MDPI

Article

Gearbox Fault Identification Model Using an Adaptive Noise Canceling Technique, Heterogeneous Feature Extraction, and Distance Ratio Principal Component Analysis

Cong Dai Nguyen [1]🆔, Cheol Hong Kim [2]🆔 and Jong-Myon Kim [3,4,*]🆔

[1] Faculty of Radio-Electronic Engineering, Le Quy Don Technical University, Hanoi 10000, Vietnam; dainc@lqdtu.edu.vn
[2] School of Computer Science and Engineering, Soongsil University, Seoul 06978, Korea; cheolhong@ssu.ac.kr
[3] Department of Electrical, Electronics, and Computer Engineering, University of Ulsan, Ulsan 44610, Korea
[4] PD Technology Cooperation, Ulsan 44610, Korea
[*] Correspondence: jmkim07@ulsan.ac.kr; Tel.: +82-52-259-2217

Abstract: Using an adaptive noise canceling technique (ANCT) and distance ratio principal component analysis (DRPCA), this paper proposes a new fault diagnostic model for multi-degree tooth-cut failures (MTCF) in a gearbox operating at inconsistent speeds. To account for background and disturbance noise in the vibration characteristics of gear failures, the proposed approach employs ANCT in the first stage to optimize vibration signals. The ANCT applies an adaptive denoising technique to each basic frequency segment in the whole frequency response of vibrations. Following that, a novel DRPCA is used to extract the discriminating low-dimensional features. The DRPCA initially determines each feature's relative proximity to fault categories by computing the average Euclidian distance ratio between similar and dissimilar classes. The most discriminatory features with the lowest dimensions are selected, as determined by principal component analysis (PCA). The new DRPCA is created by combining distance ratio–based feature inspection with PCA. The optimal feature set containing the most discriminative features is then fed to the support vector machine classifier to identify multiple failure categories. The experimental results indicate that the proposed model outperforms the state-of-art approaches and offers the highest identification accuracy.

Keywords: fault diagnosis; feature extraction; gearbox fault identification; adaptive noise canceling technique; principal component analysis; support vector machine

1. Introduction

Gear failures are the most common problems in gearboxes, which are used in many types of machinery, automobiles, and wind turbines [1–4], because of difficult and constant working conditions. A gearbox failure can cause severe system failures, financial losses, and workforce risks. Thus, the early detection of gearbox failings is essential. The condition-based monitoring method provides maintenance tasks based on gearbox data and can extend gearbox lifespans while reducing maintenance costs [5]. The most effective technique for diagnosing gearbox failure is vibration-based condition monitoring [6]. The gearbox vibration signal includes tooth meshing harmonics, sideband frequencies, and free oscillations. In the ideal case, failure-related oscillations are blended sideband frequencies [7,8]. However, system interconnections, mechanical systems (resonances caused by a shaft, bearing, gear, etc.), and background noise are major sources of interference [9,10]. Vibration signals have noisy components that dominate and distort the failure-related ingredients with their random magnitudes and occurrences. To identify gear failures, the raw vibration signal must be optimized to reduce uncertainty about protruding fault-associated elements, such as meshing frequency harmonics and sideband frequencies.

For rotating-machine fault diagnosis, many signal processing techniques have been proposed to examine the signal in multiple domains. The Hilbert transform (HT), window bandpass filters, short-time Fourier transform, empirical mode decomposition (EMD), and wavelet transform (WT) are some of those techniques [11–15]. Hybrid techniques use WT-HT and EMD-WT [16,17]. All those techniques reduce the interference noise to some extent, but they degrade the amplitudes of the sideband and meshing frequency harmonics, distorting the failure signature information associated with gear failures in the vibration signals. Therefore, those approaches are unlikely to be useful in identifying the various types of multi-degree tooth-cut failure (MTCF) in a gearbox using fluctuating speed operations. Nguyen et al. [18,19] demonstrated that their adaptive noise reduction method effectively reduced noise in a raw vibration signal. They used the resulting signal to identify failures in gearboxes with multi-level tooth cut faults. Their adaptive model works by searching an optimal set of Gaussian parameters linked to filter weights across the whole frequency range of a vibration signal. However, a non-stationary and complicated vibration signal can be caused by both random noise and gear stiffness variation. Meshing harmonics, sideband components, and random noise are all present in gearbox vibration signals, with different energy distributions in many frequency segments along the frequency spectrum. Therefore, applying a single optimized Gaussian parameter set to all frequency ranges is unsatisfactory. To analyze vibration signals from MTCF failures, we propose an adaptive noise canceling technique (ANCT). The proposed method adopts an adaptive denoising technique (ADT) from [18] to analyze the frequency domains segmentally. The optimized vibration signals provide a wealth of information about gear failures after they've been transformed by the ANCT. That is, the ANCT significantly reduces noise while retaining the original failure information. The proposed gearbox fault diagnostic approach uses the ANCT outputs to configure feature pools and identify failure types.

In terms of feature engineering and classification, the fault-relevant elements in the vibration signals of an MTCF gearbox are too similar because the failure categories are all reflected in the vibration characteristics in the same way. Thus, the output signals of the ANCT must be evaluated in an enhanced manner. Numerous approaches are used for feature extraction, including statistical feature calculation, complex envelope analysis, and wavelet packet analysis [20–23]. We hypothesize that expanding the discrimination representation of a failure diagnostic system beyond a single feature model will minimize the possibility of overlooking significant information in the data. In other words, we deem it preferable to generate as much information about the process state as possible, even if it is tainted by redundancy, and then filter out the most useful aspects in a subsequent information processing phase. Therefore, we use three feature computation techniques (statistics, wavelets, and envelopes) in this study to configure a heterogeneous feature pool (HFP). Although the HFP increases the effectiveness of gearbox fault expression for fault detection, the large dimensionality of the feature pool poses a difficulty for machine learning. To address that issue, feature selection [24], which is a critical technique in artificial intelligence, can be applied to reduce the dimensionality of the HFP and optimize system accuracy. This paper analyzes a collection of potential features using multivariate search methods, with the goal of minimizing duplication in the final, optimized feature set. Thus, we choose the most differentiating features from the HFP, which include both relevant and redundant information, by developing a search algorithm. From the many types of analysis available, including independent component analysis (ICA) [25], linear discriminant analysis (LDA) [26], genetic algorithm (GA) [27], and principal component analysis (PCA) [28,29], the PCA technique leverages eigenvalues from the covariance matrix of the variables to explore the statistical structures of the data and choose the principal elements (those with the highest eigenvalues). Both supervised and unsupervised PCA techniques were investigated in [29] for condition-monitoring of rotating machines. According to that analysis, PCA performed better than the other dimensional reduction techniques for failure condition classification. However, PCA does not consider the distance between the attributes of different failure classes. To address this issue, we propose a new

feature selection approach, the distance ratio principal component analysis (DRPCA). This new technique starts by performing a PCA on the HFP to compute the average Euclidean distance parameters between features pertaining to the same and diverse failure categories. The novel DRPCA combines a relative distance–based evaluation with a PCA.

DRPCA configures a feature set of the most distinct features, which is then used as input data for machine learning–based classification. Compared with other artificial intelligence techniques, support vector machines (SVMs) have higher generalization, allowing them to diagnose mechanical defects in a rotating machine with high accuracy [30]. SVMs can also effectively separate nonlinear datasets using hyperplane mapping functions [31]. To classify multiclass data, SVMs use the one against one, one against all, or hierarchical strategies. The most accurate approach is a one against one SVM (OAOSVM) [32], which we thus use in our proposed model.

Moreover, gear fault diagnosis models have been proposed to identify the MTCF gearbox under variable speed conditions, and they validated the superiority of the classification accuracies [19,33]. These models were developed using a combination of adaptive denoising approaches and deep neural network architectures (DNAs). Accordingly, the results of the adaptive denoising process were then used to input to the deep neural network for the enhancement operation through three execution stages: (1) a high dimensional feature pool configuration by the automated feature extraction, (2) a fine-tuning process for selection of the most discriminative features, and (3) a gear fault identification process by the classifying layer in the deep network hierarchy. The constructed DNAs have effectively performed feature engineering and classification to obtain highly accurate classification outcomes. However, since DNAs use more computing resources than machine learning classifiers, they may not be appropriate for certain applications, particularly in real-time observation systems. Therefore, we propose the gearbox fault identification model to perform a manual feature engineering process (i.e., HFP + DRPCA) and a machine learning–based classification approach (OVOSVM) for achieving the precise identification of the fault categories.

The following outlines the contributions of this paper.

(1) ANCT is a new approach for signal processing that uses an adaptive method to identify the appropriate Gaussian parameter set for each basic frequency segment of a vibration frequency domain. The optimized vibration signal outputs of the ANCT contain failure-related information with decreased disturbance noise.

(2) Various feature models (statistical computation, wavelet basis decomposition, and complex envelope decomposition) are used to extract features from the ANCT results. The resulting HFP offers as many failure signatures as feasible for the condition monitoring process; it can also represent various fault types.

(3) The DRPCA chooses the most distinctive features in the HFP to identify fault categories. To compute the distance ratio between features of the failure types, the DRPCA first performs principal component analysis, examines the distance between the features of different failure types, and then selects the features with the highest relative distance ratio to create an optimal feature set with decreased dimensions.

(4) To evaluate the proposed technique, vibration signals from an MTCF gearbox with various types of failures were acquired from a real-world experimental testbed.

The remainder of this article is organized as follows: Section 2 discusses the underlying approaches used in this study; Section 3 explains the experimental testbed for an MTCF gearbox and vibration measurements; Section 4 constructs the proposed diagnostic model for an MTCF gearbox; Section 5 details the steps of performance assessment and the experimental results obtained from the proposed model; and Section 6 summarizes the findings of this study.

2. The Background of the Techniques

2.1. Adaptive Noise Filtering Technique

The adaptive noise filtering technique uses two key operational functions: digital filtering and adaptive processing [34], as shown in Figure 1. The adaptive method adjusts the coefficient parameters of a digital filter to reduce the output error for denoising.
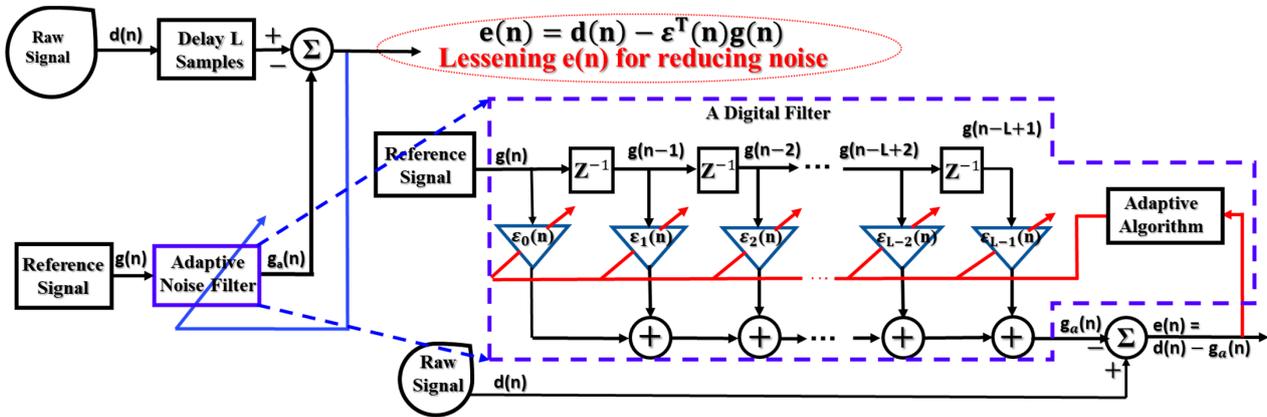

Figure 1. The functional structure of an adaptive noise filtering technique.

