

Article

# Application of Hybrid Model between the Technique for Order of Preference by Similarity to Ideal Solution and Feature Extractions for Bearing Defect Classification

Chun-Yao Lee <sup>1,\*</sup> , Truong-An Le <sup>2</sup>  and Chung-Yao Chang <sup>1</sup><sup>1</sup> Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan 320314, Taiwan<sup>2</sup> Department of Electrical and Electronic Engineering, Thu Dau Mot University, Thu Dau Mot 75000, Vietnam\* Correspondence: [cyl@cycu.edu.tw](mailto:cyl@cycu.edu.tw)

**Abstract:** This paper describes a development that offers new opportunities for detecting faulty bearings. Prioritization is based on the technique for order of preference by similarity to the ideal solution (TOPSIS) for the most discriminative features in the faulty bearing dataset. The proposed model is divided into three steps: feature extraction, feature selection, and classification. In feature extraction, variational mode decomposition (VMD) and fast Fourier transform (FFT) are used to extract features from the measured signal of the test motors and use the symmetrical uncertainty (SU) value for calculation, reducing the redundancy of data. In terms of feature selection, the TOPSIS method is used instead of the traditional filtering method, which is applied to analysis and decision making, and important features are selected from seven filtering methods. Finally, in order to validate the classification ability of the proposed model, k-nearest neighbors (KNN), support vector machine (SVM), and artificial neural networks (ANN) are used as independent classifiers. The effectiveness of the proposed model is evaluated by applying two bearing datasets, namely the current dataset of motor vibration signals and the dataset of bearing motors provided by Case Western Reserve University (CWRU). The results show that the comparison of the proposed model with other models shows the feasibility of this study.

**Keywords:** bearing fault diagnosis; feature selection; TOPSIS; feature extraction**MSC:** 68T07

**Citation:** Lee, C.-Y.; Le, T.-A.; Chang, C.-Y. Application of Hybrid Model between the Technique for Order of Preference by Similarity to Ideal Solution and Feature Extractions for Bearing Defect Classification.

*Mathematics* **2023**, *11*, 1442. <https://doi.org/10.3390/math11061442>

Academic Editors: Heung Soo Kim, Salman Khalid, Ananda Shankar and Prashant Kumar

Received: 20 February 2023

Revised: 13 March 2023

Accepted: 14 March 2023

Published: 16 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Rolling bearings are an important part of industrial machinery and equipment and have been widely used in this field. Rolling bearings need to bear the weight of rotating machines and ensure normal operation [1]. However, the bearing suffers vibration, wear, corrosion, fatigue, etc., during operation. When the bearing fails, the vibration will increase when it is fatigued, and its working efficiency will be reduced and even cause casualties. Therefore, bearing fault detection can sense the subtle vibration of the bearing, and the running state of the bearing can be judged through simple data, which can be seen at a glance [2–4]. Fault vibration of rolling bearings is caused by two kinds of faults: local faults and distributed faults. The ultimate goal of vibration monitoring is to know when the bearing needs to be replaced by tracking the state of the bearing. According to statistics, about 30% to 40% of mechanical failures in rotating equipment using rolling bearings are caused by rolling bearings [5,6]. Therefore, early fault diagnosis is an important topic in a wide range of applications to improve the safety and reliability of rotating electrical machines [7]. The purpose of this study is to propose an efficient method for bearing fault diagnosis based on machine learning. The bearing fault diagnosis model includes five steps: raw data acquisition, data preprocessing, feature extraction, feature selection, and fault mode recognition [8]. When the rolling bearing fails, it sends out vibration and impact

