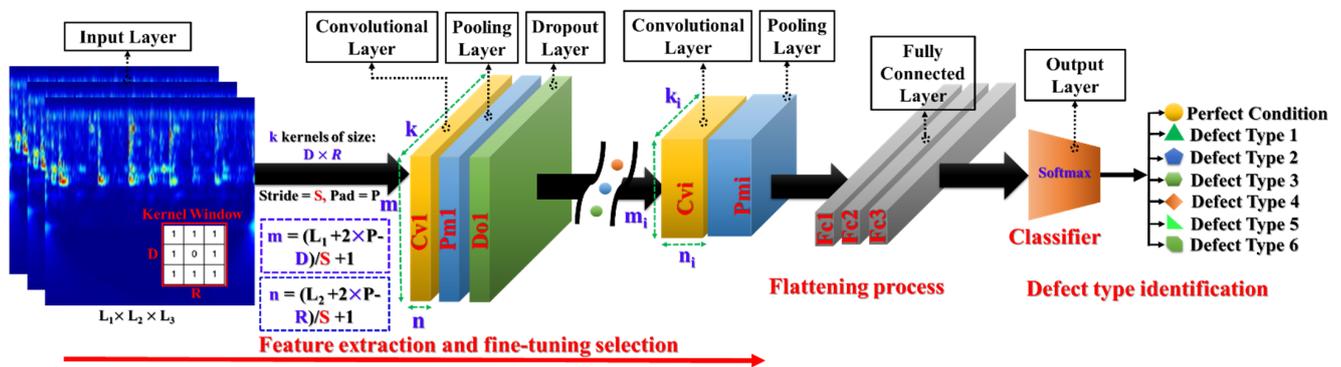


Figure 5. Description of typical DCNA.

The convolutional layer (Cv) is responsible for the latent feature engineering processing. The Cv performs feature mapping through its layers for the extraction of representative attributes from input images that contain key information about gear states. To demonstrate the feature mapping process, we consider two consecutive layers: j^{th} and $(j + 1)^{\text{th}}$ convolutional layers. There are k filters (or kernels), with the size of $m \times n$, which are utilized for extracting features from the output of the j^{th} layer. The output space of the j^{th} layer, with dimensions of $m \times n$, is locally swept to convolve with each filter of $D \times R$ size, using w training weights for adjustment. Then, each result, which corresponds to a single kernel, is added in scale computation with bias b , and functionalized by activation functions of nodes in the $(j + 1)^{\text{th}}$ layer, these are normally non-linear functions, such as the rectified linear unit function (ReLU), used to perform non-linear feature mapping through layers. Assuming that the parameter used in the convolutional calculation is a unity, then feature space with a dimension of $(m - D + 1) \times (n - R + 1)$ is formed corresponding to each filter. In general, the i -th feature mapping space (fms) of the convolutional layer k can be formulated as follows:

$$fms_i^k = RL\left(\sum_{r \in A^{k-1}} fms_r^{k-1} \otimes w_i^k + b_i^k\right), \quad (8)$$

with RL as the ReLU function: $RL(x) = \max(0, x)$.

Where, w_i^k and b_i^k are the sets of weights and bias for the i^{th} filter in layer k , \otimes indicates the convolution operator, A^{k-1} denotes all feature mapping spaces in the $(k - 1)^{\text{th}}$ layer. The feature spaces become more separable as it goes from lower convolutional layer to bottleneck layer network.

Typically, the pooling layer (Pm) is used next to each convolutional layer for the down-sampling process. It scans the whole range of a feature mapping space sequentially, and then applies the pooling operation on a defined pooling region by a non-overlapping searching method. The pooling operation that is most commonly used is the mean average, or maximum value in the defined pooling area [41].

Usually, many incorporated pairs of convolutional and pooling layers are employed in DCNA. After the final convolutional layer or pooling layer, several fully connected layers (Fc) are used to expand deep representation feature mapping spaces, as well as the concatenation of feature mapping spaces into a feature vector. Finally, the represented feature vectors are provided as an input to non-linear nodes for classifying the features into their corresponding categories (the fault states of a gearbox). The SoftMax function is typically used as the final activation function in the classification layer for classifying the input data into their corresponding categories.

The learning process of the DCNA is based on the optimization of the loss function of the reconstruction error. The loss function is the function of the training error, which is the difference between predicted output (\hat{y}_q) and actual output (y_q). It can be presented as follows:

$$e(n) = \frac{1}{2} \sum_{q=1}^K (y_q^n - \hat{y}_q^n)^2 \quad (9)$$

Here, K signifies the number of neurons, and n is the order of repetitive steps. The major purpose of the training process in building the DCNA is to fine-tune its parameters, converging to reduce $E(n)$ through a back-propagation process based on the stochastic gradient descent method [52].

4. The Accurate and Stable MDTF Fault Identification Framework and Its Experimental Evaluation

The key aim of this study was to identify defect types of MDTF gearbox systems under variable speed conditions. As mentioned in Section 1, it has been observed that the existing models might not be able to differentiate those fault types due to the similar behavior of different degrees of tooth fault reflected in the vibration spectrum. To address this issue, in this paper, a new gearbox fault diagnosis scheme has been proposed. Figure 6 provides a block diagram of the proposed framework. From Figure 6, it can be seen that the proposed method consists of four main steps: (1) sensors and data acquisition (DAQ), (2) LADT, (3) WVI, and (4) DCNA. The preliminary section covered the main steps of the proposed method. This section will provide the experimental validation of the proposed method.

4.1. The Gearbox Testbed and Data Acquisition

A gearbox testbed, self-developed at the Ulsan Industrial Artificial Intelligence laboratory, for acquiring vibration data is shown in Figure 7. The testbed can be explained as follows: an AC motor is directly connected to the pinion wheel through the drive shaft (DS), whereas the gear wheel is fixed with a non-drive shaft (NDS) and the adjustable blades (the load). The pinion wheel with 25 teeth, whose length is 9 mm, and the gear wheel (38 teeth) are engaged with each other and housed in the gearbox, creating a gear reduction ratio of 1:1.52. The rotation movement (torque) of the load is provided by the motion of the AC motor through the gearbox. Therefore, the rotational speed of the pinion wheel is equal to the rotational speed of AC motors, and the gear frequency is calculated by the pinion frequency and the gear ratio. The vibration sensor (the accelerometer) is placed at the end of the NDS, 72.5 mm from the gear wheel. The rotational speed of the DS (a pinion frequency) is measured by the displacement transducer, which is mounted for tracking the hole in DS once per revolution. The data acquisition system, which is the PCI-2 data acquisition board, is connected to the accelerometer (622B01) to measure and digitize vibration signals, and to store digital vibration samples. The specifications of the accelerometer, speed sensor, and data acquisition system are given in Table 1.