Assuming that a coefficient vector of an L-tap finite impulse response digital filter is denoted by $\varepsilon(n)$ (i.e., $\varepsilon(n) = [\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{L-1}]^T$), a raw signal $d(n)$ and a reference signal $g(n)$ (those are the discrete time signals, and n indicates a time variable in the discreteness) and output error $e(n)$ can be calculated according to [34], as follows:

$$e(n) = d(n) - \varepsilon^T(n)g(n) \tag{1}$$

The mean square error (MSE) is used to determine the convergence condition for the output error:

$$
\begin{aligned}
J &\equiv E\{e^2(n)\} = E\{[d(n) - \varepsilon^T(n)g(n)]^2\} \\
J &= \varepsilon^T(n)C\varepsilon(n) - 2P\varepsilon^T(n) + E\{d^2(n)\}
\end{aligned}
\tag{2}
$$

where $C \equiv E\{g(n)d(n)\}$ indicates the cross-correlation parameters of a reference signal and a raw signal, and $P \equiv E\{g(n)g^T(n)\}$ signifies the autocorrelation parameter of a reference signal. As shown in Equation (2), the MSE is a quadratic function, so there exists a unique global minimum corresponding to an optimal coefficient vector. The least mean square (LMS) process is the technique most often used to reduce the output error [35], and the weight parameters can be adjusted as follows:

$$\varepsilon(n+1) = \varepsilon(n) + 2\lambda g(n)\varepsilon(n) \tag{3}$$

Here, the convergence factor (λ), which is used as a step-size parameter to estimate the minimum MSE, can be adjusted between about 0 and $(LS_r)^{-1}$ (S_r is the average power of a reference signal $g(n)$, and L signifies the order of a digital filter).

2.2. Wavelet Packet Transform (WPT)

WPT is a variant of the discrete wavelet transform that exhibits the capacity to decompose signal information in the high-frequency region in depth [36]. WPT divides a signal into several sub-bands, with great resolution in both the low- and high-frequency domains, by using a sequence of scaling analysis (low-pass) and wavelet analysis (high-pass) filters, resulting in 2^l sub-bands, where l is the depth level. The following formula encapsulates the WPT function [36]:

$$W_{\delta,\rho}^n(t) = 2^{\frac{\delta}{2}} W^n\left(2^\delta t - \rho\right), \tag{4}$$

where ρ and δ are the transition and scale parameters, respectively. The scaling and mother wavelet functions are assigned as $n = 0$ $(W_{0,0}^0(t) = \Phi(t))$, and $n = 1$ $(W_{0,0}^1(t) = \Psi(t))$, respectively. Here, $\Phi(t)$ and $\Psi(t)$ are a scaling function and a wavelet function that form the multi-resolution analysis technique in WPT. The scaling function represented by the low-pass filters is used to analyze the input signals in the low-resolution approximation spaces, while the wavelet function is applied to refine those approximation spaces in the higher frequency zones, implemented by the band-pass filters [36].

Then, other wavelet packet functions can be calculated through a recursive process, which is represented as

$$W_{\delta,\rho}^{2n}(t) = \sqrt{2}\sum_{\rho=-\infty}^{\rho=\infty}h(\rho)W^n(2t - \rho), \tag{5}$$

$$W_{\delta,\rho}^{2n+1}(t) = \sqrt{2}\sum_{\rho=-\infty}^{\rho=\infty}g(\rho)W^n(2t - \rho). \tag{6}$$

Here, $h(\rho)$ and $g(\rho)$ are low- and high-pass quadrate mirror filters, respectively, which are associated with the selected mother wavelet and scaling functions. Therefore, the wavelet coefficients of a signal $s(t)$ obtained by the WPT method can be calculated by the following inner product formula:

$$\theta_\delta^n(\rho) = \ <s(t), W_{\delta,\rho}^n> \ = \int_{-\infty}^{\infty} s(t), W_{\delta,\rho}^n(t)dt. \tag{7}$$

Using Equations (5)–(7), WPT decomposes each node into two sub-bands using high- and low-pass filters formed by the wavelet packet coefficients. Each coefficient $\theta_\delta^n(\rho)$ is the specific parameter for each resolution in the frequency domain and thus represents the scaling parameter and oscillation component in the observed signal.

2.3. Complex Envelope Analysis

The HT is used to analyze the complex envelopes of a signal to expand it into a complex space and generate an analytical signal in the new time domain. If $y(t)$ represents the analytical signal of the observed signals $s(t)$ in the time domain, then it can be expressed as follows [37]:

$$y(t) = s(t) + j\widetilde{s}(t), \tag{8}$$

where $j = \sqrt{-1}$, and $\widetilde{s}(t)$ is defined as the convolution of $s(t)$ and $1/\pi t$ in the time domain, as follows:

$$\widetilde{s}(t) = s(t) * \frac{1}{\pi t} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t - \tau}d\tau. \tag{9}$$

In the frequency domain, its frequency spectrum can be calculated as

$$\widetilde{S}(\omega) = -jsgn(\omega)S(\omega); -jsgn(\omega) = \begin{cases} j, \ with \ \omega > 0 \\ 0, \ with \ \omega = 0 \\ -j, \ with \ \omega < 0 \end{cases} \tag{10}$$

The complex envelope analysis is then performed by computing the attributes of the observed signal in the complex spaces defined by Equations (8)–(10).

2.4. Principle Component Analysis

PCA is a common procedure for shrinking the dimensionality of a feature pool by assessing the variances of many extracted features to identify which attributes have the largest effect on the data structure. By using this statistical technique, it is possible to exclude weaker-distinctive elements from a high-dimensional pool while retaining the most fault-related representational components (i.e., the dimensionality of the feature pool is reduced). In other words, the initial coordinate space of the observed data maps onto the direction of increasing variance. The computation process of PCA, which is presented in [38], is explained in detail by the following formulas.

Consider $S \in \mathbb{R}^{m \times p}$ as a feature vector with zero mean and unit variance that has been normalized from an original feature vector consisting of m rows of samples and p variable columns. The covariance matrix C can be computed as

$$C = \frac{1}{m-1} S^T S \qquad (11)$$

By performing single-value decomposition for matrix C, it can be represented as

$$C = V \Delta V^T, \qquad (12)$$

where V is a matrix of eigenvectors of C, and Δ signifies a diagonal matrix, which consists of the downward order eigenvalues of C (i.e., $\lambda_i \geq \lambda_{i+1} \geq .. \geq \lambda_p \geq 0$). Then, by producing the transfer matrix $\mathcal{L} \in \mathbb{R}^{m \times k}$ ($k < p$) with k eigenvectors (k columns of V) equivalent to the desired number of primary eigenvalues (k), the dimension of the observed variables can be reduced ($k < p$). Therefore, several formulae represent the manipulating performance of a PCA:

$$\left|\begin{array}{l} K = S\mathcal{L} \\ \hat{S} = K\mathcal{L}^T \\ E = S - \hat{S} \\ S = K\mathcal{L}^T + E. \end{array}\right. \qquad (13)$$

where K is a scoring matrix, $\mathcal{L}$ is a loading matrix, $\hat{S}$ indicates a principal source of the original feature space, and E represents the redundancy in the process. It is essential to select k to filter the principal components.

3. The Experimental Gearbox Test-Rig and Dataset Description

Figure 2 depicts the experimental test equipment used to obtain vibration data from a gearbox system. The functional structure of the test rig is shown in Figure 2a; the gearbox transfers rotational motion from a three-phase motor to adjustable blades (i.e., the load) through a 1:1.52 reduction ratio. The non-drive shaft is used to connect the load and a gear wheel. The driveshaft connects the pinion wheel directly to the three-phase motor. The pinion wheel and gear wheel are engaged by the teeth (each tooth is 9 mm in length), as indicated in Figure 2b. The vibration sensor is positioned at the non-drive end to monitor the vibration characteristics of the gearbox. The displacement transducer is mounted on the driveshaft to track the rotational speeds of the pinion wheel. The vibration dataset in this study was collected using a PCI-based data acquisition device that digitizes vibration signals from the accelerometer. Table 1 summarizes the specifications for the vibration sensor, transducer, and acquisition board.

Real-world gearbox systems suffer two major sorts of gear failures: those induced during manufacture or installation (e.g., incorrect tooth sketching, wheel alignment, or parallelism) and those that occur during operation (e.g., tooth pitting, tooth spalling, tooth cracking, tooth breakage). Gear failures were simulated in this testbed by cutting one tooth with several degrees of tooth length (9 mm) to generate multi-degree tooth cut failures of 6.6% (0.6 mm), 10% (0.9 mm), 20% (1.8 mm), 30% (2.7 mm), 40% (3.6 mm), and 50% (4.5 mm), as shown in Figure 3. These failures were generated to approximate the operational problems that occur in actual gearbox systems as a result of long-term rotational performance fatigue.
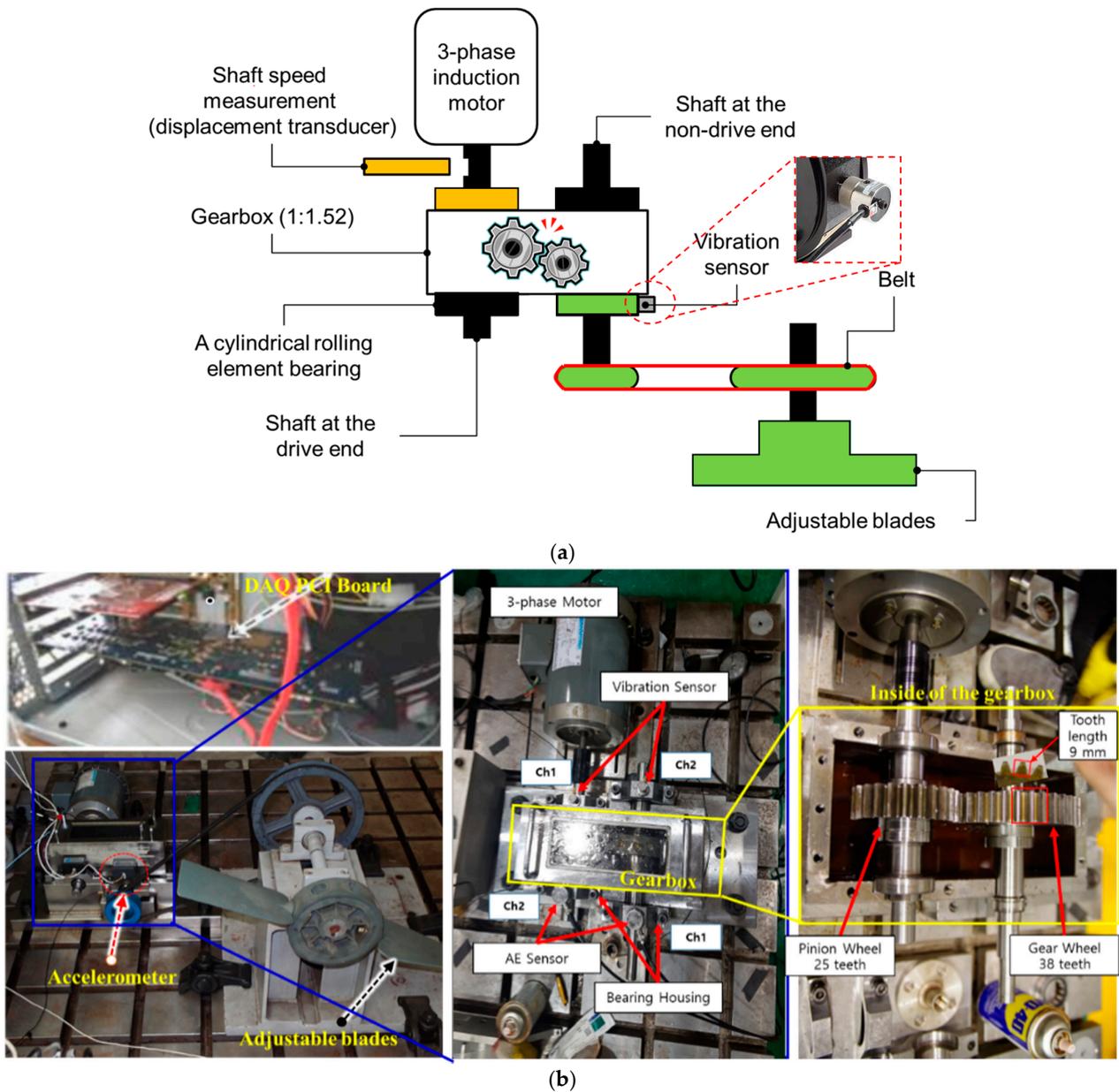
(**a**)

(**b**)

Figure 2. The experimental gearbox test rig: (**a**) functional structure block diagram and (**b**) actual arrangement of the test rig.