signals, and the collection of vibration signals can directly reflect the operating status of the faulty mechanical equipment. With the technical improvement of feature extraction methods, the dimension or category of extracted feature vectors is increasing, and there are more and more irrelevant and redundant feature vectors in high-dimensional feature sets, which may affect the accuracy rate. Therefore, some dimensionality reduction strategies are used in feature extraction to select sensitive features, which will also affect the diagnosis results as computational efficiency. There are several feature extraction methods [9,10]: the more famous ones are envelope analysis, such as principal component analysis [11,12], distance estimation techniques, symmetric uncertainty (SU) methods [13], and empirical mode decomposition (EMD) [14]; variational mode decomposition (VMD) [15]; and fast Fourier transform (FFT) methods [16,17]. EMD technology has high precision in signal extraction and fast convergence speed, which is very suitable for bearing fault diagnosis. However, in extracting the signal, VMD can set the desired mode number and avoid the endpoint effect similar to EMD decomposition through mirror extension [18]. After the feature extraction step, the feature set has been preliminarily formed, but there are still many redundant features in the extracted datasets, especially in high-dimensional datasets. These redundant features lead to a decrease in the accuracy of the final classification. Therefore, a feature selection step is required, which is an intermediate step between data extraction and classification. What it does is preprocess the dataset to further select its features before sending it to the classifier. Feature selection has flourished, and algorithms have gained attention in solving dataset optimization problems. Regarding fault classification methods, some artificial intelligence methods [19] include support vector machine (SVM) [20,21], k-nearest neighbors (KNN) [22], fuzzy inference system [23], random forest (RF) [24], and artificial neural network (ANN) [25–30]. Ultimately, bearing health depends on these technologies.

Since a single feature selection method cannot provide excellent classification accuracy in the case of different faulty bearings, to improve the performance of a new method for bearing defect classification, the proposed method relies on an ideal solution with an order preference similar to the ideal technical solution (TOPSIS) [31]. The information in the original data can be fully utilized, and the results can accurately reflect the gap between evaluation schemes. The basic process is to use the cosine method to find the optimal solution and the worst solution in the finite solution based on the normalized original data matrix and then calculate the distances between each evaluation object and the optimal solution and the worst solution, respectively, and calculate the relative closeness of each evaluation object to the optimal solution as the basis for evaluating the advantages and disadvantages. Although TOPSIS avoids the subjectivity of data, has no strict requirements on data components and samples, and is more flexible in describing the comprehensive impact of multiple indicators, it needs the data of each indicator, and it is difficult to choose the corresponding quantitative weight. It needs to be determined in order to accurately describe the impact of weights. The purpose of this research is to try to apply the TOPSIS method to the feature selection of filters and try to improve the lack of filter classification accuracy. For the stability and accuracy of the final feature extraction, this study uses the SU method to delete redundant features and combines six feature selection methods for weight distribution. For specific research, six typical feature selection methods are considered, which are important for TOPSIS. The construction of the method is challenging, because not every filter method is suitable for different data, and these filter methods are combined by the TOPSIS method, which will lower the recognition rate. The number of indicators selected is appropriate, which can describe the influence of the indicators well and overcome the blind spots of the TOPSIS method. This study utilizes multiple methods, namely, RF [24], minimum redundancy maximum relevance (mRMR) [32], correlation-based feature selection (CFS) [33], F-score (FS) [34], Pearson correlation coefficient (PCC) [35], and item variance (TV) [33]; in addition, we also utilize three well-known classifications, the SVM [20,21], KNN [22], and ANN [26–30]. After extracting the original data by VMD-FFT, the correlation coefficient is calculated by the SU method to reduce the uncertainty

characteristics. The selection filtering method is used as the selection weight because the calculation speed of this method is very fast and can be applied to various machine learning models.

In this paper, an optimal feature selection method in TOPSIS is proposed, which also removes redundant features to make the method more refined. The optimization trade-off between detection accuracy, processing speed, and flexibility is the main consideration of this method. The results through the classifier show that the best features can be selected quickly and provide stable and real-time performance.

The main contributions of this paper are summarized as follows:

- (1) In the research of filtering methods, combining the advantages of different filtering methods makes the accuracy rate more stable and effectively removes irrelevant and redundant functions.
- (2) We introduce a new orientation-adaptive feature extraction method. This paper proposes a fault diagnosis method based on TOPSIS. Compared with the traditional signal method, this method takes advantage of the combination of multiple feature extractions to fill in the order of the newly sorted features, thereby improving the stability of feature recognition.
- (3) In this paper, a hybrid model for rolling bearing fault identification is established, and the model is tested using existing equipment and fault data under different conditions to achieve accurate fault diagnosis and compare the currently available algorithms. The proposed model has a lower computational cost.

The structure of this paper is as follows: Section 2 describes the basics and workflow involved in the TOPSIS hybrid model. Section 3 introduces the method for measuring motor signals in the bearing motor dataset and the Case Western Reserve University (CWRU) dataset. In Section 4, the results of three classifiers in four different models of the two faulty bearing datasets in Section 3 are discussed, and the information and performance are analyzed and evaluated in detail. Finally, the optimal model is determined based on these values.