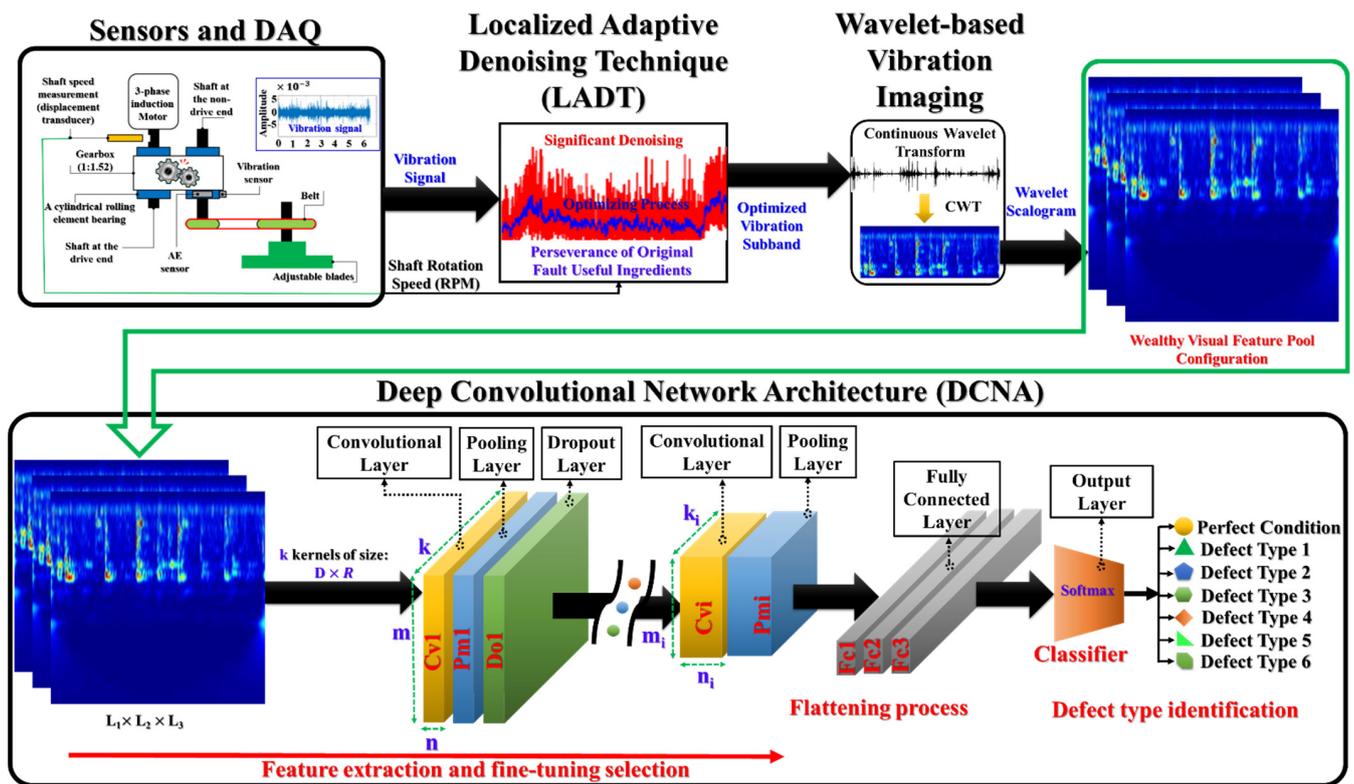


Figure 6. Block diagram of the proposed accurate and stable MDTF gear fault identification framework.

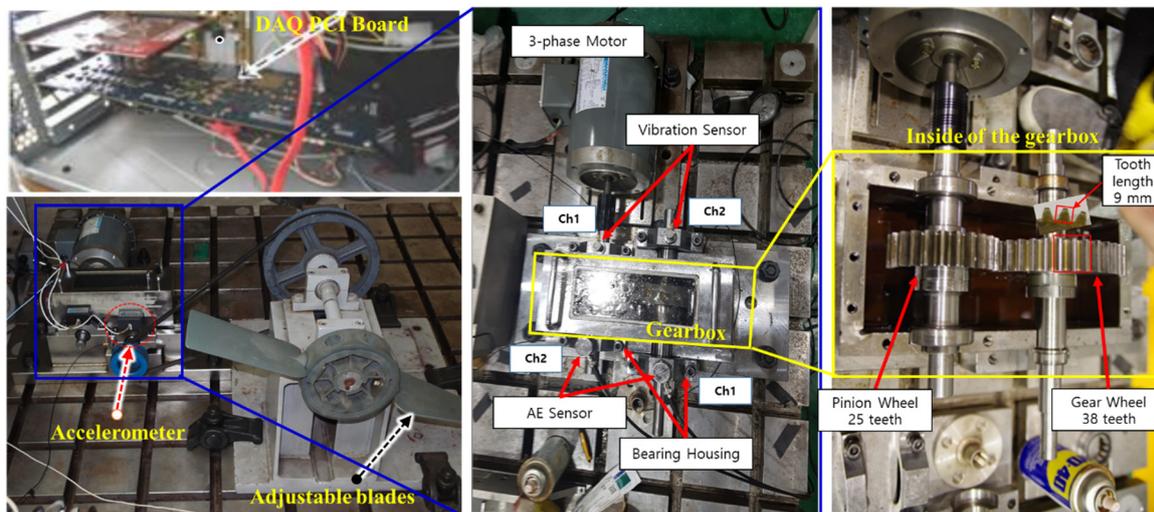


Figure 7. Gearbox experimental testbed.

The MDTF gearbox was created by cutting one tooth, mounted on the gear wheel, to different degrees. Figure 8 shows the degrees of cut teeth and the vibration signals obtained under each condition for all observed defect types in this study, including a normal or perfect condition gear (PC), 6.6% degree of tooth defect (DT1), 10% degree of tooth defect (DT2), 20% degree of tooth defect (DT3), 30% degree of tooth defect (DT4), 40% degree of tooth defect (DT5), and 50% degree of tooth defect (DT6). These multiple degree tooth faults were seeded for simulation of the same behavior of the gear defects caused by long-term operation of a gearbox system (e.g., tooth spalling, tooth cracking, worn tooth, etc.). The vibration characteristic for fault states of a gearbox was analyzed in detail in Section 2.

Table 1. Specifications of the sensors and data acquisition system.

Devices	Specification
Vibration sensor (Accelerometer 622B01)	Sensitivity (V/g): 10.2 mV/(m/s ²)
	Operational frequency range: 0.42 to 10 kHz
	Resonant frequency: 30 kHz
	Measurement range: ± 490 m/s ²
4-Channel DAQ PCI Board	18-bit 40 MHz AD conversion, a sampling frequency of 65.536 kHz is used for each of two channels simultaneously
Displacement transducer	Distance from the head of a transducer to a hole: 1.0 mm
	Diameter of a hole: 12.80 mm
	Sensitivity: 0 to -3 dB
	Frequency response: 0–10 kHz

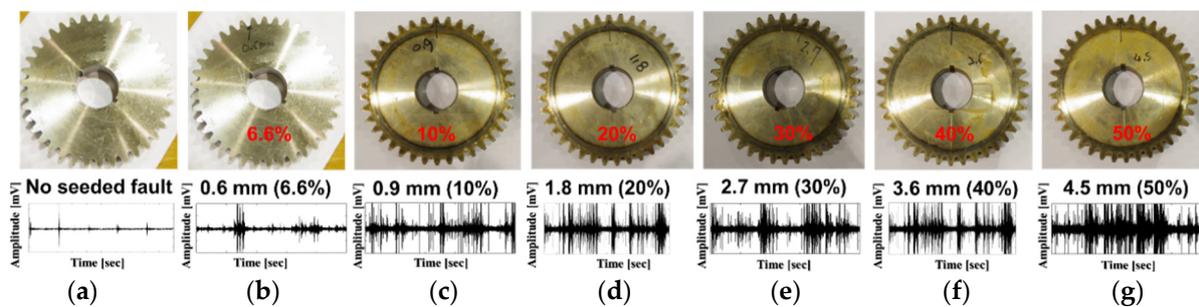
**Figure 8.** The observed defect types of the multi-degree tooth faults on the gear wheel and examples of vibration signals at 600 RPM: (a) PC, (b) DT1, (c) DT2, (d) DT3, (e) DT4, (f) DT5, and (g) DT6, respectively.

Table 2 demonstrates the configuration of the dataset used in this paper. The data acquisition system converts the analog vibration signal to a digital vibration signal with a sampling frequency of 65,536 Hz. Each sample is one second long, termed a one-sec sample. A total of 200 samples were collected under each defect condition with variable rotating speed (four shaft rotational speeds are evaluated in this study). Therefore, there are 800 samples for each defect condition, and a total of 5600 samples are extracted from this testbed.

4.2. LADT Performance for Effective Noise Removal of Vibration Signals of a MDTF Gearbox under Variable Speed Conditions

The raw vibration signals were digitized at a high sampling frequency of 65,536 Hz in order to gather rich discrete vibration samples, and to capture the extent of feasible defect-related components in each one-sec vibration signal. The vibration data collected from the gearbox contain fault-related information and interference noise. By sampling the vibration signal at a frequency of 65,536 Hz, the frequency spectrum of a discrete vibration sample is from 0 Hz to 32,768 Hz (according to the Nyquist–Shannon sampling theorem). However, the accelerometer is capable of sensing the vibration oscillations in the frequency range of 0.42–10,000 Hz (Table 1). Thus, the fault-related information is in the frequency range of 0.42–10,000 Hz. Therefore, rather than providing the raw vibration signal to LADT, the vibration signal is pre-processed by performing down-sampling using a low-pass filter to avoid aliasing [22]. After performing down-sampling, the vibration sub-bands are obtained, which have the time length of one second, the sampling frequency of 21,845 Hz (65,536/3), and frequency range from 0–10,922 Hz.

Table 2. The configuration of the MDTF gearbox dataset.