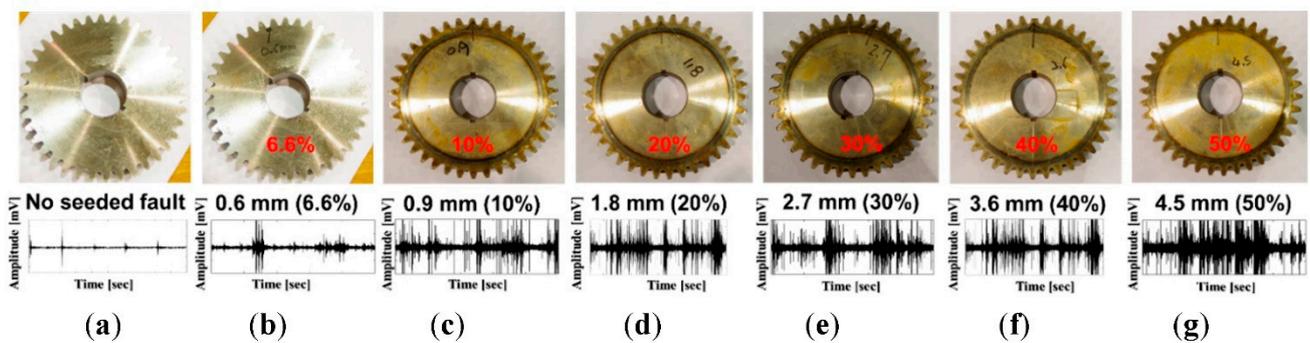(**a**)　　(**b**)　　(**c**)　　(**d**)　　(**e**)　　(**f**)　　(**g**)

Figure 3. The multi-degree tooth cut failures of the gear wheel and the vibration signals at 600 RPM: (**a**) NF, (**b**) F1, (**c**) F2, (**d**) F3, (**e**) F4, (**f**) F5, and (**g**) F6.

Table 1. The specifications of the sensors and the acquisition module.

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

Analog vibration signals were produced by an accelerometer and transformed into digital vibration samples at a sampling frequency of 65.536 kHz. Each sample was one-second long, and we captured 65,536 data points. The acquisition method was performed 150 times for each of the seven failure conditions and alternated over four rotational speeds (300 RPM, 600 RPM, 900 RPM, and 1200 RPM), which produced the design for the experimental dataset used in this study (Table 2).

Table 2. The vibration dataset for the MTCF gearbox obtained from the test rig and data acquisition equipment.

Gearbox Failure Condition	Description	Number of Samples for Each Rotation Speed (RPM)				Sampling Frequency (Hz)
		300	600	900	1200	
Non-Failure (NF)	Normal or perfect gearbox	150	150	150	150	65,536
Failure Type 1 (F1)	6.6% of tooth length (0.6 mm/9 mm)	150	150	150	150	65,536
Failure Type 2 (F2)	10% of tooth length (0.9 mm/9 mm)	150	150	150	150	65,536
Failure Type 3 (F3)	20% of tooth length (1.8 mm/9 mm)	150	150	150	150	65,536
Failure Type 4 (F4)	30% of tooth length (2.7 mm/9 mm)	150	150	150	150	65,536
Failure Type 5 (F5)	40% of tooth length (3.6 mm/9 mm)	150	150	150	150	65,536
Failure Type 6 (F6)	50% of tooth length (4.5 mm/9 mm)	150	150	150	150	65,536

The vibration samples gathered fully reflect the vibrational characteristics of the gearbox. Specifically, a linear and regularly oscillating vibration signal is derived from a gearbox system in the absence of failure [39]. As helpful identifying information, the frequency spectrum comprises the tooth meshing frequency (which reflects the stiffness of the gearbox) and its harmonics. When a pair of gears rotates across a defective tooth, the angular velocity changes abruptly. When a gearbox fails, its rigidity fluctuates, complicating the signal structure (e.g., making it nonlinear and nonstationary). In other words, the nonlinear and nonstationary properties of a faulty gearbox are shaped by its angular

acceleration, as recorded by a vibration sensor. The frequency spectrum of the vibration signals contains the meshing frequency harmonics, sideband frequency tones surrounding each harmonic, and other oscillations, which can each exhibit amplitude-phase modulation characteristics [40]. The frequency spectrum analysis of vibration samples taken from a gearbox rotating at 900 RPM under non-failure and failure type 3 conditions is shown in Figure 4.
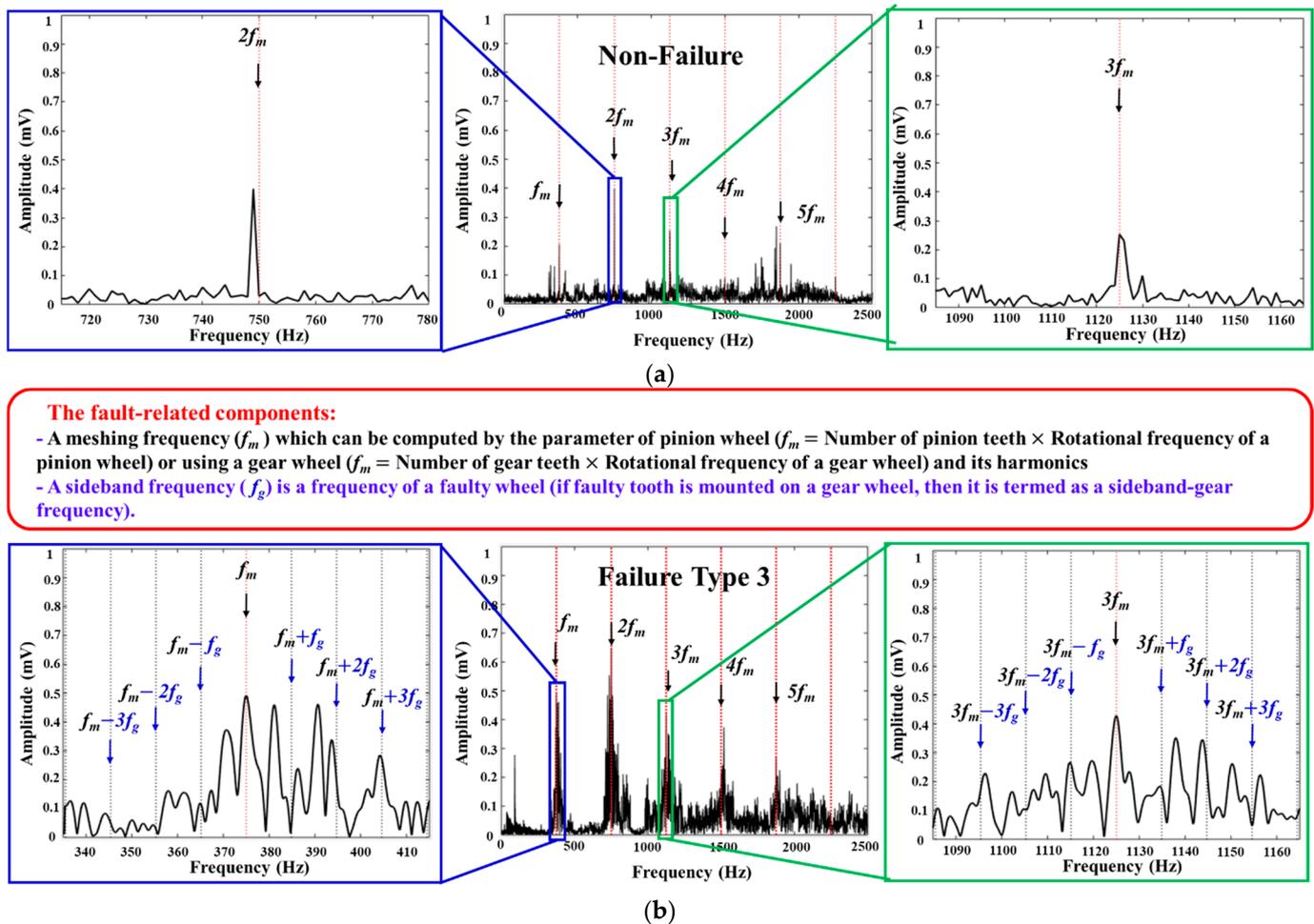

(a)

The fault-related components:
- A meshing frequency (f_m) which can be computed by the parameter of pinion wheel (f_m = Number of pinion teeth × Rotational frequency of a pinion wheel) or using a gear wheel (f_m = Number of gear teeth × Rotational frequency of a gear wheel) and its harmonics
- A sideband frequency (f_g) is a frequency of a faulty wheel (if faulty tooth is mounted on a gear wheel, then it is termed as a sideband-gear frequency).

(b)

Figure 4. The frequency spectrum analysis of vibration samples acquired from the non-failure (**a**) and failure type 3 (**b**) conditions at 900 RPM.

4. The Proposed Gearbox Fault Diagnosis Model

The novel fault diagnosis approach presented in this article was used to categorize seven different types of gear failure in an MTCF gearbox running at various rotational speeds. The processing flow shown in Figure 5 is defined by the function blocks applied. The data acquisition system described in Section 3 collected vibration signals to create the dataset for the experiment. Each vibration sample was collected by sampling at the high frequency of 65.536 kHz for one second (the time required to complete a few rotation cycles) to generate a rich sample and capture as much failure-related information as feasible. The ANCT module processes those raw vibration signals. In the first step, a low-pass filter and three-times down-sampling of the raw vibration signals [18] are used to produce vibration samples with real operating frequency tones from 0–10 kHz, which the accelerometer can detect (Table 1). This step attempts to eliminate redundancy in the frequency range of the raw vibration signal introduced by the high-speed sampling procedure. Later, as shown in Figure 5, effective denoising approaches, multiple domain feature extraction, DRPCA-based feature selection, and an SVM classifier are used to construct the proposed failure

identification model. The following sub-sections contain complete functional descriptions of those modules.
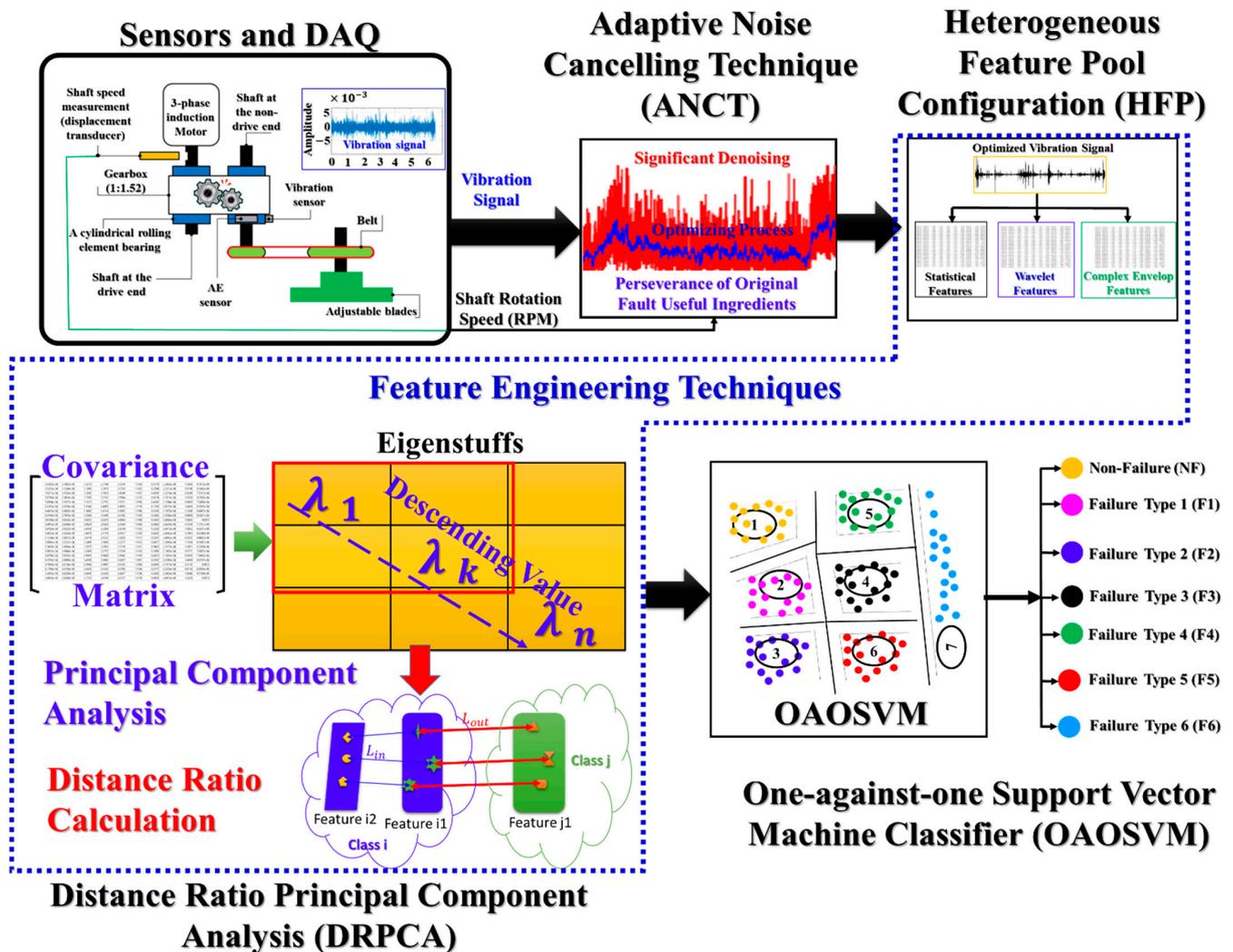

Figure 5. The processing flow and functional blocks in the proposed gearbox fault diagnosis method.