## 2. Hybrid Models

### 2.1. Variational Mode Decomposition (VMD)

Variational pattern decomposition has previously been published (Dragomiretskiy, 2014). VMD is a novel adaptive nonrecursive signal decomposition methods for enhancing sequence stability [15]. In addition to VMD, it can specify the number of modes desired for the outcome. The IMF decomposed by its method has an independent center frequency and shows the characteristics of sparsity in the frequency domain, which has the characteristics of sparse research. In the process of solving the IMF, the endpoint effect similar to that in the EMD decomposition is avoided by means of mirror extension. Selecting an appropriate value of  $K$  can effectively avoid modal aliasing. This study used VMD to decompose the ball bearing test data series for normal and faulty bearings. The VMD structure is as follows: For VMD, the preset  $K$  value determines the number of IMF components. The sum of the IMFs is the original signal. The constraints are expressed by (1) and (2):

$$\min \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega t} \right\|^2 \tag{1}$$

$$\text{s.t. } \sum_k u_k = f \tag{2}$$

where  $\{u_k\} = \{u_{\text{IMF1}}, \dots, u_{\text{IMFK}}\}$  contains local characteristic signals of different time scales of the original signal;  $\{w_k\} = \{w_1, \dots, w_k\}$  represents the center frequency of each IMF component;  $\sum_k u_k$  represents the sum of all modal components; and  $f$  represents the time series of ball bearing test data for decomposed normal and faulty bearings.

VMD solution: a secondary penalty factor  $\alpha$  and the introduction of a Lagrangian penalty operator  $\lambda(t)$  to transform the constrained variational problem given by (1) and (2) into an unconstrained variational problem where the extension of the Lagrangian expression is given in (3) below:

$$L \{u_k\} \{w_k\}, \lambda(t) = \alpha \sum_k \|\partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega t}\|_2^2 \tag{3}$$

The Alternating Direction Multiplier Algorithm [2] is used to solve the variational problem given by (3), above which produces the alternately updated  $\overline{u_k^{n+1}}(w)$  and  $w_k^{n+1}$  expressions given in (4) and (5):

$$\overline{u_k^{n+1}}(w) = \frac{\overline{f}(w) - \sum_{i \neq k} \overline{u_i}(w) + \overline{\lambda}(w)/2}{1 + 2\alpha(w - w_k)^2} \tag{4}$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\overline{u_k}(w)|^2 dw}{\int_0^\infty |\overline{u_k}(w)|^2 dw} \tag{5}$$

where  $\overline{u_k^{n+1}}(w)$  is the Wiener filter of  $\overline{f}(w) - \sum \overline{u_i}(w)$  that yields  $w_k^{n+1}$ , the corresponding power spectrum of the centroid modal function.

The VMD model is as follows:

Step 1: Initialize  $u_k, w_k, \lambda$  and  $n = 0$ .

Step 2:  $n = n + 1$  (number of iterations).

Step 3: Update  $u_k$  and  $w_k$  according to the VMD algorithm formula.

Step 4: Update the Lagrange multiplier  $\lambda$  according to the relevant algorithm.

Step 5: Know until a certain condition is met (judged by the similarity coefficient), stop the iteration; otherwise, go to Step 2.

Step 6: As  $k = k + 1$ , subtract the decomposed mode from the source signal and use it as the source signal for the next cycle; go to Step 1.

### 2.2. Fast Fourier Transform (FFT)

FFT [16,17] is based on the mathematics of discrete Fourier transform (DFT), which utilizes the symmetric properties of discrete Fourier transform complex multiplication on the complex plane, and multiple multiplications with symmetric properties are combined into one item, so it can effectively reduce the number of mathematical calculations and obtain more efficient calculations without changing the original mathematical model structure. This calculation method was proposed by Cooley and Tukey in 1965. In the field of digital signal processing, the data can be converted from the time domain waveform to the frequency spectrum through the Fourier transform. The fast Fourier transform model is defined as follows:

DFT is obtained by decomposing a series of values into components of different frequencies. The definition of DFT is expressed as (6):

$$X(k) \equiv X(e^{\frac{j2\pi k}{N}}) = \sum_{n=0}^{N-1} x(n)W_N^{kn}, k = 0, 1, \dots, N - 1 \tag{6}$$

It is further expressed as the sum of the following odd-numbered terms plus the sum of the even-numbered terms.