Gearbox Defect Type	Description	Number of One-Second Samples for Specific Rotation Speed (RPM)				Sampling Frequency (Hz)
		300	600	900	1200	
Perfect Condition (PC)	Normal or perfect gearbox	200	200	200	200	65,536
Defect Type 1 (DT1)	6.6% degree of tooth defect (0.6 mm/9 mm)	200	200	200	200	65,536
Defect Type 2 (DT2)	10% degree of tooth defect (0.9 mm/9 mm)	200	200	200	200	65,536
Defect Type 3 (DT3)	20% degree of tooth defect (1.8 mm/9 mm)	200	200	200	200	65,536
Defect Type 4 (DT4)	30% degree of tooth defect (2.7 mm/9 mm)	200	200	200	200	65,536
Defect Type 5 (DT5)	40% degree of tooth defect (3.6 mm/9 mm)	200	200	200	200	65,536
Defect Type 6 (DT6)	50% degree of tooth defect (4.5 mm/9 mm)	200	200	200	200	65,536

The vibration sub-bands are provided as an input to LADT for reducing the noise and enhancing the useful fault-related information, which represents multi-degree tooth fault behaviors. LADT applies ANR-GRS to each PFS. Through localized adaptive optimization, the new denoising methodology tries to find the localized optimal parameter set of the noise-simulated reference signal, which is appropriate to each specific PFS. The outputs of LADT are the optimized vibration sub-bands, which maintain the original defect-related frequency tones (meshing frequency harmonics and sideband frequencies) and reduced background noise. Those defect representative ingredients are key factors for identifying the defect types of an MDTF gearbox under the condition of the variable speed. Those fault types proceed as analogous behaviors reflecting the vibration characteristic, the differences between them might be degrees of amplitudes of informative tones, their proportions, or occurrence events in the tiny range of separation. Thus, the image-based enriching feature pool configuration methods are needed to sort them out for condensation.

4.3. Wealthy Feature Pool Configuration Based on VWI

In this step, a continuous wavelet transform is applied to the noise-free optimized sub-bands obtained from the LADT. The wavelet-based transforming method is used to convert time-domain optimized sub-bands to the scalograms for the enriched visualized features pool. The CWT method in this paper employs the Morlet wavelet, which is the most effective technique in the fault diagnosis approach of the rotation machines [53], as a mother wavelet function. The signal is decomposed up to 16 octaves. Based on experiments, the optimal value of voices per octave parameter was chosen as 16. The wavelet coefficients for each input sub-band, which are derived from applying the wavelet family functions in Formula (5), are used to obtain a scalogram, which is an energy distribution map of the input sub-band on a time–frequency scale. Those scalogram images of the vibration sub-bands are reshaped by the size of $224 \times 224 \times 3$ for compatibility with the input layer of DCNA in the next classification step and packed to configure the enriched visual image feature pool.

4.4. DCNA Construction

In this study, the contents of the enriched visualized feature pool, which are called WVIs are obtained from the CWT of low-noise optimized vibration sub-bands, are provided as an input to DCNA. The WVI contains fault-related information in the form of edges, lines, curves, spots, or pixels with various intensities (which are represented by the R, G, and B channels of the RGB image). The DCNA is used primarily to recognize images. Figure 9 demonstrates the architecture of the proposed DCNA used in this study. The proposed DCNA has fifteen layers, including five convolutional layers (Cv), three pooling layers (Pm), two drop-out layers (Do), three fully connected layers (Fc), one input layer, and one terminal output layer (Os). The DCNA makes a start with an input layer of size $224 \times 224 \times 3$, according to the size of RGB images (224×224 indicates the values of length and width, and 3 denotes three R, G, B channels of the input image). Next, the features are extracted from fault-related images by the first convolutional layer with 96 kernels of size $11 \times 11 \times 3$ and the stride of 4. The results of the first convolutional calculation are feature spaces of size $54 \times 54 \times 96$. After the first convolution layer (Cv1), the max-pooling layer (Pm1) is applied for down-sampling. Moreover, the drop-out layer (Do1) is located in series to extenuate the over-fitting issue [50]. The second convolutional layer has 256 filters of size $5 \times 5 \times 48$, and it is followed by pooling and dropout processing layers. The Cv3 and Cv4 layers consist of 384 filters with a size of $3 \times 3 \times 256$ and 384 kernels with a size of $3 \times 3 \times 384$, respectively. Next, Cv5 is down-sampled by the third max-pooling layer (Pm3), composed of 256 of $3 \times 3 \times 384$ kernels. All of the max-pooling layers employ 3×3 filters with a stride of 2. The output of the third max-pooling is used as an input to the fully connected layers (Fc1, Fc2, Fc3). Fc1 tries to implement a flattening process to convert all feature matrices ($6 \times 6 \times 256$) from the output of layer Pm3 to the feature vectors ($1 \times 1 \times 4096$) through its operation as a weighted sum with bias terms. These output feature vectors then are passed through the activation function ReLU and input to the next layer (Fc2). The second fully connected layer, which is the penultimate layer, includes 1000 neurons and functions, similarly to Fc1, to output feature vectors of size $1 \times 1 \times 1000$. The last flattened layer, Fc3, including 7 neurons, which are the SoftMax activation functions, is the classification layer. It operates at a terminal spot of the DCNA for estimation of the probabilities of the categories.

In this paper, the fifteen-layer DCNA has been conducted based on the original AlexNet architecture [54], with some modifications for this specific application. The AlexNet model has already achieved better feasibility than other models for recognizing images. This model has implemented training for 1.2 million high-resolution pictures of ImageNet for classification of up to 1000 differential species targets in the contest of LSVRC-2010 by training of 650 thousand neurons and 60 million parameters, with many optimizing processes in the network architecture. In our research, we have replaced two normalization layers with two drop-out layers in order to improve the capability of over-fitting avoidance [50,55]. Moreover, the last fully connected layers (Fc3), which include 1000 neurons from the original AlexNet, are replaced by the same fully connected layers with a reduced number of neurons (7), for suitable application in our research with seven classifying categories. The detailed description of the proposed DCNA is shown in Table 3.

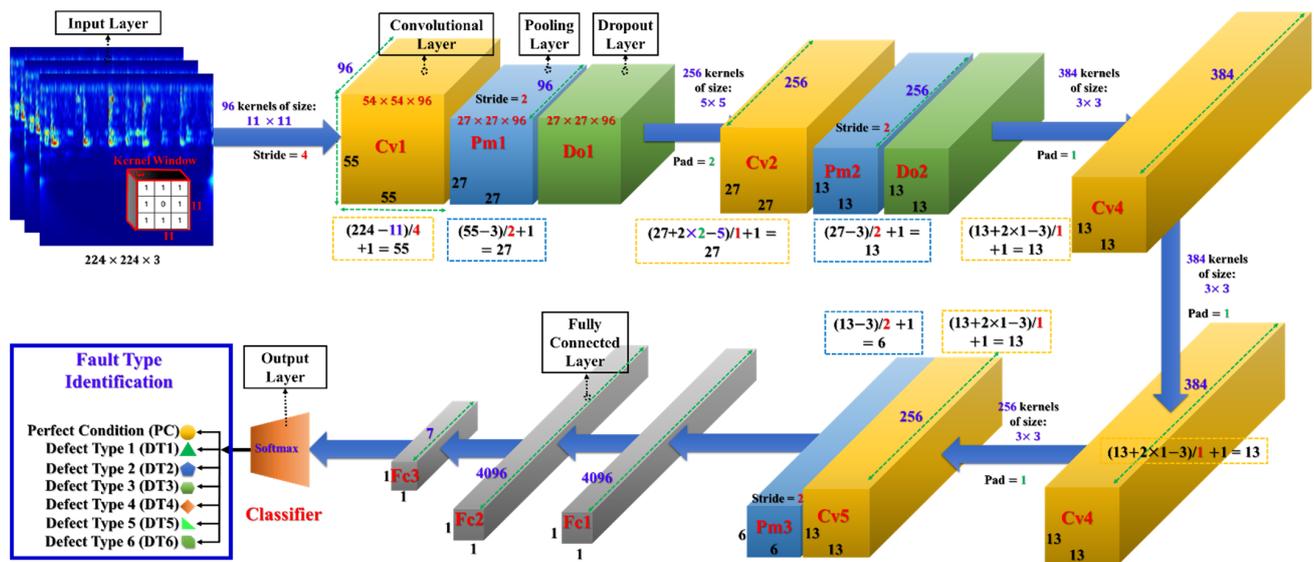


Figure 9. The applied DCNA model for implementing the fault-type identification in this study.

Table 3. The structural elements of the proposed DCNA.