4.1. ANCT

The ANCT is based on the adaptive noise filtering technique outlined in Section 2 and is used to reduce the noise and redundancy of a vibration signal. Accordingly, the raw signal input is the vibration signal, and the reference signal is composed of the signal that simulates the noise oscillations in a gearbox vibration signal, termed the noise-pretended signal (NPS) (Figure 1). Two essential steps operate in the ANCT: the generation of NPSs and the execution of adaptive algorithms to access and reduce noise in multiple segments of the vibration frequency domain. The adaptive denoising techniques (ADTs) developed and verified in [18,19,33] have shown outstanding results in removing noise and protecting the failure-related signal. The noise in the gearbox vibration signal is categorized into two forms using those methods: white noise caused by the measuring system (e.g., data acquisition systems, electronic devices) and waveband noise induced by unrelated mechanical component resonances. Thus, as shown in Figure 6a, the NPS is formed by combining two functions of a uniform random distribution to emulate white noise and a Gaussian distribution to imitate the analogous behavior of waveband noise. Additionally, the parameters (i.e., mean and standard deviation) of the Gaussian signal, which are functions of the rotating speed variable, are associated with filter weights to

be adjusted by the LMS adaptive algorithm [19]. The following formula describes the Gaussian function part of the NPS:

$$G_{gs}(m) = \sum_{m=1}^{M} e^{-\frac{(m-F_{ct})^2}{2\sigma^2}}. \tag{14}$$

where M denotes the number of sideband frequency segments or meshing frequency harmonics. F_{ct} and σ are, respectively, the mean and standard deviation, which are proportional to the rotational speed of the shaft. As a consequence, when a gearbox rotates at different speeds, a Gaussian signal is created that can provide access to the frequency gap between two successive sideband frequency tones.
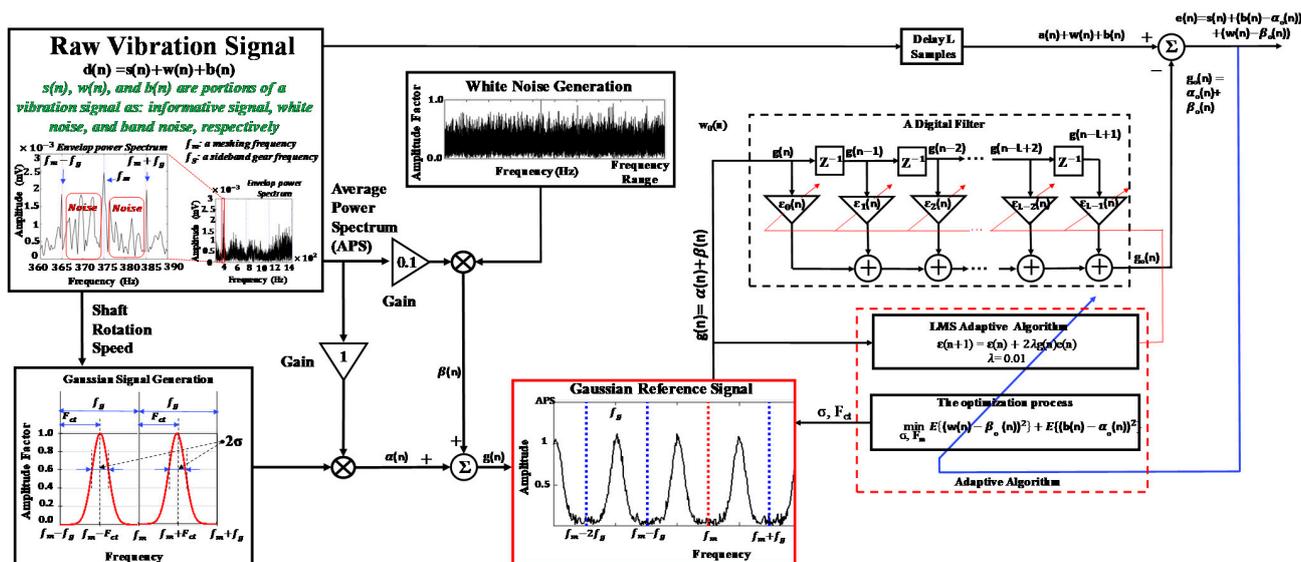
(**a**)
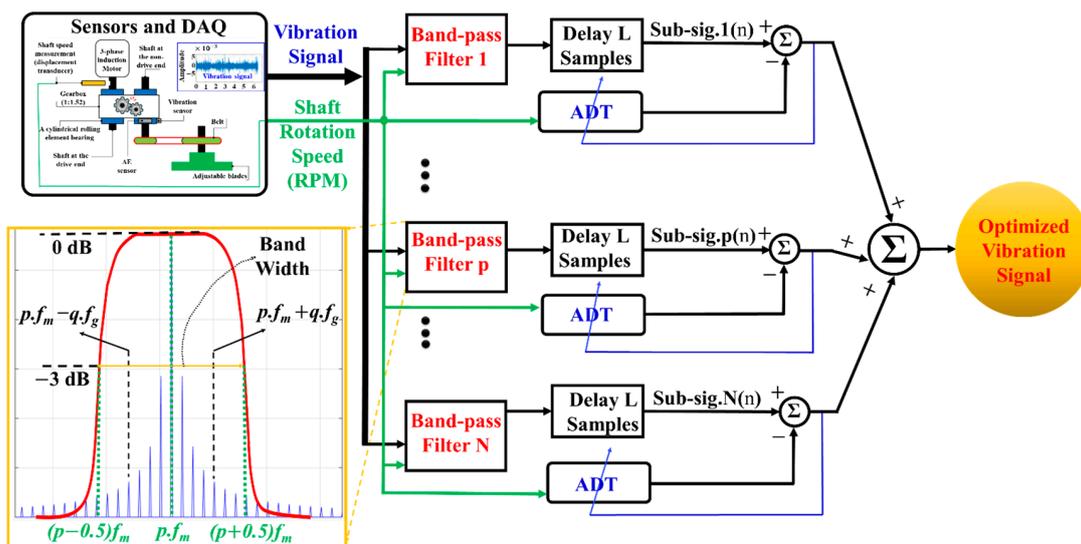
(**b**)

Figure 6. The processing flows and functional blocks of the ADT method (**a**) and the ANCT (**b**).

Next, the LMS technique is used to adaptively adjust the parameters (F_{ct}, σ) of each generated NPS in conjunction with the filter weights to achieve the optimized parameter set [18]. Figure 6a illustrates the signal processing flow and function blocks of the ADT

method. As shown in Figure 6a, the overall procedure of the ADT for each vibration sample is to determine the general optimal parameter set for an NPS coupled to an optimal filter coefficient vector for its full frequency domain (0–10 kHz). Nonetheless, the frequency spectrum of a vibration signal is composed of several basic frequency segments (BFSs), which have different energy distributions. Thus, we propose the ANCT to segment the vibration frequency domain into numerous BFSs (the segment width of each BFS equals a meshing frequency value and is centered on a meshing frequency harmonic) and then apply the ADT method [19] to each BFS to obtain an optimal parameter set for each BFS, thereby producing a series of optimized parameter sets for each vibration signal. The ANCT output is the summation of all the optimized sub-signals, which have been adaptively denoised using the series of segment-observed optimal parameter sets. Thus, the frequency range of the resultant signal of the ANCT is reconstructed in the same range as the input vibration signal. In that way, the ANCT can outperform previous ADT approaches in terms of noise reduction effectiveness and maintaining the original failure-related frequency tones for heterogeneous feature extraction in subsequent processes. The function block schematic of the ANCT is shown in Figure 6b. The ANCT partitions a vibration signal into N sub-signals, each with a frequency spectrum of one BFS (i.e., the frequency range of each sub-signal, which equals the range of BFS or the value of a meshing frequency), using N band-pass filters (Filter 1, Filter 2, . . . , Filter p, . . . , Filter N) simultaneously. Each filter is an IIR Chebyshev Type-I bandpass filter [41] with an order of 30, a bandwidth that is equal to the meshing frequency, and a band-pass frequency range in the consecutive series of the frequency domain of a vibration signal. N is the quotient of the maximum frequency of vibration signal and the meshing frequency. The ADT is then applied to each sub-signal to produce optimally segmented sub-signals. The output signal of the ANCT is then formed by averaging the N-optimized sub-signals. This output signal is referred to as the *optimized vibration signal* and is used to configure the HFP to explore failure-related information across a multiplicity of observation domains.

4.2. Heterogeneous Feature Pool Configuration

4.2.1. Statistical Feature Calculation

The optimized vibration signal, i.e., the output of the ANCT, mostly contains intrinsic defective symptoms of the MTCF gearbox and is statistically computed and extracted in the two domains of time and frequency [42]. The calculation produces twenty-one features—three features in the frequency domain and eighteen features in the time domain—for each optimized vibration signal. Those features might contain some discriminant failure features that can be deemed subservient or congruent in the feature selection stage, as described in Table 3.

Table 3. The statistical features extracted in the time and frequency domains.

Features	Equations	Features	Equations	Features	Equations
Peak	$\text{Max}(|s|)$	Shape factor	$\frac{s_{rms}}{\frac{1}{N}\sum_{n=1}^{N}|s_n|}$	Mean $(\bar{s})$	$\frac{1}{N}\sum_{n=1}^{N}s_n$
Root mean square	$\sqrt{\frac{1}{N}\sum_{n=1}^{N}s_n^2}$	Entropy	$-\sum_{n=1}^{N}p_n.log_2(p_n)$	Shape factor square mean root	$\frac{s_{srm}}{\frac{1}{N}\sum_{n=1}^{N}|s_n|}$
Kurtosis	$\frac{1}{N}\sum_{n=1}^{N}\left(\frac{s_n-\bar{s}}{\sigma}\right)$	Skewness	$\frac{1}{N}\sum_{n=1}^{N}\left(\frac{s_n-\bar{s}}{\sigma}\right)^3$	Margin factor	$\frac{max(s)}{s_{smr}}$

Table 3. *Cont.*

Features	Equations	Features	Equations	Features	Equations
Crest factor	$\dfrac{Max(\lvert s \rvert)}{s_{rms}}$	Square mean root	$\left(\dfrac{1}{N} \sum\limits_{n=1}^{N} \sqrt{\lvert s_n \rvert} \right)^2$	Peak to peak	$\max(s)-\min(s)$
Clearance factor	$\dfrac{Max(\lvert s \rvert)}{s_{smr}}$	5th normalized moment	$\dfrac{1}{N} \sum\limits_{n=1}^{N} \left(\dfrac{s_n - \bar{s}}{\sigma} \right)^5$	Kurtosis factor	$\dfrac{Kurtoris}{s_{rms}^4}$
Impulse factor	$\dfrac{Max(\lvert s \rvert)}{\frac{1}{N} \sum_{n=1}^{N} \lvert s_n \rvert}$	6th normalized moment	$\dfrac{1}{N} \sum\limits_{n=1}^{N} \left(\dfrac{s_n - \bar{s}}{\sigma} \right)^6$	Energy of signal	$\sum\limits_{n=1}^{N} s_n^2$
Frequency center (FC)	$\dfrac{1}{N_f} \sum\limits_{f}^{N_f} S(f)$	Root mean square frequency	$\sqrt{\dfrac{1}{N_f} \sum\limits_{f}^{N_f} S(f)^2}$	Root variance frequency	$\sqrt{\dfrac{1}{N_f} \sum\limits_{f}^{N_f} (S(f) - FC)^2}$

Here is an input signal (i.e., optimized sub-band), N is the total number of samples, $S(f)$ is the magnitude response of the fast Fourier transform of the input signal s, N_f is the total number of frequency bins, $\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (s_n - \bar{s})^2}$, and $p_n = \dfrac{s_n^2}{\sum_{n=1}^{N} s_n^2}$.