$$X(k) = \sum_{r=0}^{N/(2-1)} x(2r)(W_N^2)^{rk} + \sum_{r=0}^{N/(2-1)} X(2r+1)(W_N^2)^{rk} = G(k) + W_N^k \cdot H(k) \tag{7}$$

In  $W_N^2 = e^{-\frac{j2\pi}{N}} = W_{N/2}$ ,  $G(k)$  and  $H(k)$  are, respectively, the DFT of the following two  $N/2$  points:

$$G(k) = \sum_{n=0}^{N/(2-1)} x(2n)W_{N/2}^{kn} \tag{8}$$

$$H(k) = \sum_{n=0}^{\frac{N}{2}-1} x(2n + 1)W_{\frac{N}{2}}^{kn} \tag{9}$$

For each pair of  $X(k)$  and  $X(k + N/2)$ , as long as the even part is known, another even part is found. The odd part is the same.

### 2.3. Feature Extraction Process

VMD minimizes the sum of the estimated bandwidths of each mode, where each mode is assumed to be a finite bandwidth with a different center frequency. To solve this variational problem, an alternating direction multiplier method is used to continuously update each mode, and its center frequency is gradually demodulated to the corresponding baseband; finally, each mode, that is, the corresponding center frequency, is extracted together. We extract the max, min, mse, rsm, and mean values from each of the eight IMFs decomposed by VMD. The features extracted in VMD are F1~F40. The eight IMFs decomposed by VMD extract the maximum value, minimum value, and average value through FFT analysis. The features extracted in the FFT are F41~F80. Eighty features were extracted in feature extraction. Its calculation method is shown in Figure 1.

### 2.4. Symmetric Uncertainty (SU) Value Feature Selection

Information entropy is the average amount of information contained in each received message, proposed by Shannon in 1948 [36]. Messages represent events, samples, or features from a distribution or data stream. Another characteristic of the source is the probability distribution of the sample. The probability distribution of events and the amount of information for each event constitute a random variable, and the mean (i.e., expectation) of this random variable is the average (i.e., entropy) of the amount of information generated by this distribution. The calculation formula of entropy is shown in (10), and its information gain is shown in (11):

$$H(t) = -k \sum_i^N P(X_i)I(X_i) \tag{10}$$

$$I(X) = \sum_i^N \ln P(x_i) \tag{11}$$

where  $P$  is the probability mass function of  $x$ ,  $k$  is a proportional constant corresponding to the chosen metric, and  $i$  is the information body of  $x$ .

For the decision-making system, the greater the probability of occurrence of each piece of information, the smaller  $H(X)$  will be. This means that the greater the regularity that appears in the information, the smaller the degree of uncertainty, which means that the decision-making system has a high probability of presenting correct information; on the contrary, if the amount of information is too large, it means that the amount of information in the system is full of various information, so there will be a phenomenon of information overload. This is the most direct and fastest way to judge. By calculating the correlation coefficient, the correlation between them can be quickly obtained. However, if the correlation is used to select features, it often leads to the tendency to select features with larger values [4]. Therefore, this study uses the method of SU value to calculate the degree of correlation between the feature and the target, where SU is expressed as (12):

$$SU(X, Y) = \frac{I(X)}{H(X)+H(Y)} \tag{12}$$

The above expression is interpreted as the form of normalized information gain by definition, and the nonlinear related information variable defined by information entropy is used to reconstruct the degree of correlation between nonlinear random variables. Among them, the SU value is used to calculate the symmetric uncertainty. Its concept is similar to information acquisition, but the value range is between 0 and 1 (0 means that X has nothing to do with Y, and 1 means that knowing Y can accurately predict X).