Layers	Operating Parameters	Number of Kernels	Kernel Size	Stride	Padding
Input layer	$224 \times 224 \times 3$				
1st Convolutional (Cv1)	$55 \times 55 \times 96$	96	11×11	4	0
1st Max Pooling (Pm1)	$27 \times 27 \times 96$	96	3×3	2	0
1st Dropout (Do1)	$27 \times 27 \times 96$				
2nd Convolutional (Cv2)	$27 \times 27 \times 256$	256	5×5	1	2
2nd Max Pooling (Pm2)	$13 \times 13 \times 256$	256	3×3	2	0
2nd Dropout (Do2)	$13 \times 13 \times 256$				
3rd Convolutional (Cv3)	$13 \times 13 \times 384$	384	3×3	1	1
4th Convolutional (Cv4)	$13 \times 13 \times 384$	384	3×3	1	1
5th Convolutional (Cv5)	$13 \times 13 \times 256$	256	3×3	1	1
3rd Max Pooling (Pm3)	$6 \times 6 \times 256$	256	3×3	2	
1st Fully Connected (Fc1)	$1 \times 1 \times 4096$				
2nd Fully Connected (Fc2)	$1 \times 1 \times 4096$				
3rd Fully Connected (Fc3)	$1 \times 1 \times 7$				
Output	SoftMax Nodes				

4.5. The Experimental Classification for an MDTF Gearbox under Variable Speed Conditions

The DCNA performs a fault-classifying process based on the input WVI imaging data for the MDTF gearbox under varying speed conditions. To verify the performance of the proposed DCNA for identifying seven MDTF fault types under varying speed conditions, we conducted an experimental setup of two scenarios, as shown in Table 4. In Scenario 1, all vibration data for four speeds were observed for classification. While in Scenario 2, four experiments were performed based on varying speed-related data. The configuration of the testing and training datasets for both of the scenarios is given in Table 4.

Table 4. Description of the dataset for training and testing with RPM in the experiment setup.

Scenarios	The Experiment	Number of Samples	The RPM of Data Samples
Scenario 1	Experiment 0	Training samples: 3360	60% of All four speeds dataset
		Testing samples: 2240	40% of All four speeds dataset
Scenario 2	Experiment 1	Training samples: 2800	The shaft speeds: 300 RPM, 600 RPM
		Testing samples: 1400	The shaft speed: 900 RPM
	Experiment 2	Training samples: 2800	The shaft speeds: 600 RPM, 900 RPM
		Testing samples: 1400	The shaft speed: 1200 RPM
	Experiment 3	Training samples: 2800	The shaft speeds: 900 RPM, 1200 RPM
		Testing samples: 1400	The shaft speed: 300 RPM
	Experiment 4	Training sample: 2800	The shaft speeds: 1200 RPM, 300 RPM
		Testing samples: 1400	The shaft speed: 600 RPM

For each speed (a total of four speeds: 300 RPM, 600 RPM, 900 RPM, and 1200 RPM), there were a total of 1400 one-second samples for all gear fault types (there were seven defect types or seven categories, PC, DT1, DT2, DT3, DT4, DT5, DT6, and each of them was acquired by sampling for one second, repeated 200 times, to achieve 200 one-second samples). All these samples were first preprocessed using LADT. Next, the output optimized sub-bands obtained from LADT were converted by the CWT method to attain the enriched feature scalogram images. That speed-related image subset was used as input data for the DCNA. For each experiment, we used two speed-related datasets (2800 samples) to train the proposed DCNA several times with multitudinous epochs, targeted to optimize the network parameters based on minimizing orientation of the loss function (Equation (9)), and the dataset of another speed (1400 samples) was used as the testing dataset of the constructed model. These processes were circularly acted based on four speed-related datasets to conduct all four experiments.

5. Results and Discussion

This section principally validates the proposed fault identification framework constructed in Section 4 for an MDTF gearbox under inconsistent rotational speeds based on the data collected from a real-world testing platform. The effectiveness of this model is entirely evaluated based on the following operations: LADT, visual enriching feature configuration (WVI's), and fault identification based on DCNA.

5.1. Experimental Verification of the Effective Performance of LADT and Wealthy Feature Pool Configuration Created by WVI

As explained in the introduction section, the real-world gearbox vibration signals originally contain informative components and random background noise. The disturbance noises appear randomly, and they can affect the informative components. Thus, in the raw form of the vibration signal, it is very difficult to separate the original informative components from the background noise. Furthermore, the operation behaviors of MDTF gear faults reflected in the vibration signal are too similar. In other words, to discriminate these kinds of faults, the use of enhanced techniques is required. The LADT approach is the key technique of this study for effective noise cancellation and for separating the original fault-related components from the high noise vibration signals. Before being fed to LADT, the raw vibration signals, gathered from the experimental gearbox testbed, were processed by down-sampling and low-pass filtering to attain vibration signals with the frequency range of 0–10,922 Hz, according to the real frequency working range of the acceleration sensor for removing the redundancy fractions. These output signals are named raw-filtered vibration signals. LADT tries to divide each raw-filtered vibration signal into many sub-signals, so that their frequency spectrums are principal frequency segments, by

applying the series of the non-overlapped band-pass filters along the frequency spectrum of the vibrations signal (0–10,922 Hz). Next, the ANR-GRS technique [22] is applied on each principle frequency segment to achieve a locally optimized sub-band from each input sub-signal based on the localized optimal parameters. The final optimized output of the LADT module is a summation of all locally optimized sub-bands corresponding to each input vibration signal.

The visual analysis of frequency spectrums of three vibration signals (a raw-filtered vibration signal, the output signals of the ANR-GRS module, and optimized sub-bands from the LADT module) are illustrated in Figure 10.

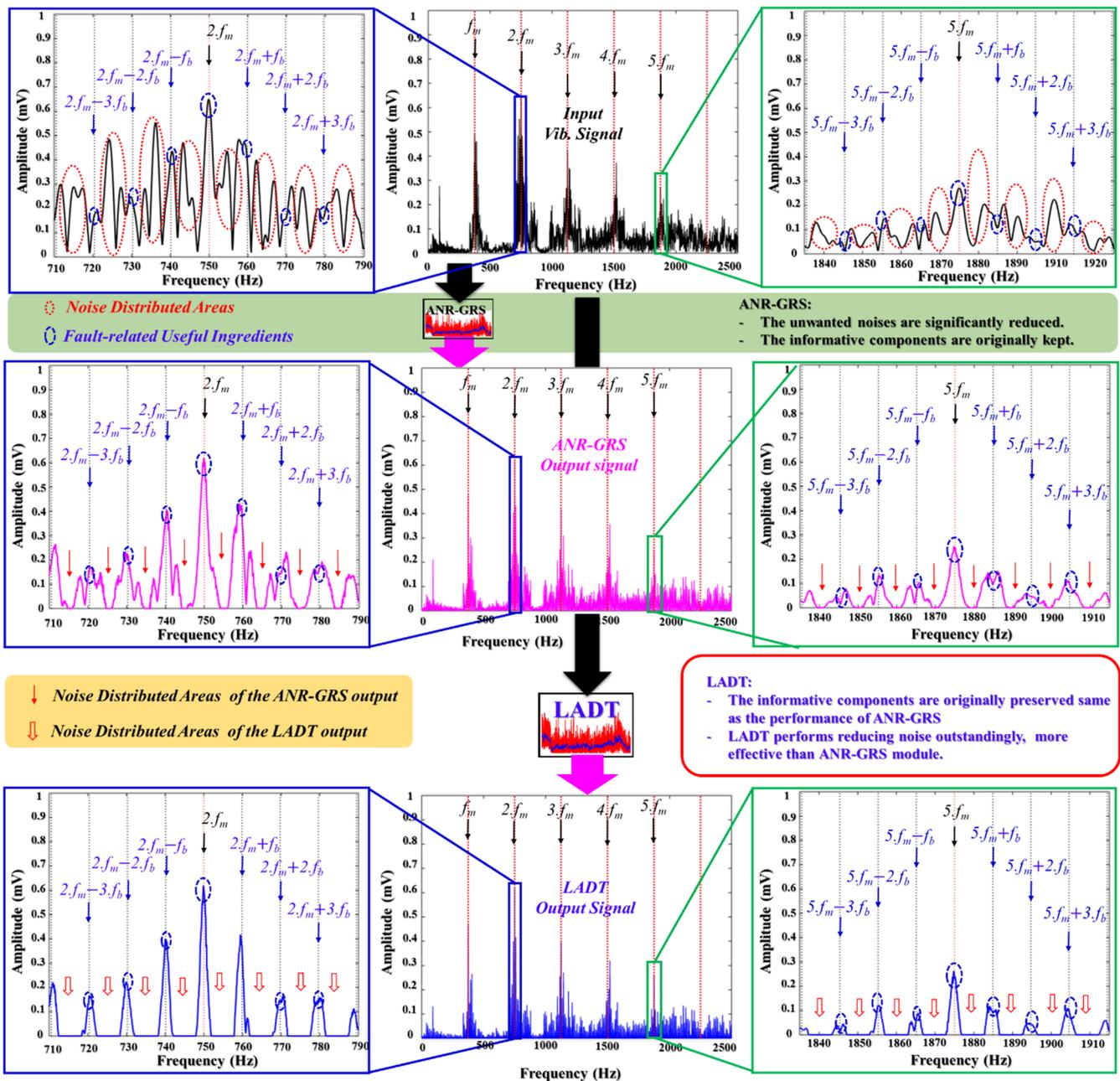


Figure 10. The frequency spectrum analysis of the input and output signals of LADT in comparison with the performance of ANR-GRS for an example vibration signal of DT3 at 900 RPM.