4.2.2. Wavelet Package Decomposition (WPD)

As shown in Section 2, WPT is more efficient than other wavelet techniques in analyzing signals in both the low- and high-frequency regions. The entropy and relative energy of the nodes, decomposed by WPT, reveal significant information about the impulsive oscillations in a vibration signal [36] and indicate the reflected vibration of various gear failure conditions. In general, the most difficult aspects of WPD are finding and choosing the mother function as well as the level of the vibration signal decomposition. The feature extraction results are greatly influenced by the selection. The method to find the best wavelet function for analysis of a vibration signal in a rotating machine fault diagnosis system, which was proposed by Rafiee et al. [43], demonstrated that the Daubechies family functions are more effective than other family functions, and the optimal level value is four. Therefore, in this study, we use WPD to execute four-level WPT and produce sixteen wavelet nodes (Figure 7a), which contribute wavelet-based features to the HFP by computing the entropy and relative energy of each node using the following two equations [44].

$$RE_{node}(k) = \frac{\sum_{l=1}^{L} \alpha_{kl}^2}{\sum_{p=1}^{P} \sum_{l=1}^{L} \alpha_{pl}^2} \tag{15}$$

$$E_{node}(k) = -\sum_{l=1}^{L} p_k(l) . log_2(p_k(l)); \; where \; p_k(l) = \frac{\alpha_{kl}^2}{\sum_{l=1}^{L} \alpha_{kl}^2} \tag{16}$$

where $RE_{node}(k)$ and $E_{node}(k)$ are the relative energy and entropy of node k, respectively. L and P signify the total number of wavelet coefficients (i.e., α_{kl}) for each node and the number of all decomposed nodes ($P = 16$), respectively.

The proper mother function of Daubechies 20 exhibits oscillation behaviors. These approximately homologous oscillations can find the resonance of the tiny variations in vibration, which are characteristic of the investigated failure types, as illustrated in Figure 7b. As a result, the WPD uses Daubechies 20 as the mother wavelet function to decompose each optimized vibration signal into 32 wavelet-based features for the HFP.
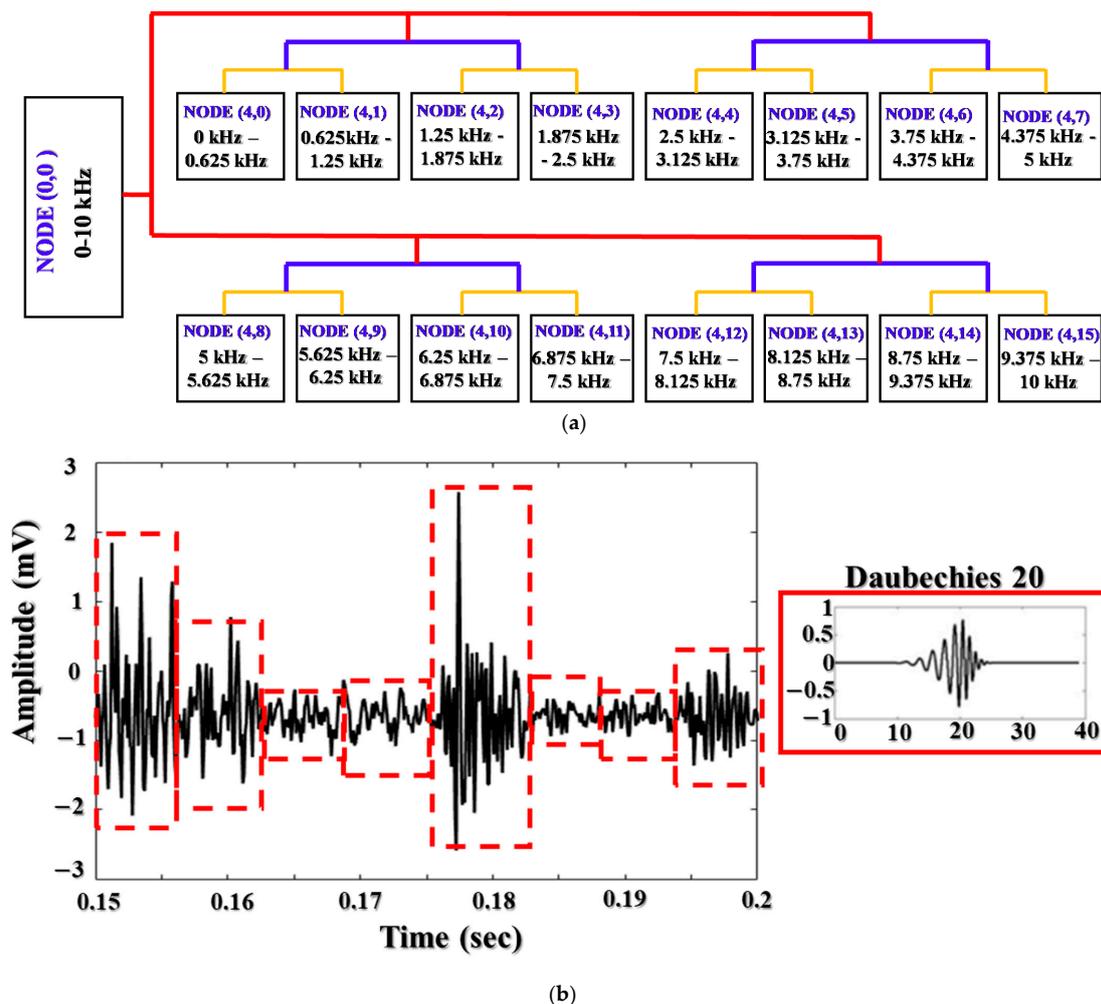
Figure 7. Four-level wavelet package decomposition (**a**) and the analogous oscillation patterns of a vibration signal and the Daubechies 20 function (**b**).

4.3. The Novel Distance Ratio Principal Component Analysis

Thus, multiple domain extraction from the output signal of the ANCT is used to configure the HFP in a high-dimensional feature space. The cardinality of the HFP is 123:21 statistical characteristics, 32 wavelet features, and 70 complex envelope features. However, the HFP contains redundant attributes that cannot be used to distinguish the different types of MTCF gearbox defects and therefore reduce the classification accuracy of the diagnostic model. To solve this problem, we propose a new DRPCA approach for choosing the discriminative features that express the fault type specifications. The DRPCA is processed in the following stages.

Stage 1. The average Euclidian distance (L_{in}) to features of the same failure type is calculated [45]:

$$L_{in} = \frac{1}{P.Q}\Sigma_{p=1}^{P}\Sigma_{q=1}^{Q}D_{Euc}(p,q); \tag{17}$$

Stage 2. According to [45], the average Euclidian distance (L_{out}) between a specific feature vector of failure type i and other feature vectors of failure type j ($i \neq j$, and $1 \leq i, j \leq$ T) is computed:

$$L_{out} = \frac{1}{P.T.Q}\Sigma_{t=1}^{T}\Sigma_{p=1}^{P}\Sigma_{q=1}^{Q}D_{Euc}(t,p,q); \tag{18}$$

Stage 3. The distance ratio by normalizing the quotient of the two distances attained from Equations (17) and (18) is determined:

$$R_d = \frac{\frac{L_{out}}{L_{in}}}{\max\left(\frac{L_{out}}{L_{in}}\right)}. \tag{19}$$

where $D_{Euc}(u,v) = \sqrt{\sum_{k=1}^{K}(u_k - v_k)^2}$ is the Euclidian distance between two feature vectors; P is the total number of features (P = 123, the dimension of the HFP); T denotes the number of failure types (T = 7); and Q signifies the observed sample count for each failure type. From Equation (19), R_d is stated in the range of [0, 1] and enlarges as L_{out} increases and L_{in} decreases. This tendency indicates the specification of discriminative features because it results in less variation within a single failure type and much higher variation across different types.

Stage 4. Apply a PCA (explained in Section 2.4) to search the features with $R_d \geq 0.9$:

$$DRPCA = PCA(optimal\ features(R_d \geq 0.9)). \tag{20}$$

The DRPCA provides a collection of the most advantageous features, which are the most discriminative features in the HFP, to configure a new feature pool, dubbed the *optimal feature set*.

4.4. Multi-Class Support Vector Machine to Identify Failure Conditions

SVMs were first developed to classify binary data on the basis of statistical processes and quadratic function learning theory. They work by looking for the special space with the biggest separation between two binary classes in the observed dataset [46].

We assume the need to train for a binary dataset with N samples of $\{(\alpha_n, \gamma_n),$ n = 1, 2, ... , N\}$, where $\alpha_n \in \mathbb{R}^Q$, Q is the dimension of a feature vector, and the category labels are γ_n ($\gamma_n \in \{-1, +1\}$). The special space, designated by w, can be obtained by increasing the margin width and decreasing the structural risk, as in the following expression [47]:

$$(w, b) = \underset{w,b}{\operatorname{argmin}} \frac{1}{2} w^T w + R \sum_{n=1}^{N} \zeta_n, \tag{21}$$

subject to $\gamma_n(w^T \Phi(x_n) + b) \geq 1 - \zeta_n$ and $\forall n = 1, 2, \ldots , N; -\zeta_n \leq 0, \forall n = 1, 2, \ldots , N$, where b denotes bias; $\zeta = \{\zeta_1, \zeta_2, \ldots, \zeta_N\}$ indicates a collection of slack variables; R is the trade-off coefficient; and $\Phi(x_n)$ is the expanded representation space of the feature vectors. By applying the Lagrange duality method [48] using Lagrange multipliers (μ_n, μ_k), the solution for Equation (21) can be transformed into the following:

$$\underset{\mu}{\operatorname{Argmax}} w(\mu) = \sum_{n=1}^{N} \mu_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{N} \mu_n \mu_k \gamma_n \gamma_k \Phi^T(\alpha_n) \Phi(\alpha_k) \tag{22}$$

subject to: $\sum_{n=1}^{N} \mu_n \gamma_n = 0, 0 \leq b_n \leq R, \forall n = 1, 2, \ldots , N$. Here, α_n and α_k are feature vectors of the input training data, which can be mapped into a new feature space with a larger dimensionality using the function $\mathcal{F}(\alpha_n, \alpha_k) = \Phi^T(\alpha_n)\Phi(\alpha_k)$, which is called a kernel function, such as a sigmoid, Gaussian, radial base, linear, or polynomial function. To categorize the dataset into several classes (this article uses seven failure classes), the initial binary SVM can be improved by using one of several network architectures, including hierarchical, one-against-all, and one-against-one. The OAOSVM architecture showed reliable classification capability [32,49]. Thus, we applied it to identify the MTCF failure types.

5. Performance Evaluation Results and Discussion

In this section, we evaluate the performance of the proposed diagnosis model for an MTCF gearbox under varying speed conditions through three key processes: signal processing, feature pool configuration, and discriminant feature–based failure type identification.