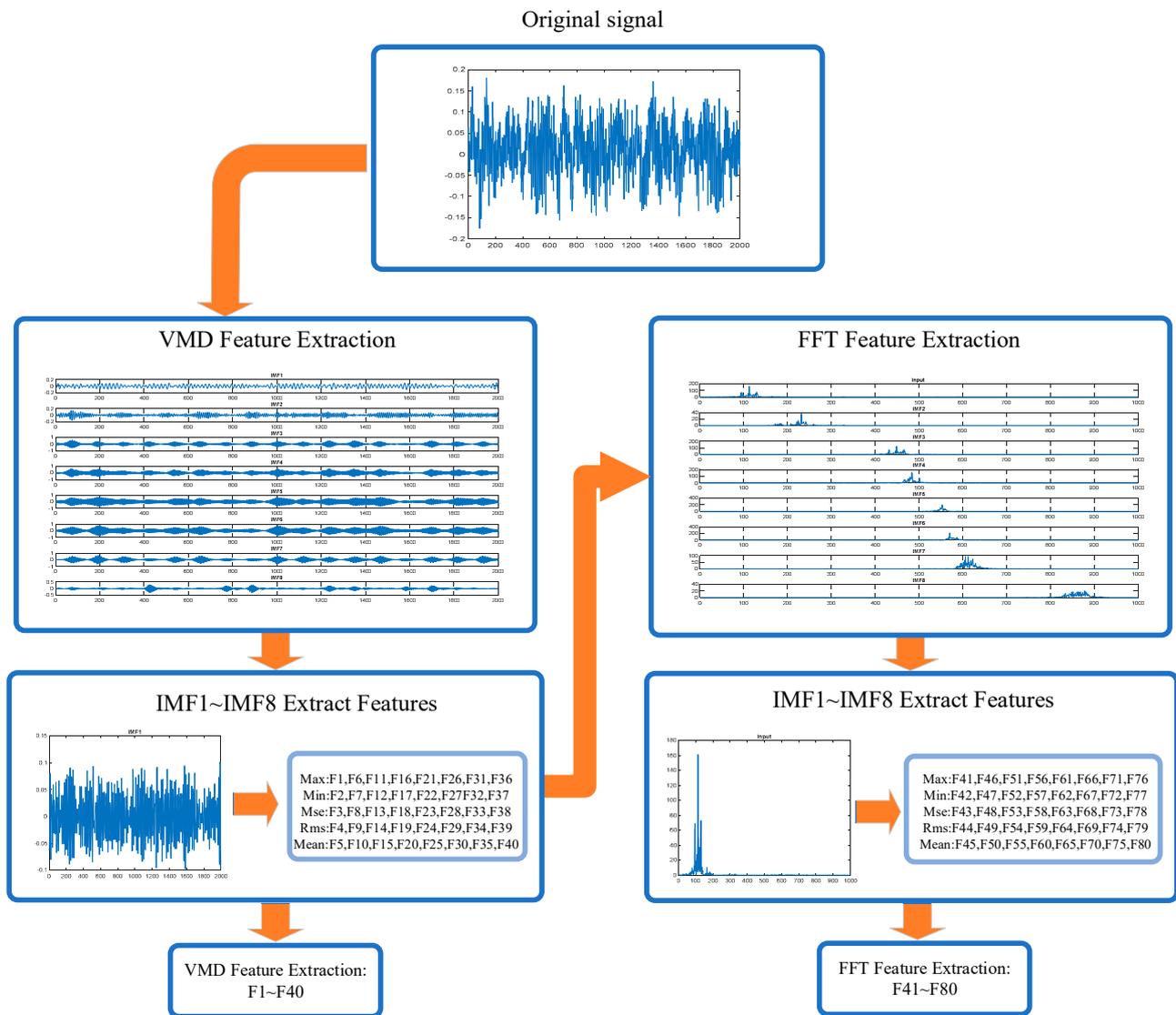


Figure 1. Feature extraction method flowchart.

2.5. Symmetric Uncertainty Method Feature Selection Process

Using the SU value of the fault type, the features (F1–F80) captured by VMD-FFT are sorted in descending order from the highest to the lowest correlation. During the screening process, the method compares the two features and retains the correlation with the target. Higher features use the features with higher correlation to complete the screening. The features with less influence are deleted by threshold setting, and the original 80 features are paired down to 60 features to obtain a feature subset and then feature comparison. This method will reduce the time complexity and realize the function of filtering while calculating to achieve the effect of speeding up the operation and improving the accuracy rate.

2.6. Feature Selection Method Process

In the filter feature selection method (FFS) [37], seven feature selection methods are used, as shown in Table 1 below. The FFS method is often used as a preprocessing step to rank the importance of variables in regression or classification problems. It selects features based on scores in various statistical tests and indicators of correlation. The algorithm of the FFS method has strong versatility, saves the training steps of the classifier, and has low algorithm complexity, so it is suitable for large-scale datasets and can quickly remove a large number of irrelevant features. It is very suitable as a prefilter for features. However, each type of method has a different calculation method to evaluate the weights of the features. This makes it difficult to make a final decision. Therefore, the main idea of this study is to apply TOPSIS to evaluate the priority order from the results of filter feature selection methods. SU is a preprocessor to remove redundant features before the remaining features are evaluated. In this study, a method with TOPSIS was proposed to improve the stability of the results of the FFS method (Table 2). Using the TOPSIS method to evaluate the current signals and select the best ranking of the signals for classification (Figure 2). The TOPSIS method is described in detail in the next section.