As shown in Figure 10, the superiority of the localized adaptive process of the LADT module for denoising is proven. Here, the noise disturbance areas, which were circled by red dotted lines in the spectrum of a raw-filtered vibration signal that inputs to ANR-GRS and LADT modules, were mostly removed in the output of the ANR-GRS module and LADT module (the spaces with red narrows in the output spectra of ANR-GRS and LADT modules). However, the output signal of LADT indicated outstanding efficiency in reducing noise relative to the ANR-GRS module; the noise areas of the second and fifth principal frequency segments (the segment contains the second and fifth harmonics of the meshing frequency) in the output sub-band of the LADT module were much lower than those in the output signal of ANR-GRS. This verifies the effectiveness of the localized adaptive optimization process of the LADT scheme. In addition, the fault-related components, which were marked by blue-dotted circles in the input and output of LADT, were exactly the same. In other words, the LADT approach reduces noise in the largest amount possible by obeying the principal rule of a condition-monitoring fault diagnosis system to preserve the original fault-informative elements, such as sideband frequency tones and meshing frequency harmonics inside of the raw vibration signals.

The output optimized sub-bands from LADT were then converted to visualized feature spaces, for better expression of defect-related components induced by vibration characteristic of MDTF defect types in the time–frequency domain, using the proposed WVI method. Similarly to the example signal in Figure 4 (Section 3.2), the wavelet-based vibration images carried the defect-correlated factorials and exposed the attributes through color images. Figure 11 demonstrates the scalograms of the seven defect types of a gearbox under four rotational speeds. Through visualization, the scalograms of the same defect type under four rotational speeds showed the proximate parallel zones with the different energy levels. In addition, the energy of the useful components (pixel illuminations) has grown according to the uptrend of rotational speeds. Those discriminate notifications were quantized in the massive process of feature extraction and optimization achieved from DCNA performance.

5.2. DCNA-Based Identification Performance Analysis

By applying the LADT method, the noise components of the vibration signals were mostly removed. The wealthy feature pool configuration based on CWT, then, translated the output of LADT as insignificant-noise vibration sub-bands to the scalogram images. These scalogram images carried enough information for fault discrimination. The wavelet-based vibration image datasets were used as input datasets for DCNA for the classification task. First, the proposed network tried to perform Scenario 1 to discover the effect of the quantity of input data on the time consumption and classification accuracy. The dataset, which contained all four speeds and seven categories, was randomly split into the training set and validation set. Each input sample was a colorized image with dimensions of $224 \times 224 \times 3$, which met the demand of the input layer size of the proposed DCNA. From the numerous proportions of the training set, the computational consumptions and accuracies are listed in Table 5. It shows that when 50% to 60% of total samples were used for training, the best performances were obtained (by high accuracies in the acceptable time consumption) in the observed quantities. Thus, a ratio of 60% was used in this study.

In Scenario 2, four experiments (in Table 4) were executed in this study to analyze the accuracy and reliability of the proposed framework for an MDTF gearbox under differential speed conditions. In each of the four experiments, the training dataset was composed of two different speed samples (2800 samples), and the data samples of the validation set contained samples collected at speeds that differed from that of the training dataset (1400 samples). Following Scenario 2, the speed-varying datasets were alternately used for the training and testing process over a total of four observed rotational speeds in this paper. The learned features of the activation processes in different layers of the applied network model can be seen in Figure 12. From the input RGB image of the defect type 3 with a speed of 600 RPM (Figure 12a), through the beginning steps of the high-dimensional feature extraction process, performed by 96 kernel filters (Figure 12b) of the first convolutional layer (Cv1), the feature images of the Cv1 of one channel are shown in Figure 12c. With the help of this process, the one time-frequency domain vibration image is mapped to 96 feature images for observing the defect-related elements in high-dimensional feature spaces. Next, the several mapping values in feature images are reduced by the max-pooling layer (Pm1) as shown in Figure 12d. Thus, the feature image in Figure 12d is inclined to be viewed more dubiously and softly than Figure 12c. From Figure 12e to Figure 12h, the complex learned feature images from Cv2 to Cv5 of an example channel are demonstrated the impacts of the kernels of those layers. After flowing through Cv and Pm layers of the applied DCNA network, the learned feature maps were flattened as feature vectors. Those feature vectors, which were outputs of the final fully connected layer (Fc3), were then used as input of a SoftMax layer or output layer for clustering.

The t-SNE (t-stochastic neighbor embedding) approach is popular in deep networks for exploring the feature spaces. Figure 13 depicts the three-dimensional distribution of the output feature vectors from the Fc3 layer according to seven defect categories through four experiments. As shown in Figure 13, the samples of the same defect type were close to each other, separate from the samples of another defect type. The clear discrimination between defect types verifies the high accuracy and stable capability of the proposed framework through the condition of the inconsistent speed. Based on this, the classification process can identify the defect types of an MDTF gearbox more easily.

Moreover, the confusion matrix, which is shown in Figure 14, provided perfect performance (100% accuracy) of fault identification for seven defect types of the experimental MDTF gearbox under variable speed conditions through the four experiments in Scenario 2.

Table 5. The classification accuracy and time consumption for various size of the training set.

Training Size (Percentages of 5600 Samples)	Number of Epochs	Time Consumption (Second)	Overall Classification Result (%)
1680 Samples (30%)	160	105.101	89.51
2240 Samples (40%)	200	138.276	94.63
2800 Samples (50%)	210	147.846	99.79
3360 Samples (60%)	250	165.569	100
3920 Samples (70%)	300	375.497	100
4480 Samples (80%)	360	458.990	100
5040 Samples (90%)	410	546.832	100

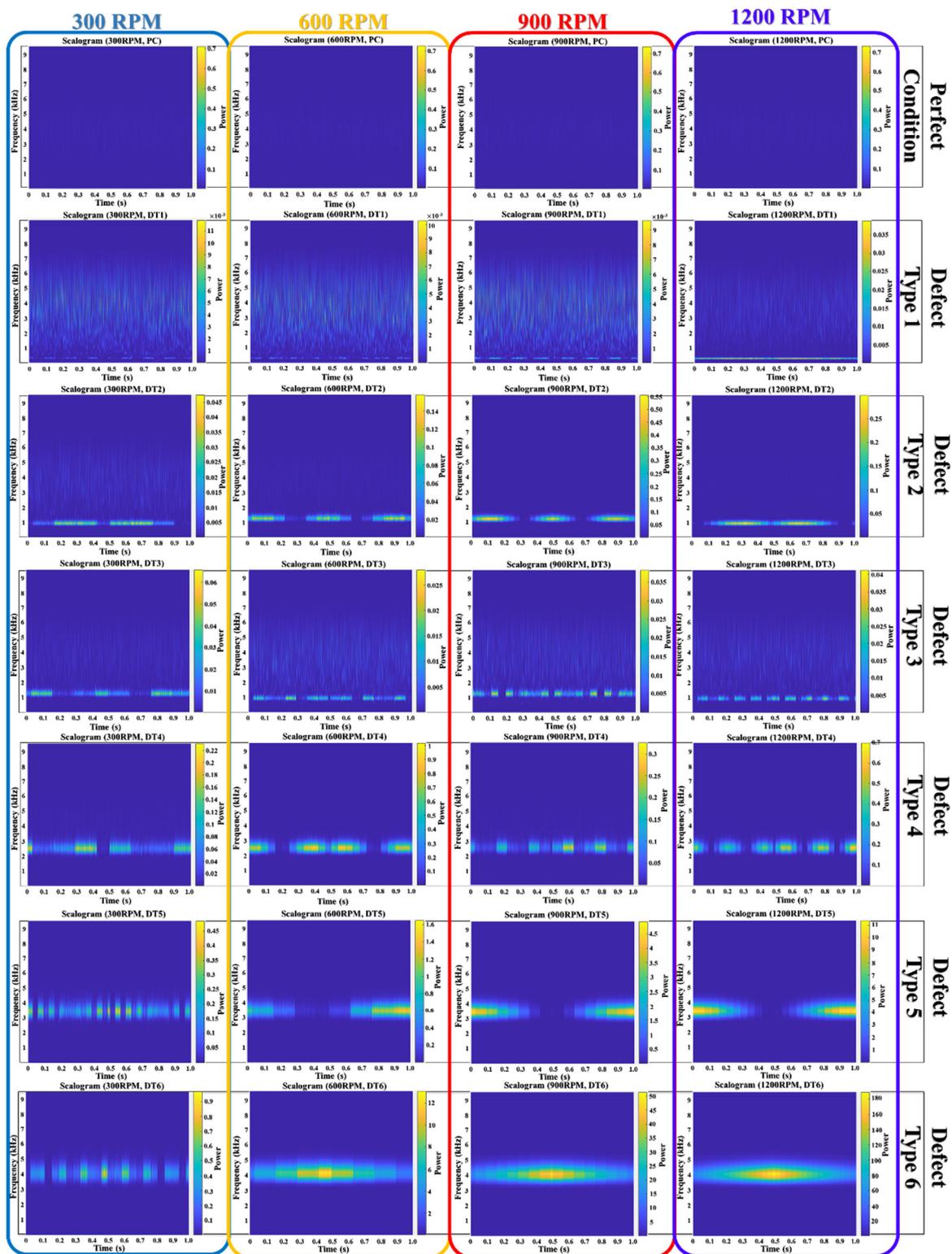


Figure 11. Frequency spectrum analysis of the vibration sub-band (for fault state D2 at 900 RPM) in the comparison between an input and output sub-band of the ANC module.