5.1. The Effectiveness of the ANCT Performance

Compared with conventional methods such as the HT, WT, window bandpass filter, and empirical mode decomposition, the ADT approach proposed in [18] demonstrates superior ability to process gearbox vibration signals to reduce noise and preserve useful fault information. Thus, the proposed ANCT was developed to enhance the denoising performance of the ADT by partitioning a vibration spectrum into many BFSs and then optimizing the vibration signal for each BFS using the segmented optimum NPS and filter coefficient set obtained from the ADT technique. To validate the efficacy of the ANCT method, the spectrum analysis of three vibration signals is shown in Figure 8: a raw vibration signal, an ADT output, and an ANCT output. In Figure 8, the dotted green circles represent items associated with failure-related frequency tones that were previously redeemed in the ADT and ANCT outputs. Furthermore, the ANCT has a considerably greater denoising capacity than the ADT; the noise regions in the ANCT output (shown by the dotted red circles) are far more degraded than those in the ADT output. As a result, ANCT outperforms ADT and other conventional signal processing techniques. In summary, the ANCT achieves considerable noise reduction while preserving the failure-related signatures, in this case the tooth meshing frequency harmonics and sideband frequency tones, in the raw signals. ANCT processed the vibration data of an MTCF gearbox with seven different failure categories over four different rotation speeds to provide optimized vibration signals. Those signals were then used to configure the HFP (which produces 123 features for each sample), as described in Section 4.2.
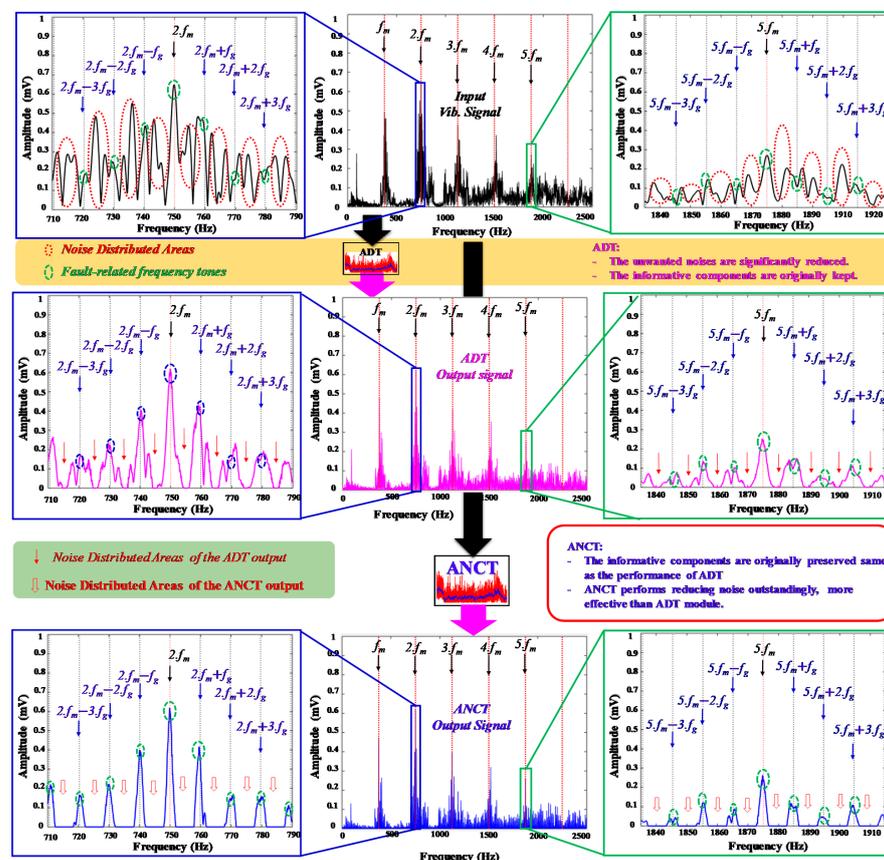

Figure 8. Visualized comparison of the ADT and ANCT performance.

5.2. DRPCA-Based Feature Selection and Classification Results

The applied diagnostic model uses a novel DRPCA procedure to capture the most discriminant features from the HFP in a smaller dimension. The PCA technique, as discussed in Section 2.4, is used to identify the importance of dynamically connected parameters or principal components, represented by a covariance matrix of eigenvectors. The PCA approach was used in this experiment to identify the attributes that produce primary elements with an eigenvalue greater than 80%. This value was chosen to capture the informative features that associated with the first 30 principal components, whose eigenvalues were greater than the remainder of the components. Experimental assessments in [28,50] reveal that the features with the greater eigenvalues contain the most fault-related information in the sequence of the all input features. The resulting principal components are then estimated using the distance ratios (according to the process given in Section 4.3). As a result of computing the DRPCA, features with substantial distance ratios ($R_d \geq 0.9$) are selected as the most discriminative features. The new optimal feature set is composed of three outcome features (designated Feature 1, Feature 2, and Feature 3), which resulted from a statistic model, wavelet decomposition, and complex envelope analysis, respectively. In Figure 9, a 3D figure based on those three optimal features illustrates the spatial distribution zones of samples of the seven MTCF failure states under the four different rotation speeds. As shown in Figure 9, the samples belonging to the same category are clustered together, and the samples belonging to different fault types are clearly separated in the feature space. The distance ratio thresholds are experimentally selected to obtain the number of the resultant features that are capable of representing failure types separately. The visualized spatial distribution of the categories in Figure 9 demonstrates that the threshold ($R_d = 0.9$) is an appropriate selection for configuring the optimal feature subset of the most discriminative features in this study. The DRPCA output is then used as the input data for the OAOSVM classifier, which categorizes the data according to the failure types (i.e., NF, F1, F2, F3, F4, F5, and F6).

The classification evaluation of the proposed model was implemented using two experiments. In the first experiment, the optimal feature sets of all the samples of the seven failure types under four rotation speed conditions were combined to create a new feature set with the size of 4200×3. That new feature set was then randomly split into a training set and a testing set using a ratio of 7:3, respectively, to investigate the general capability of the proposed model. The second experiment was run to assess the robustness of the proposed gearbox fault diagnosis method in identifying failure types under inconstant rotational speed conditions. Therefore, the speed-related datasets (i.e., optimal feature sets) were used differently for the training set (two speed-related datasets) and the testing set (a single speed-related dataset). For instance, a training set was created by merging two speed-related datasets (300 RPM and 600 RPM for a dimension of 2100×3), and then a testing set was configured using the 900 RPM dataset (for a dimensionality of 1050×3). The four speed-related datasets (300 RPM, 600 RPM, 900 RPM, and 1200 RPM) were alternately selected to configure the training and testing sets for the classifier, creating four classifying executions.

K-fold cross validation (KCV) was used to estimate the generalized performance [51] of the proposed classifier. The training set was randomly dispensed into K subsets (K = 10 in this study), and then the OAOSVM classified K times based on K subsets to estimate the average classification accuracy (some of the K subsets were used for the training process, and the rest were used for the validation process). The average classification accuracy of the training model with KCV was calculated using the outcomes of the proposed training model on a validation set:

$$C_{KCV} = \frac{1}{K} \Sigma_{k=1}^{K} \left(\Sigma_{t=1}^{T} \frac{S_{t,k}^{TP}}{S_k} \right) \times 100\% \tag{23}$$

where S_k is the whole quantity of samples in a validation set at the kth iteration; $S_{t,k}^{TP}$ is the number of true positives (i.e., samples of failure type t that are correctly assigned to category t).

The proposed fault diagnosis model was trained for various times (N) with the training set and testing sets of two experiments, and then the average identification accuracy (*AIA*) was summarized using Equation (24):

$$AIA = \frac{1}{N}\sum_{n=1}^{N}\frac{\sum_{t=1}^{T}S_{t,n}^{TP}}{S_n} \times 100\%, \qquad (24)$$

where the whole quantity of true positives for failure type t at the nth evaluation process ($S_{t,n}^{TP}$) and the total number of the testing set for the nth process (S_n) (the output results from the testing process) are used to calculate *AIA*. The proposed learning model architecture using an OAOSVM is depicted in Figure 10.
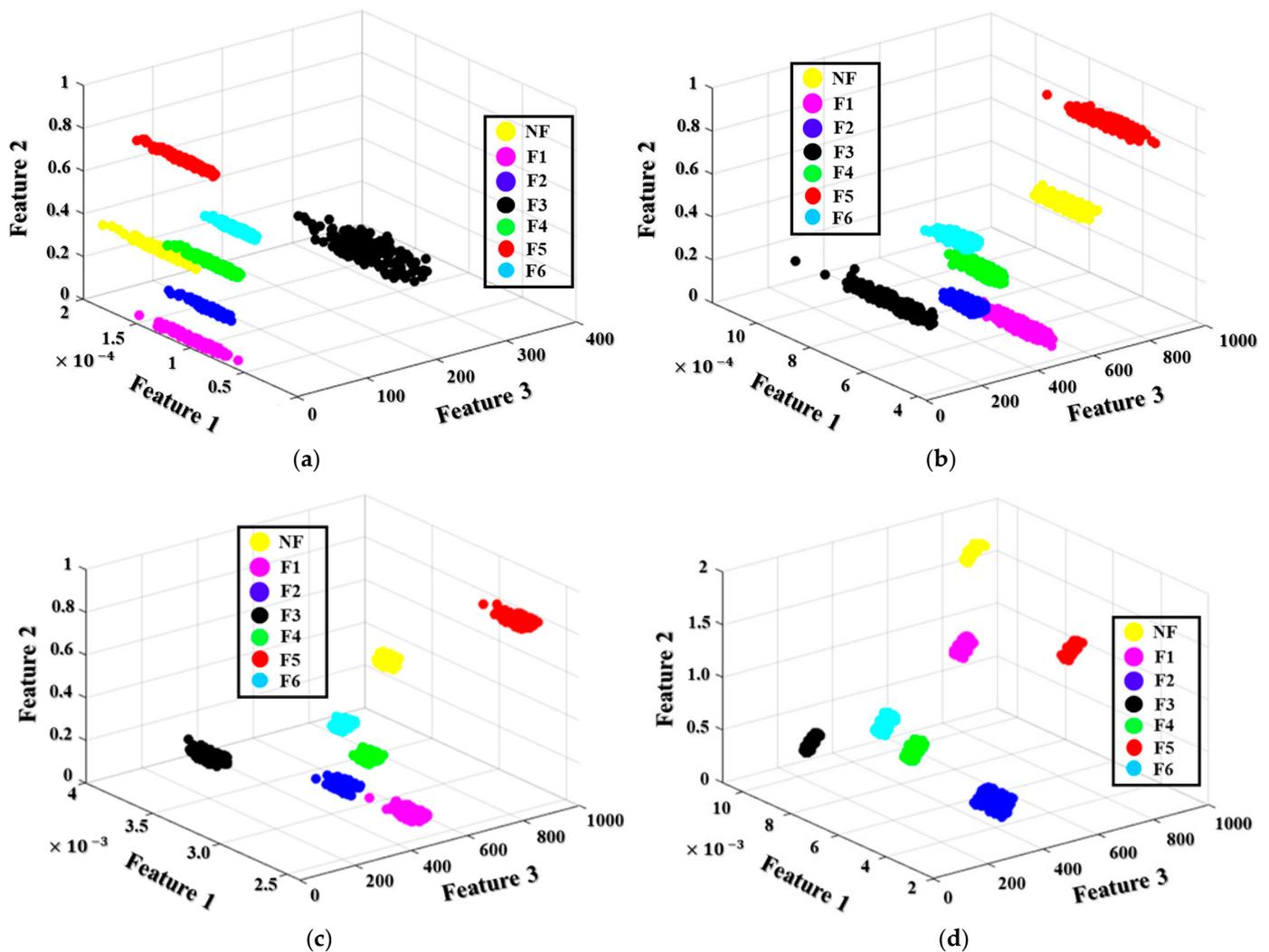

Figure 9. Three-dimensional visualization of samples of the seven failure types under four rotational speeds: (**a**) 300 RPM, (**b**) 600 RPM, (**c**) 900 RPM, and (**d**) 1200 RPM.
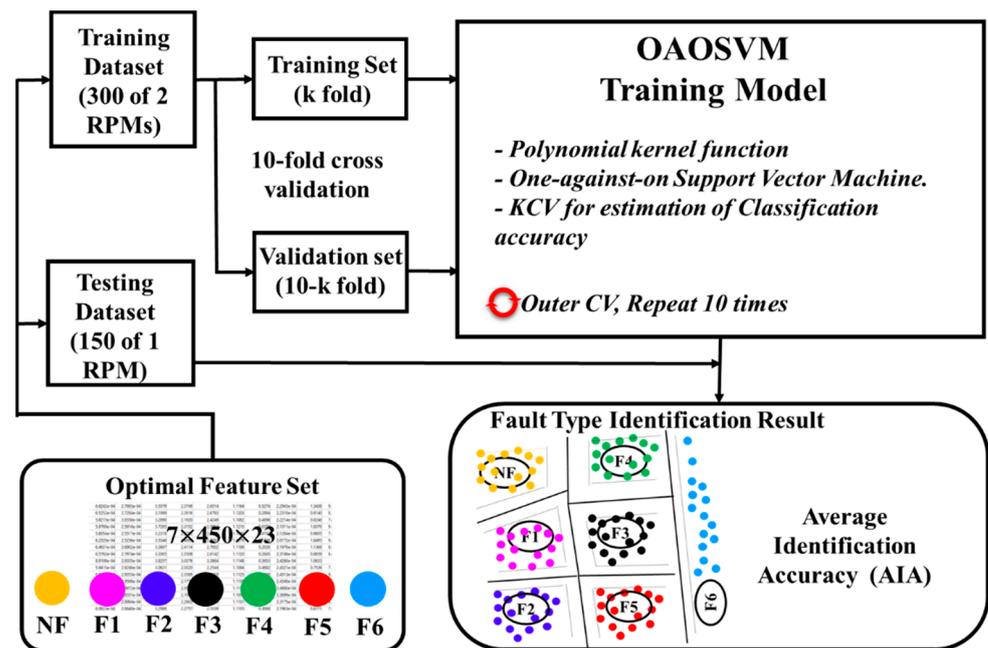
Figure 10. The learning model architecture using the one-against-one SVM.