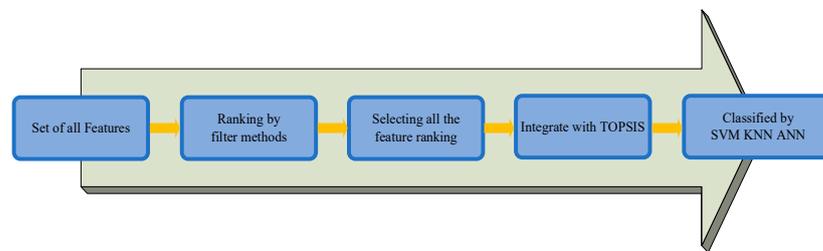


Figure 2. Schematic of the filter method proposed data analysis procedure.

Table 1. Filter methods for feature selection.

Methods	Advantages	Disadvantages	References
RelieFF (RF)	The algorithm is relatively simple High operating efficiency	The limitation of the algorithm is that it cannot effectively remove redundant features.	[24]
Minimum redundancy feature selection (mRMR)	Maximize the correlation between features and categorical variables Minimize the correlation between features and features	Does not take into account the correlation between features	[32]
Symmetrical uncertainty (SU)	Select a subset of features that are highly correlated with the category	A feature that has a high correlation with the target variable but little correlation with other features	[13]
Correlation-based feature selection (CFS)	Contains a subset of features that are highly correlated with the class but not correlated with each other	No interaction with classifiers, ignoring feature correlations	[33]
F-score (FS)	The accuracy rate can judge the total correct rate	In the case of unbalanced samples, it is not a good indicator to measure the results.	[34]
Pearson correlation coefficient (PCC)	The relationship of variables can be measured numerically and is directional	This method cannot be used to refine and solidify the relationship between variables to form a model.	[35]
Term variance (TV)	The method independently measures the relationship between each feature and the response variable.	The relationship between variables is not considered, so there will be redundant variables between numbers.	[33]

**Table 2.** Feature selection ranking collection.

Selected Type	Selected Methods
A	RF, mRMR, SU, CFS, FS, PCC, TV
B	RF, mRMR, CFS, FS, PCC, TV
C	RF, mRMR, CFS, FS, PCC
D	RF, CFS, FS, PCC, TV

2.7. The Feature Selection in TOPSIS

The TOPSIS method is a sequence optimization technique for ideal target similarity, and it is a very effective method in multiobjective decision analysis, proposed by Hwang and Yoon in 1981 [31]. Calculate the distance between each evaluation target and the ideal solution and anti-ideal solution, and use this as the basis for evaluating the target.

The workflow of the TOPSIS method consists of the following seven steps [38].

Step 1: Generating an *m-by-n* evaluation matrix contains *m* alternatives  $A_1, A_2; \dots, A_m$ , with each evaluation matrix assessed by *n* local criteria  $C_1, C_2; \dots; C_n$ .

Step 2: Normalizing the decision matrix:

$$u_{ij} = \frac{X_{ij}}{\sqrt{\sum_{k=1}^m x_{kj}^2}} ; i = 1, \dots, m; j = 1, \dots, n \tag{13}$$

where  $x_{ij}$  is the score of alternative  $A_i$  concerning criterion  $C_j$ .

Step 3: Calculating the weighted normalized decision matrix, for which its values  $V_{ij}$  are computed as (14):

$$V_{ij} = W_i \times u_{ij}; j = 1, 2, \dots, m; i = 1, 2, \dots, n \tag{14}$$

Let  $W_i = [w_1, w_2, \dots, w_n]$  be the vector of local criteria weights satisfying  $\sum_{i=1}^n W_i = 1$ .