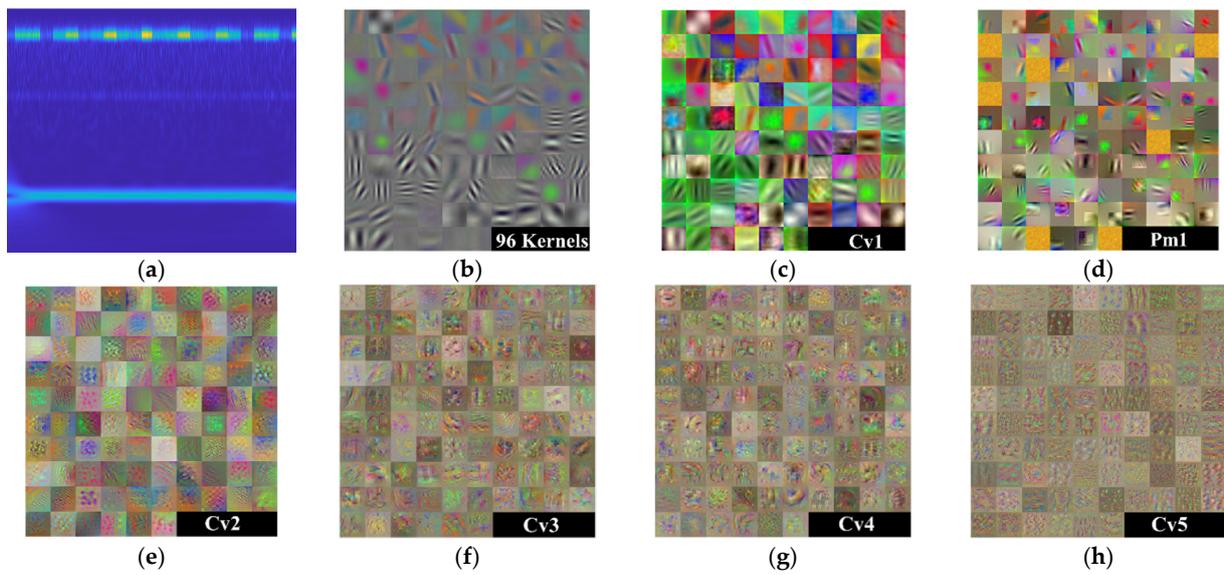


Figure 12. The flowing learned feature images through layers of the proposed DCNA for one example channel, here, (a) RGB input image, (b) the 96 kernels of size 11×11 (c), the feature images of the Cv1, (d) the feature images of the Pm1, (e) the feature images of the Cv2, (f) the feature images of the Cv3, (g) the feature images of the Cv4, the feature images of the Cv5 (h).

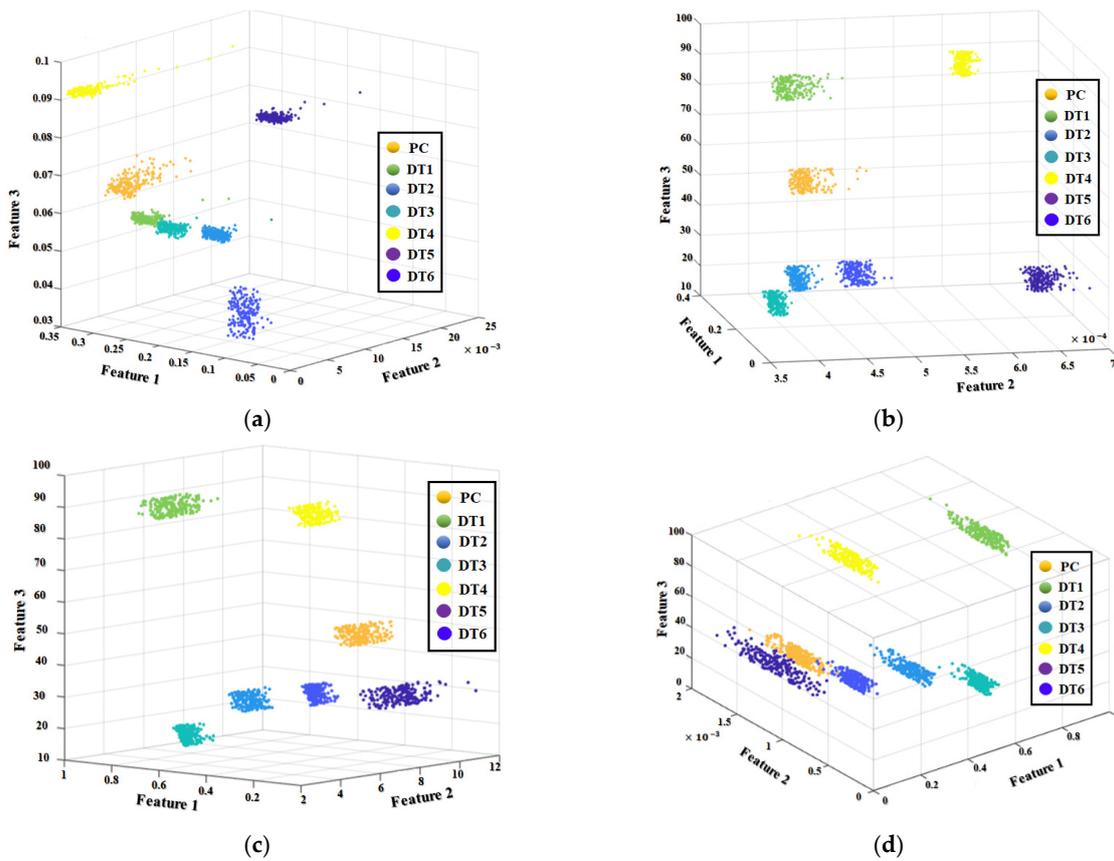


Figure 13. Three-dimensional clustering spaces of the four experiments: (a) Experiment 1, (b) Experiment 2, (c) Experiment 3, (d) Experiment 4.

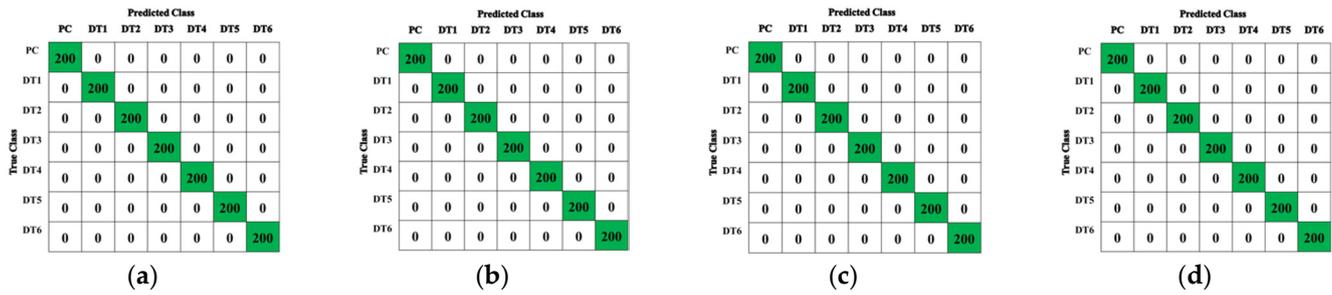


Figure 14. The confusion matrices of the experimental scenario 2: (a) Experiment 1, (b) Experiment 2, (c) Experiment 3, (d) Experiment 4.