To validate the efficacy of our novel fault diagnosis method, we compared its performance with those of conventional fault diagnosis models, including M1 (ADT + HFE + DRPCA + OAOSVM), M2 (ANCT + HFE + PCA + OAOSVM), M3 (ANCT + HFE + GA + OAOSVM), M4 (ANCT + HFE + ICA + OAOSVM), M5 (ANCT + HFE + LDA + OAOSVM), M6 (LHTIS + HFE + DRPCA + OAOSVM), and M7 (WSET + HFE + DR-PCA + OAOSVM). M1 is constructed by replacing the ACNT module of the proposed method with the ADT [18], which outperformed conventional signal processing methods for gearbox vibrations signals to explore the effect of noise reduction on the classification process. The other models were used to examine the ability of the DRPCA-based feature selection method by changing the proposed framework to use simple PCA (M2), GA [45] (M3), ICA (M4), and LDA (M5) instead of DRPCA. In addition, the other vibration signal analysis methods were applied as a replacement for the ANCT module in the proposed framework to establish models of M6 and M7. The local maximum high order time iterative synchro-squeezing (LHTIS) [52] and wavelet-based synchro-extracting transform (WSET) [53] techniques were used in M6 and M7, respectively.

The classification results of the proposed and reference models are tabulated in Table 4. The *AIA* results of M1 showing 14.31–24.37% are less than those of the proposed method. This is because the random disturbance noise elements in the vibration signals disorganize the feature engineering process and cause misclassification. ANCT provides better denoising performance than that specified in Section 5.1, demonstrating its high efficiency. In addition, when other signal processing approaches (LHTIS and WSET) are replaced with ANCT, the classification accuracies of the models are significantly lower than those of the proposed identification framework, showing 46.49–51.2% of M6 and 30.88–53.5% of M7. Furthermore, our DRPCA-based feature selection method outperforms M2 (13.75–31.9%), M3 (8.1–14.5%), M4 (17.79–30.18%), and M5 (17.83–36.57%), which verifies that the discriminant fault representation features of the optimal feature pool are very important for distinguishing the failure types of an MTCF gearbox under variable operating speeds. Overall, DRPCA is a suitable feature selection method for establishing the gearbox fault diagnosis model in this paper.

Table 4. Diagnostic performance results from the proposed and reference models in different experiments.

Experiment	Dataset (Samples)		Average Identification Accuracy (AIA%)							
	Training Set	Testing Set	Proposed	M1	M2	M3	M4	M5	M6	M7
Experiment 1	2940 (70% of dataset)	1260 (30% of dataset)	100	75.63	72.79	85.50	81.68	70.18	62.70	68.54
Experiment 2	2100 (300 RPM, 600 RPM)	1050 (900 RPM)	100	80.31	72.54	86.65	69.82	63.43	63.51	46.50
	2100 (600 RPM, 900 RPM)	1050 (1200 RPM)	100	79.17	81.87	89.51	72.85	75.31	61.39	51.29
	2100 (900 RPM, 1200 RPM)	1050 (300 RPM)	100	81.71	86.25	87.30	81.21	67.62	48.80	69.12
	2100 (1200 RPM, 300 RPM)	1050 (600 RPM)	100	85.69	68.10	91.90	75.49	82.17	52.10	55.37

5.3. Discussion

The proposed model outperforms conventional state-of-the-art methods for identifying the failure types of an MTCF gearbox operating in inconstant speed conditions, yielding an average identification accuracy of 100% through two experiments. Denoising complex gearbox vibration data using a robust technique is critical for effective condition monitoring systems. Disturbance noise has a detrimental effect on feature engineering and identification performance. We developed our novel fault diagnosis model for an MTCF gearbox operating at changing speeds using ANCT for denoising, HFP and DRPCA for configuring the optimal feature set, and an OAOSVM classifier. During research and testing, our proposed approach achieved the highest classification performance, demonstrating its applicability.

In order to precisely identify the MTCF failure types in a gearbox system, the disturbance noise in the non-linear and non-stationary vibration signals should be eliminated in the first stage. As illustrated in Figure 4, the noise frequency tones are distributed in the whole range of the frequency spectrum, especially in the vicinity of the defect frequency tones. The high noise amplitudes might lead to misclassification of the failure categories and a deterioration in fault identification accuracy. The key technique of the ADT and ANCT methods is to use the adjustable Gaussian windows to adaptively access a zone between two consecutive sideband frequency tones along whole vibration spectrum to remove noise elements. Experimental results show that the ADT and ANCT outperforms the other conventional signal analysis approaches in [18] as well as M6 and M7 in reducing noise and preserving the failure-related ingredients. In addition, the proposed ANCT outperforms ADT in term of noise elimination, as shown in Figure 8. The next step of the fault diagnosis development process is the feature engineering and classification approach. In this paper, manual feature extraction and feature selection techniques are utilized to identify the most discriminant features that can be used as input to the machine learning–based classification algorithm. Experimental results show that the combination (HEF + DRPCA + OAOSVM) is the optimal solution for identifying seven MTCF gearbox failure types under different speed conditions, as shown in Table 4.

On the other hand, the deep neural network architectures are also applied for automated feature extraction, feature selection, and classification in the posterior phase of the signal analysis. The sensitive and stable fault diagnosis frameworks are proposed by combining ADT and deep learning architecture; the stacked spare autoencoder-based deep neural network (SSA-DNN) in [19] (this model is referred to as model M8) and the deep convolutional neural network (DCNA) in [33] (M9) provide successful results (100%) for the identification seven failure types of an MTCF gearbox under variable speed operations. However, the computational time of the deep learning architectures in M8 and M9 requires more time than that of the proposed model (ANCT + HFE + DRPCA + OAOSVM), as shown in Figure 11.
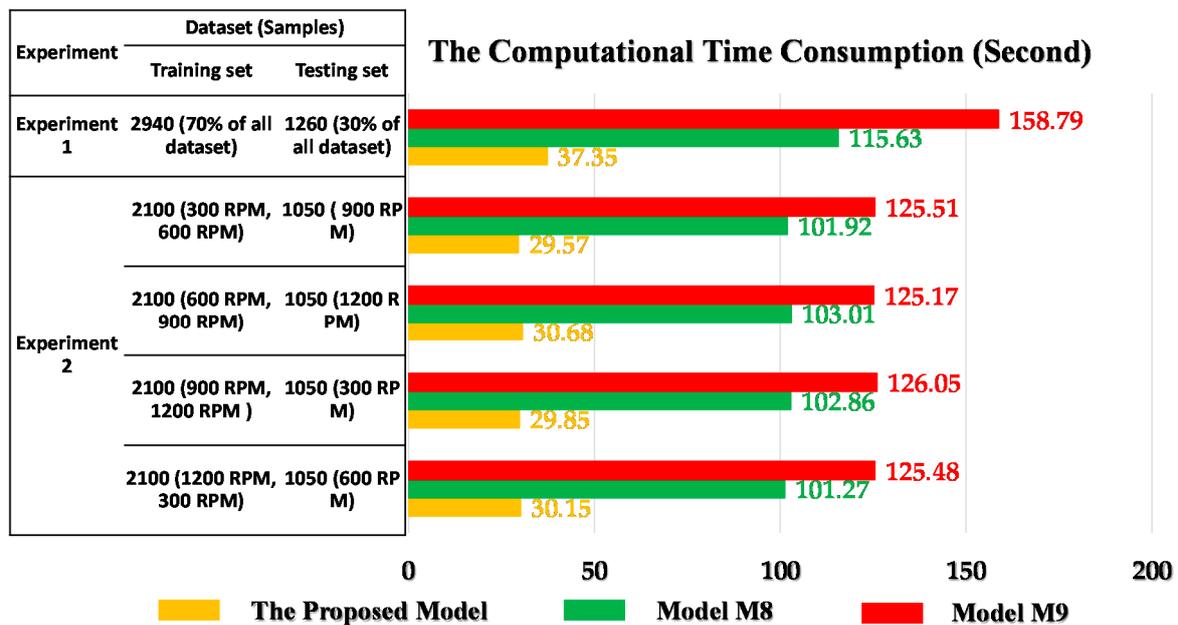
| Experiment | Dataset (Samples) | | The Computational Time Consumption (Second) |
	Training set	Testing set	
Experiment 1	2940 (70% of all dataset)	1260 (30% of all dataset)	158.79 / 115.63 / 37.35
Experiment 2	2100 (300 RPM, 600 RPM)	1050 (900 RPM)	125.51 / 101.92 / 29.57
	2100 (600 RPM, 900 RPM)	1050 (1200 RPM)	125.17 / 103.01 / 30.68
	2100 (900 RPM, 1200 RPM)	1050 (300 RPM)	126.05 / 102.86 / 29.85
	2100 (1200 RPM, 300 RPM)	1050 (600 RPM)	125.48 / 101.27 / 30.15

The Proposed Model   Model M8   Model M9

Figure 11. Visual expression of the computational time in the comparison of three models: the proposed model, M8, and M9.

6. Conclusions

In this paper, a novel model was proposed for identifying the fault types of multi-degree tooth-cut failures (MTCFs) in a gearbox system operating under inconsistent speed conditions.

(1) Vibration signals acquired from gearboxes have noisy components that dominate and distort the failure-related signal information, such as meshing frequency harmonics and sideband frequencies. To extract this information, a novel adaptive noise reduction method, ANCT, was proposed. ANCT uses an adaptive method to identify the appropriate Gaussian parameters for each segment of the vibration frequency domain. ANCT significantly reduces the noise in the vibration signal while retaining maximal failure-related information.

(2) Since multi-level tooth-cut failures are essentially the same type of faults that only differ in their relative size, a heterogeneous feature pool was constructed by calculating more than a hundred statistical parameters of the denoised vibration signal in multiple domains using wavelet packet decomposition and complex envelope decomposition to ensure the collection of maximal information on each type of MTCF fault.

(3) To reduce the dimensionality of the feature pool and select the most discriminative features for identifying the MTCF faults, a novel feature selection method, DRPCA, was proposed. DRPCA combines principal component analysis with relative distance ratio analysis of features of different fault types. The optimal feature set is constructed by selecting features with the highest relative distance ratio. This provides lower dimensionality, thereby improving the diagnostic performance of the OAOSVM, which was employed as a classifier.

(4) Finally, the performance of the proposed methodology was evaluated using a real-world experimental testbed and two different experiments based on the operational speed, where the vibration data were collected. In the first experiment, datasets for all the operational speeds were merged into a single set. Then, training and test subsets were constructed by randomly collecting features from this set. In the second experiment, the training dataset consisted of data recorded with one speed, while the testing dataset consisted of data recorded with other speeds. The proposed model outperformed the state-of-the-art approaches, with an average identification accuracy

of 100% in both experiments. Moreover, our proposed model showed three times the speed-up over the relevant models, including the deep neural architectures.