Step 4: Determining the positive ideal  $I (A^+)$  and negative ideal  $(A^-)$  solutions as (15)–(18):

In the proposed method, all criteria are considered as benefits; therefore,  $J'$  is empty, and (15) and (16) can be reduced to (17) and (18):

$$A^+ = \{v_1^+, \dots, v_n^+\} = \{(\max_i V_{ij} | j \in J), (\min_i V_{ij} | j \in J')\} \tag{15}$$

$$A^- = \{v_1^-, \dots, v_n^-\} = \{(\min_i V_{ij} | j \in J), (\max_i V_{ij} | j \in J')\} \tag{16}$$

$$A^+ = \{v_1^+, \dots, v_n^+\} = \{(\max_i V_{ij} | j \in J)\} \tag{17}$$

$$A^- = \{v_1^-, \dots, v_n^-\} = \{(\min_i V_{ij} | j \in J)\} \tag{18}$$

Step 5: Measuring the Euclidean distances between each alternative and both the positive and negative ideal, which are calculated as (19) and (20):

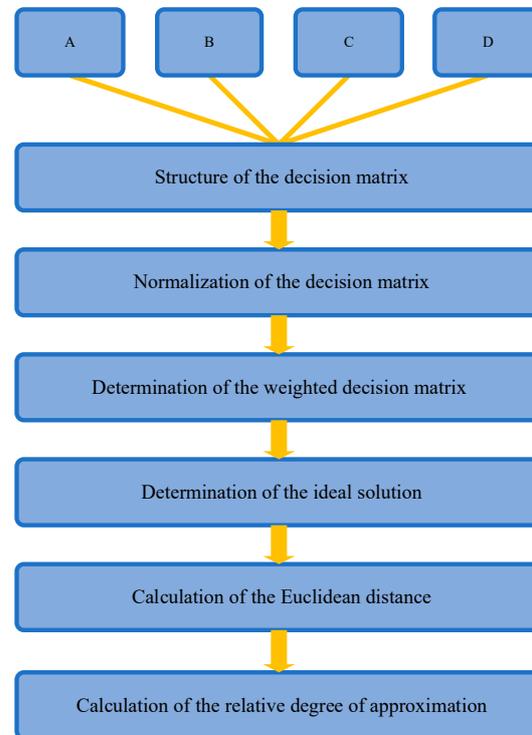
$$P_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2} ; i = 1, 2, \dots, m \tag{19}$$

$$P_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2} ; i = 1, 2, \dots, m \tag{20}$$

Step 6: Computing the relative closeness to the ideal solution as (21):

$$H_i = \frac{P_i^-}{P_i^+ + P_i^-} ; i = 1, 2, \dots, m; 0 \leq H_i \leq 1 \tag{21}$$

Step 7: Ranking alternatives based on the H value of each parameter.  $H_i = 1$  indicates the highest rank, and  $H_i = 0$  indicates the lowest rank. Its calculation method is shown in Figure 3.



**Figure 3.** The TOPSIS feature ranking process integrates different FFS methods ranking.

### 2.8. Performance Measures

The k-fold cross-validation test, independent dataset test, subsampling test, and jackknife cross-validation test are four schemes widely used in statistical classification to check the performance of predictive models [39]. The jackknife method is widely used to estimate the generalization ability of predictive models [40,41]. However, this is time-consuming. To save computation time, ten-fold cross-validation was used in this study.

We next investigated the performance of predictive models. In k-fold cross-validation, the data are divided into  $k$  subsets, each using one of the  $k$  subsets and  $k - 1$  subsets as test and training data, respectively.

### 2.9. Classification

In this study, we used three classification models, including the SVM, KNN, and ANN classifiers, to compare the bearing motor dataset and the CWRU dataset differences in the various fault types and normal motors and obtain accurate results through the analysis software MATLAB.

Support vector machine (SVM) [20,21]—An SVM is a supervised learning model and a machine learning model of related learning algorithms. It has relative advantages for problems such as small samples, nonlinearity, high dimensionality, and local minimum points. This is the method to use for classification or regression. Given a group of classified data, the SVM can obtain a set of models through training. Then, if there is unclassified data, the support vector machine can use the previously trained model to predict the category of this data, making it a nonprobabilistic binary linear classifier. Classification decisions are made through linear combinations of features. The characteristics of objects are usually described as eigenvalues, and in vectors, as eigenvectors. Since the support vector machine must first have classified data for training when building a model, the support vector machine is one of the methods of supervised learning.