For robustness analysis of the proposed methodology, a comparison was made between the proposed method and existing state-of-the-art methods such as ANR-GRS + SFE + GA + KNN (Fw1), LADT + GA + KNN (Fw2), LADT + FSE + SVM (Fw3), ANR-GRS + CWT + DCNA (Fw4), LADT + STHT + DCNA (Fw5). Those are explained in detail as follows:

- (1) ANR-GRS + SFE + GA + KNN (Fw1): This framework used the denoising method as an adaptive noise-reducer-based Gaussian reference for optimizing vibration signals. Next, the handcraft feature extraction technique was used to extract the statistical features in the time and frequency domain (SFE: statistical feature extraction). The achieved feature pool, then, was processed by a feature-selection-method-based genetic algorithm (GA) to fetch the most discriminate features in preparation for input into the learning model as k-nearest neighbors (KNN). KNN performed fault classification based on the selected features (reduced dimensionality) to identify the gear defect types for validating the accuracy of the constructed model (Fw1). The details of Fw1 can be found in [56].
- (2) LADT + GA + KNN (Fw2): To validate the improved denoising technique, the LADT module was used instead of the ANR-GRS module in the Fw1 to construct the Fw2.
- (3) LADT + FSE + SVM (Fw3): This observed framework was created to explore the noise reduction proficiency of LADT, incorporating the high-dimensional feature pool, which can be well-classified by a support vector machine (SVM). The proposed denoising approach (LADT) in this study was applied to optimize vibration signals. The FSE step tried to configure the feature pool. Then, an SVM was utilized to process fault diagnosis by using the extracted features to input learning data [22].
- (4) ANR-GRS + CWT + DCNA (Fw4): By implementing this framework, the effectiveness of the LADT module was straightforwardly compared to the initial adaptive noise technique (ANR-GRS). In this situation, we only replaced the LADT module with ANR-GRS.
- (5) LADT + STHT + DCNA (Fw5): This framework was implemented by using short-time Fourier transform (STHT) to extract the visualized image features as spectrogram images. It was used for comparison with the proposed scheme in the process of enriching feature extraction.

Those methodologies were selected to evaluate the performance of the proposed method in terms of the improvement of LADT for denoising in comparison with the initial method (ANR-GRS), the effective performance between the automatic feature engineering methods (feature extraction, feature selection, and classification) based on DNN from the enriched feature pool (CWT + DCNA), handcraft-method-aided shallow neural networks (SFE + GA + KNN, SFE + SVM), and the effect of enriching feature pool configuration methods (CWT and STHT).

To evaluate the proposed method against the reference methods, the overall classification accuracy (R_f) for each framework was calculated using Equation (10).

$$R_f = \frac{\sum TP}{\sum TS} \cdot 100\% \quad (10)$$

where, $\sum TP$ denotes the summation of the true positives and $\sum TS$ refers to the total number of samples used in the classifying process. Each framework was executed several times to achieve the average results of overall classifying accuracies for seven defect types. The classification results of all frameworks through two scenarios are shown in Table 6.

Table 6. The overall identification accuracies of the compared frameworks through two scenarios.

Scenarios	Experiment	Average Classification Accuracies of Frameworks (%)					
		Fw1	Fw2	Fw3	Fw4	Fw5	Proposed Fw
Scenario 1	Experiment 0	62.18	65.13	54.71	83.50	91.68	100
	Experiment 1	53.43	57.35	72.54	86.65	88.82	100
Scenario 2	Experiment 2	45.31	51.43	68.78	81.51	86.85	100
	Experiment 3	57.62	67.71	74.19	88.30	90.21	100
	Experiment 4	50.17	58.69	72.70	85.90	89.49	100

As can be seen from Table 6, the LADT approach performed denoising better than the ANR-GRS method in the three frameworks: Fw1, Fw2, Fw3; however, the identification accuracy results were lesser than the proposed method from 54.69% to 25.81% due to the limitations of those frameworks in engaging with handcraft feature extraction and shallow learning networks. The different results (from 12.7% to 18.49%) between Fw4 and the proposed framework in this paper confirm the high improvement in denoising delicacy of LADT. The Fw5 results (from 13.15% to 8.32% as lower) demonstrate that the wavelet-based vibration imaging to configure the wealth feature pool achieved a better performance than using STFT. By comparative analysis, the applied framework in this paper outperformed the defect type identification for an MDTF gearbox under variable speeds condition in comparison with those state-of-the-art frameworks, yielding an average classification performance of 100% during two scenarios.

To establish an accurate fault identification framework, an effective denoising technique for the complex gearbox vibration signals is critically needed. The disturbance noises in the vibration signals make the subsequent processes of feature engineering and classification less effective. Therefore, this paper combined LADT for highly effective denoising, VWI for wealthy visual feature pool configuration, and DCNA for high dimensional and automated feature extraction, feature-optimizing selection and classification, and to draw the accurate and stable fault identification framework for an MDTF gearbox under inconsistent speed conditions. Through analysis and experimentation, our proposed methodology achieved the highest classification result, verifying the effectiveness of the proposed model.

6. Conclusions

This paper proposed an accurate and stable fault diagnosis framework for multi-degree tooth faults in the gearbox under variable speed conditions. The raw vibration signal obtained from the gearbox contains fault-related information and background noises. To obtain information related to multi-degree tooth faults from the vibration signal, the proposed method preprocesses the raw vibration signal by using the newly developed localized adaptive denoising technique. The localized adaptive denoising technique results in optimized vibration sub-bands with reduced noise. To obtain fault-related information in the form of a time–frequency scale image, a wavelet-based vibration imaging approach is applied to the denoised vibration signal. Finally, these wavelet-based vibration images are provided as an input to a deep convolutional neural network model for fault classification. The deep convolutional neural network is specifically developed for fault diagnosis purposes. To verify the effectiveness of the proposed method, the proposed method was applied to two different datasets. The first dataset had a fixed speed; however, the second dataset consisted of variable speed conditions. On both datasets, the proposed method

outperformed the existing state-of-the-art methods with an average classification accuracy of 100%. In the future, the goal is to apply the proposed fault diagnosis technique to the fault diagnosis of complex rotating machinery, such as centrifugal pumps.

Author Contributions: Conceptualization, C.D.N., Z.A. and J.-M.K.; data curation, C.D.N. and Z.A.; formal analysis, C.D.N., Z.A. and J.-M.K.; funding acquisition, J.-M.K.; methodology, C.D.N., Z.A. and J.-M.K.; software, C.D.N. and Z.A.; supervision, J.-M.K.; validation, C.D.N., Z.A. and J.-M.K.; visualization, C.D.N., Z.A. and J.-M.K.; writing—original draft, C.D.N. and Z.A.; writing—review and editing, J.-M.K. All authors have read and agreed to the published version of the manuscript.

Funding: Following are results of a study on the “Leaders in Industry-university Cooperation +” Project, supported by the Ministry of Education and National Research Foundation of Korea.

Data Availability Statement: The data are publicly available.