Nomenclature

ANCT	adaptive noise-canceling technique
MTCF	multi-degree tooth-cut failures
HFP	heterogeneous feature pool
EMD	empirical mode decomposition
ADT	adaptive denoising technique
LDA	linear discriminant analysis
SVM	support vector machines
WPT	wavelet packet transform
KCV	K-fold cross validation
DRPCA	distance ratio principal component analysis
PCA	principle component analysis
HT	Hilbert transform
WT	wavelet transform
ICA	independent component analysis
GA	genetic algorithm
OAOSVM	one-against-one SVM
DAQ	data acquisition
AIA	average identification accuracy

References

1. León, R.; Montaleza, C.; Maldonado, J.L.; Tostado-Véliz, M.; Jurado, F. Hybrid Electric Vehicles: A Review of Existing Configurations and Thermodynamic Cycles. *Thermo* **2021**, *1*, 134–150. [CrossRef]
2. Nie, M.; Wang, L. Review of Condition Monitoring and Fault Diagnosis Technologies for Wind Turbine Gearbox. *Procedia CIRP* **2013**, *11*, 287–290. [CrossRef]
3. 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]
4. Alban, L.E. Failures of Gears. In *Failure Analysis and Prevention*; William, T., Becker, R.J.S., Eds.; ASM International: Materials Park, OH, USA, 2002; Volume 11.
5. Sait, A.S.; Sharaf-Eldeen, Y.I. A Review of Gearbox Condition Monitoring Based on Vibration Analysis Techniques Diagnostics and Prognostics. In *Proceedings of the Conference Society for Experimental Mechanics Series*; Springer: New York, NY, USA, 2011; Volume 5, pp. 307–324.
6. 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.
7. Zhang, R.; Gu, X.; Gu, F.; Wang, T.; Ball, A.D. Gear Wear Process Monitoring Using a Sideband Estimator Based on Modulation Signal Bispectrum. *Appl. Sci.* **2017**, *7*, 274. [CrossRef]

8. Zhang, M.; Wang, K.S.; Wei, D.; Zuo, M.J. Amplitudes of Characteristic Frequencies for Fault Diagnosis of Planetary Gearbox. *J. Sound Vib.* **2018**, *432*, 119–132. [CrossRef]
9. Podulka, P. Selection of Methods of Surface Texture Characterisation for Reduction of the Frequency-Based Errors in the Measurement and Data Analysis Processes. *Sensors* **2022**, *22*, 791. [CrossRef]
10. Patil, C.R.; Kulkarni, P.P.; Sarode, N.N.; Shinde, K.U. Gearbox Noise & Vibration Prediction and Control. *Int. Res. J. Eng. Technol.* **2017**, *4*, 873–877.
11. Zhang, X.; Xu, G.; Kuang, J.; Suo, L.; Zhang, S.; Khalique, U. A Three-Phase Current Tacholess Envelope Order Analysis Method for Feature Extraction of Planetary Gearbox under Variable Speed Conditions. *Sensors* **2021**, *21*, 5714. [CrossRef]
12. La, D.; Blas, C.; Lopez-Martin, C.A.; Delgado-Prieto, M.; Rios, R.O.; Picot, A.; Martincorena-Arraiza, M.; De, C.A.; Blas, L.C.; Lopez-Martin, A.; et al. Fault Detection of Planetary Gears Based on Signal Space Constellations. *Sensors* **2022**, *22*, 366. [CrossRef]
13. Gearbox Fault, Y.; Khodaei, S.; Dao, P.B.; Qiu, L.; Yu, L.; Guo, Y.; Jiang, S.; Yang, Y.; Jin, X.; Wei, Y. Gearbox Fault Diagnosis Based on Improved Variational Mode Extraction. *Sensors* **2022**, *22*, 1779. [CrossRef]
14. Roshanmanesh, S.; Hayati, F.; Papaelias, M. Utilisation of Ensemble Empirical Mode Decomposition in Conjunction with Cyclostationary Technique for Wind Turbine Gearbox Fault Detection. *Appl. Sci.* **2020**, *10*, 3334. [CrossRef]
15. Guo, J.; Shi, Z.; Li, H.; Zhen, D.; Gu, F.; Ball, A.D. Early Fault Diagnosis for Planetary Gearbox Based Wavelet Packet Energy and Modulation Signal Bispectrum Analysis. *Sensors* **2018**, *18*, 2908. [CrossRef] [PubMed]
16. Fan, X.; Zuo, M.J. Gearbox Fault Detection Using Hilbert and Wavelet Packet Transform. *Mech. Syst. Signal Processing* **2006**, *20*, 966–982. [CrossRef]
17. Yang, Q.; An, D. EMD and Wavelet Transform Based Fault Diagnosis for Wind Turbine Gear Box. *Adv. Mech. Eng.* **2013**, *5*, 212836. [CrossRef]
18. 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]
19. 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]
20. Caesarendra, W.; Kosasih, B.; Tieu, K.; Moodie, C.A.S. An Application of Nonlinear Feature Extraction—A Case Study for Low Speed Slewing Bearing Condition Monitoring and Prognosis. In Proceedings of the 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Wollongong, NSW, Australia, 9–12 July 2013; pp. 1713–1718. [CrossRef]
21. Zhu, H.; He, Z.; Wei, J.; Wang, J.; Zhou, H. Bearing Fault Feature Extraction and Fault Diagnosis Method Based on Feature Fusion. *Sensors* **2021**, *21*, 2524. [CrossRef]
22. Liu, R.; Zhao, Z.; Kundu, P.; Antonino-Daviu, J.A.; Liu, Z.; Ding, K.; Lin, H.; He, G.; Du, C.; Chen, Z. A Novel Impact Feature Extraction Method Based on EMD and Sparse Decomposition for Gear Local Fault Diagnosis. *Machines* **2022**, *10*, 242. [CrossRef]
23. Duan, J.; Shi, T.; Zhou, H.; Xuan, J.; Zhang, Y. Multiband Envelope Spectra Extraction for Fault Diagnosis of Rolling Element Bearings. *Sensors* **2018**, *18*, 1466. [CrossRef]
24. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [CrossRef]
25. Zuo, M.J.; Lin, J.; Fan, X. Feature Separation Using ICA for a One-Dimensional Time Series and Its Application in Fault Detection. *J. Sound Vib.* **2005**, *287*, 614–624. [CrossRef]
26. Li, W.; Zhang, L.; Xu, Y. Gearbox Pitting Detection Using Linear Discriminant Analysis and Distance Preserving Self-Organizing Map. In Proceedings of the 2012 IEEE I2MTC—International Instrumentation and Measurement Technology Conference Proceedings, Graz, Austria, 13–16 May 2012; pp. 2225–2229.
27. Samanta, B. Gear Fault Detection Using Artificial Neural Networks and Support Vector Machines with Genetic Algorithms. *Mech. Syst. Signal Processing* **2004**, *18*, 625–644. [CrossRef]
28. Abouhnik, A.; Ibrahim, G.R.; Shnibha, R.; Albarbar, A. Novel Approach to Rotating Machinery Diagnostics Based on Principal Component and Residual Matrix Analysis. *ISRN Mech. Eng.* **2012**, *2012*, 1–7. [CrossRef]
29. Sakthivel, N.R.; Nair, B.B.; Elangovan, M.; Sugumaran, V.; Saravanmurugan, S. Comparison of Dimensionality Reduction Techniques for the Fault Diagnosis of Mono Block Centrifugal Pump Using Vibration Signals. *Eng. Sci. Technol. Int. J.* **2014**, *17*, 30–38. [CrossRef]
30. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. *Mech. Syst. Signal Processing* **2018**, *108*, 33–47. [CrossRef]
31. Liu, J.; Zio, E. Feature Vector Regression with Efficient Hyperparameters Tuning and Geometric Interpretation. *Neurocomputing* **2016**, *218*, 411–422. [CrossRef]
32. Hsu, C.W.; Lin, C.J. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [CrossRef]
33. 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. [CrossRef]
34. Lee, K.A.; Gan, W.S.; Kuo, S.M. *Subband Adaptive Filtering: Theory and Implementation*; John Wiley and Sons: Chichester, UK, 2009; ISBN 9780470516942.

35. Maniruzzaman, M.; Billah, K.M.S.; Biswas, U.; Gain, B. Least-Mean-Square Algorithm Based Adaptive Filters for Removing Power Line Interference from ECG Signal. In Proceedings of the 2012 International Conference on Informatics, Electronics & Vision (ICIEV), Dhaka, Bangladesh, 18–19 May 2012; Volume 2012, pp. 737–740. [CrossRef]

36. Yan, R.; Gao, R.X.; Chen, X. Wavelets for Fault Diagnosis of Rotary Machines: A Review with Applications. *Signal Processing* **2014**, *96*, 1–15. [CrossRef]

37. Yan, R.; Gao, R.X. Multi-Scale Enveloping Spectrogram for Vibration Analysis in Bearing Defect Diagnosis. *Tribol. Int.* **2009**, *42*, 293–302. [CrossRef]

38. Zhang, G.; Tang, L.; Zhou, L.; Liu, Z.; Liu, Y.; Jiang, Z. Principal Component Analysis Method with Space and Time Windows for Damage Detection. *Sensors* **2019**, *19*, 2521. [CrossRef] [PubMed]

39. 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]

40. Chaari, F.; Bartelmus, W.; Zimroz, R.; Fakhfakh, T.; Haddar, M. Gearbox Vibration Signal Amplitude and Frequency Modulation. *Shock Vib.* **2012**, *19*, 635–652. [CrossRef]

41. 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.

42. 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]

43. Rafiee, J.; Tse, P.W.; Harifi, A.; Sadeghi, M.H. A Novel Technique for Selecting Mother Wavelet Function Using an Intelli Gent Fault Diagnosis System. *Expert Syst. Appl.* **2009**, *36*, 4862–4875. [CrossRef]

44. 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]

45. Nguyen, C.D.; Prosvirin, A.; Kim, J.-M. Fault Identification of Multi-Level Gear Defects Using Adaptive Noise Control and a Genetic Algorithm. In *Proceedings of the Intelligent Human Computer Interaction*; Singh, M., Kang, D.-K., Lee, J.-H., Tiwary, U.S., Singh, D., Chung, W.-Y., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 325–335.

46. Santos, P.; Villa, L.F.; Reñones, A.; Bustillo, A.; Maudes, J. An SVM-Based Solution for Fault Detection in Wind Turbines. *Sensors* **2015**, *15*, 5627–5648. [CrossRef]

47. Manjurul Islam, M.M.; Kim, J.M. Reliable Multiple Combined Fault Diagnosis of Bearings Using Heterogeneous Feature Models and Multiclass Support Vector Machines. Reliab. *Eng. Syst. Saf.* **2019**, *184*, 55–66. [CrossRef]

48. Widodo, A.; Kim, E.Y.; Son, J.D.; Yang, B.S.; Tan, A.C.C.; Gu, D.S.; Choi, B.K.; Mathew, J. Fault Diagnosis of Low Speed Bearing Based on Relevance Vector Machine and Support Vector Machine. *Expert Syst. Appl.* **2009**, *36*, 7252–7261. [CrossRef]

49. Tomar, D.; Agarwal, S. A Comparison on Multi-Class Classification Methods Based on Least Squares Twin Support Vector Machine. *Knowl.-Based Syst.* **2015**, *81*, 131–147. [CrossRef]

50. Cao, S.; Hu, Z.; Luo, X.; Wang, H. Research on Fault Diagnosis Technology of Centrifugal Pump Blade Crack Based on PCA and GMM. *Measurement* **2020**, *173*, 108558. [CrossRef]

51. Rodríguez, J.D.; Pérez, A.; Lozano, J.A. Sensitivity Analysis of K-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575. [CrossRef] [PubMed]

52. Li, Y.; Shi, Z.; Lin, T.R.; Yu, G. An Iterative Reassignment Based Energy-Concentrated TFA Post-Processing Tool and Application to Bearing Fault Diagnosis. *Measurement* **2022**, *193*, 110953. [CrossRef]

53. Shi, Z.; Yang, X.; Li, Y.; Yu, G. Wavelet-Based Synchroextracting Transform: An Effective TFA Tool for Machinery Fault Diagnosis. *Control Eng. Pract.* **2021**, *114*, 104884. [CrossRef]