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

References

- Samuel, P.D.; Pines, D.J. A review of vibration-based techniques for helicopter transmission diagnostics. *J. Sound Vib.* **2005**, *282*, 475–508. [[CrossRef](#)]
- Nie, M.; Wang, L. Review of condition monitoring and fault diagnosis technologies for wind turbine gearbox. *Procedia CIRP* **2013**, *11*, 287–290. [[CrossRef](#)]
- Praveenkumar, T.; Saimurugan, M.; Krishnakumar, P.; Ramachandran, K.I. Fault diagnosis of automobile gearbox based on machine learning techniques. *Procedia Eng.* **2014**, *97*, 2092–2098. [[CrossRef](#)]
- Alban, L.E. Failures of gears. In *Failure Analysis and Prevention*; William, T., Becker, R.J.S., Eds.; ASM International: Almere, The Netherlands, 2002; Volume 11.
- Goyal, D.; Pabla, B.S.; Dhama, S.S. Condition monitoring parameters for fault diagnosis of fixed axis gearbox: A review. *Arch. Comput. Methods Eng.* **2017**, *24*, 543–556. [[CrossRef](#)]
- Sait, A.S.; Sharaf-Eldeen, Y.I. A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics. In *Rotating Machinery, Structural Health Monitoring, Shock and Vibration*; Springer: New York, NY, USA, 2011; Volume 5, pp. 307–324.
- Mitchell, J.S. An Introduction to machinery analysis and monitoring. *Comput. Eng.* **1991**, *10*, 314–315. [[CrossRef](#)]
- Ghodake, S.B.; Mishra, P.A.K.; Deokar, P.A.V. A review on fault diagnosis of gear-box by using vibration analysis method. *IPASJ Int. J. Mech. Eng.* **2016**, *4*, 31–35.
- Baxter, J.W.; Bumby, J.R. An explanation for the asymmetry of the modulation sidebands about the tooth meshing frequency in epicyclic gear vibration. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **1985**, *199*, 65–70. [[CrossRef](#)]
- McNames, J. Fourier series analysis of epicyclic gearbox vibration. *J. Vib. Acoust. Trans. ASME* **2002**, *124*, 150–160. [[CrossRef](#)]
- Bartelmus, W.; Zimroz, R. A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. *Mech. Syst. Signal Process.* **2009**, *23*, 1528–1534. [[CrossRef](#)]
- Kohler, H.K.; Pratt, A.; Thompson, A.M. Dynamics and noise of parallel-axis gearing. *Proc. Inst. Mech. Eng. Conf. Proc.* **1969**, *184*, 111–121. [[CrossRef](#)]
- Patil, C.R.; Kulkarni, P.P.; Sarode, N.N. Gearbox noise & vibration prediction and control. *Int. Res. J. Eng. Technol.* **2017**, *4*, 873–877.
- Randall, R.B.; Antoni, J.; Chobsaard, S. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mech. Syst. Signal Process.* **2001**, *15*, 945–962. [[CrossRef](#)]
- Randall, R.B. *Frequency Analysis*, 3rd ed.; Brüel & Kjaer: Nairobi, Denmark, 1987; ISBN 8787355078.
- Kang, M.; Kim, J.; Kim, J.M.; Tan, A.C.C.; Kim, E.Y.; Choi, B.K. Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis. *IEEE Trans. Power Electron.* **2015**, *30*, 2786–2797. [[CrossRef](#)]
- Loutridis, S.J. Damage detection in gear systems using empirical mode decomposition. *Eng. Struct.* **2004**, *26*, 1833–1841. [[CrossRef](#)]
- Zhang, C.; Peng, Z.; Chen, S.; Li, Z.; Wang, J. A gearbox fault diagnosis method based on frequency-modulated empirical mode decomposition and support vector machine. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2018**, *232*, 369–380. [[CrossRef](#)]
- Aharamuthu, K.; Ayyasamy, E.P. Application of discrete wavelet transform and Zhao-Atlas-Marks transforms in non stationary gear fault diagnosis. *J. Mech. Sci. Technol.* **2013**, *27*, 641–647. [[CrossRef](#)]
- Liu, B.; Riemenschneider, S.; Xu, Y. Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum. *Mech. Syst. Signal Process.* **2006**, *20*, 718–734. [[CrossRef](#)]
- Yang, Q.; An, D. EMD and wavelet transform based fault diagnosis for wind turbine gear box. *Adv. Mech. Eng.* **2013**, *5*, 212836. [[CrossRef](#)]
- Nguyen, C.D.; Prosvirin, A.; Kim, J.M. A reliable fault diagnosis method for a gearbox system with varying rotational speeds. *Sensors* **2020**, *20*, 3105. [[CrossRef](#)]

23. Nguyen, C.D.; Prosvirin, A.E.; Kim, C.H.; Kim, J.-M. Construction of a sensitive and speed invariant gearbox fault diagnosis model using an incorporated utilizing adaptive noise control and a stacked sparse autoencoder-based deep neural network. *Sensors* **2020**, *21*, 18. [[CrossRef](#)]
24. 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](#)]
25. Samanta, B. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mech. Syst. Signal Process.* **2004**, *18*, 625–644. [[CrossRef](#)]
26. Han, D.; Zhao, N.; Shi, P. Gear fault feature extraction and diagnosis method under different load excitation based on EMD, PSO-SVM and fractal box dimension. *J. Mech. Sci. Technol.* **2019**, *33*, 487–494. [[CrossRef](#)]
27. Gunasegaran, V.; Muralidharan, V. Fault diagnosis of spur gear system through decision tree algorithm using vibration Signal. *Mater. Today Proc.* **2019**, *22*, 3232–3239. [[CrossRef](#)]
28. Strączkiewicz, M.; Barszcz, T. Application of artificial neural network for damage detection in planetary gearbox of wind turbine. *Shock Vib.* **2016**, *2016*, 1–12. [[CrossRef](#)]
29. Caesarendra, W.; Tjahjowidodo, T. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines* **2017**, *5*, 21. [[CrossRef](#)]
30. Rioul, O.; Vetierli, M. Wavelets and signal processing. *IEEE Signal Process. Mag.* **1991**, *8*, 14–38. [[CrossRef](#)]
31. Zheng, H.; Li, Z.; Chen, X. Gear fault diagnosis based on continuous wavelet transform. *Mech. Syst. Signal Process.* **2002**, *16*, 447–457. [[CrossRef](#)]
32. Peng, Z.K.; Chu, F.L. Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography. *Mech. Syst. Signal Process.* **2004**, *18*, 199–221. [[CrossRef](#)]
33. Saufi, S.R.; Ahmad, Z.A.; Leong, M.S.; Lim, M.H. Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. *IEEE Access* **2019**, *7*, 122644–122662. [[CrossRef](#)]
34. Xu, Y.; Li, C.; Xie, T. Intelligent diagnosis of subway traction motor bearing fault based on improved stacked denoising autoencoder. *Shock Vib.* **2021**, *2021*, 1–9. [[CrossRef](#)]
35. Liu, H.; Zhou, J.; Zheng, Y.; Jiang, W.; Zhang, Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans.* **2018**, *77*, 167–178. [[CrossRef](#)]
36. Wang, W.F.; Qiu, X.H.; Chen, C.; Lin, B.; Zhang, H.M. Application research on long short-term memory network in fault diagnosis. *Proc. Int. Conf. Mach. Learn. Cybern.* **2018**, *2*, 360–365. [[CrossRef](#)]
37. Zhang, X.; Cong, Y.; Yuan, Z.; Zhang, T.; Bai, X. Early fault detection method of rolling bearing based on MCNN and GRU network with an attention mechanism. *Shock Vib.* **2021**, *2021*. [[CrossRef](#)]
38. Grezmak, J.; Wang, P.; Sun, C.; Gao, R.X. Explainable convolutional neural network for gearbox fault diagnosis. *Procedia CIRP* **2019**, *80*, 476–481. [[CrossRef](#)]
39. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [[CrossRef](#)]
40. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
41. Jing, L.; Zhao, M.; Li, P.; Xu, X. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Meas. J. Int. Meas. Confed.* **2017**, *111*, 1–10. [[CrossRef](#)]
42. Lu, C.; Wang, Z.; Zhou, B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Adv. Eng. Inform.* **2017**, *32*, 139–151. [[CrossRef](#)]
43. Jiao, J.; Zhao, M.; Lin, J.; Zhao, J. A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes. *Knowl. Based Syst.* **2018**, *160*, 237–250. [[CrossRef](#)]
44. Dalpiaz, G.; Dalpiaz, G.; Rivola, A.; Rubini, R. Dynamic modelling of gear system for condition monitoring and diagnostics. In Proceedings of the Congress on Technical Diagnostics, Bochum, Germany, 16–19 July 1996; pp. 185–192.
45. Fakhfakh, T.; Chaari, F.; Haddar, M. Numerical and experimental analysis of a gear system with teeth defects. *Int. J. Adv. Manuf. Technol.* **2005**, *25*, 542–550. [[CrossRef](#)]
46. Chaari, F.; Bartelmus, W.; Zimroz, R.; Fakhfakh, T.; Haddar, M. Gearbox vibration signal amplitude and frequency modulation. *Shock Vib.* **2012**, *19*, 635–652. [[CrossRef](#)]
47. Fan, X.; Zuo, M.J. Gearbox fault detection using Hilbert and wavelet packet transform. *Mech. Syst. Signal Process.* **2006**, *20*, 966–982. [[CrossRef](#)]
48. Houser, D.R. Gear noise and vibration prediction and control methods. In *Handbook of Noise and Vibration Control*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; pp. 847–856.
49. Lutovac, M.D.; Tošić, D.V.; Evans, B.L. *Filter Design for Signal Processing Using MATLAB and Mathematica*; Prentice-Hall: Englewood Cliffs, NJ, USA, 2001; ISBN 0201361302.
50. Dahl, G.E.; Sainath, T.N.; Hinton, G.E. Improving deep neural networks for LVCSR using rectified linear units and dropout. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8609–8613.
51. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2323. [[CrossRef](#)]

52. Aggarwal, C.C. Convolutional neural networks. In *Neural Networks and Deep Learning*; Springer: Cham, Switzerland, 2018; pp. 315–352.
53. Yoo, Y.; Baek, J.G. A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. *Appl. Sci.* **2018**, *8*, 1102. [[CrossRef](#)]
54. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
55. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
56. Nguyen, C.D.; Prosvirin, A.; Kim, J.-M. Fault Identification of multi-level gear defects using adaptive noise control and a genetic algorithm. In *Intelligent Human Computer Interaction*; Singh, M., Kang, D.-K., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 325–335.