Tuning the SVM classifier:

Using the parameter value that minimizes the cross-validation loss for the SVM and using the parameters that optimize the two-class learning, the eligible parameters are ‘BoxConstraint’, ‘KernelFunction’, ‘KernelScale’, ‘PolynomialOrder’, and ‘Standardize’.

Manually adjust the parameters of the classifier according to this scheme:

Step 1: Pass the data to the SVM and set the name–value pair parameters to ‘KernelScale’ and ‘auto’. Assume that the trained SVM model is called the SVM model. The heuristic process uses subsampling. Therefore, to reproduce the results, use rng to set the random number seed before training the classifier.

Step 2: Cross-validate a classifier by passing it to crossover. By default, the software performs 10-fold cross-validation.

Step 3: Pass the cross-validated SVM model to k-fold Loss to estimate and retain the classification error.

Step 4: Retrain the SVM classifier but adjust the ‘KernelScale’ and ‘BoxConstraint’ name–value pair arguments. As shown in Table 3.

**Table 3.** Support vector machine parameter setting.

Parameter	Parameter Value
BoxConstraint	1
KernelFunction	polynomial
KernelScale	auto
PolynomialOrder	2

k-Nearest Neighbor algorithm (KNN) [22]—The KNN is a nonparametric method for classification and regression prediction problems. It is one of the simplest of all machine learning algorithms. The sorting criteria are determined by a ‘majority vote’ of neighbors. The output of a regression model is a continuous value that is predicted to be the average of the outputs of the k-nearest neighbors. On the classification problem, the KNN adopts the majority principle and uses k-nearest neighbors to judge which group the new data belong to. The algorithm flow is very simple. The disadvantage of the KNN is that it is very sensitive to the local structure of the data, so it is extremely important to adjust the appropriate k value. First, determine the size of k. Then calculate the distance between the current new data and the adjacent data. In the third step, find out the k-nearest neighbors and see which group has the most neighbors, and then classify it into that group.

Artificial Neural Network (ANN) [26–30]—Artificial neural networks are used for supervised learning. More specifically, the ANN structure, training process, risk of overfitting, and data normalization for regression problems are analyzed. Calculations are routed through a large number of artificial neuron connections. In most cases, the artificial neural network can change its internal structure according to external information and is an adaptive system. Modern neural networks are nonlinear statistical data modeling tools, and neural networks are usually optimized by learning methods based on mathematical statistics. Artificial neural networks can have simple decision-making abilities and judgment abilities similar to human beings, and this method has more advantages than formal logical reasoning.

The different steps are described below:

Step 1: The collected input and output samples are divided into a test set and a training set. The splitting is performed randomly; usually, 80% of the samples are used for training and 20% for testing.

Step 2: The training set is subdivided into training and validation subsets. Splitting is random. Generally, 80% of the samples are used for the training subset and 20% for the validation subset.

Step 3: Set the weights and biases of the artificial neural network. In the first iteration, these values are chosen randomly. For the next iteration, a previous value of the error

metric is chosen relative to the error metric obtained in the previous iteration of the training subset.

Step 4: The error metric between the ANN output and dataset output is used to evaluate training and validation subsets. Widely used metrics for a regression ANN are mean squared error, and binary cross-entropy are used for a categorical ANN.

Step 5: Compare the error metrics for the training and validation subsets to stop training when overfitting occurs. The training error metric monitors a subset to detect training improvement when the completion metric converges and stops.

Step 6: If convergence is not achieved, the error metric is used to improve the weights and biases for the next iteration.

Step 7: After training is complete, evaluate the training and test sets and compare the resulting output with the dataset output. The training set comparison is biased because the same data are already used for training. Therefore, the test set exists and provides an unbiased validity check.

Step 8: If the performance of the artificial neural network cannot meet expectations, the ANN structure, number of hidden layers, number of neurons, or training algorithm should be reset.

ANN training is not a deterministic process due to the random splitting of datasets and random initialization bias of weights. The training subset is divided into batches for training iterations. The results of each iteration are used to improve the ANN parameters. The training process is complete when all batches are complete and all samples from the training subset have been used. The training process of an artificial neural network consists of many epochs. The number of samples per batch is a parameter that can affect the training process's quality, stability, and computational cost.

2.10. Proposed Method Process

In this study, the faulty bearing detection model is based on VMD-FFT to establish the feature extraction process and the used equations. This part of feature selection is divided into four cases of individual tests, as shown in Figure 4. The specific steps are as follows:

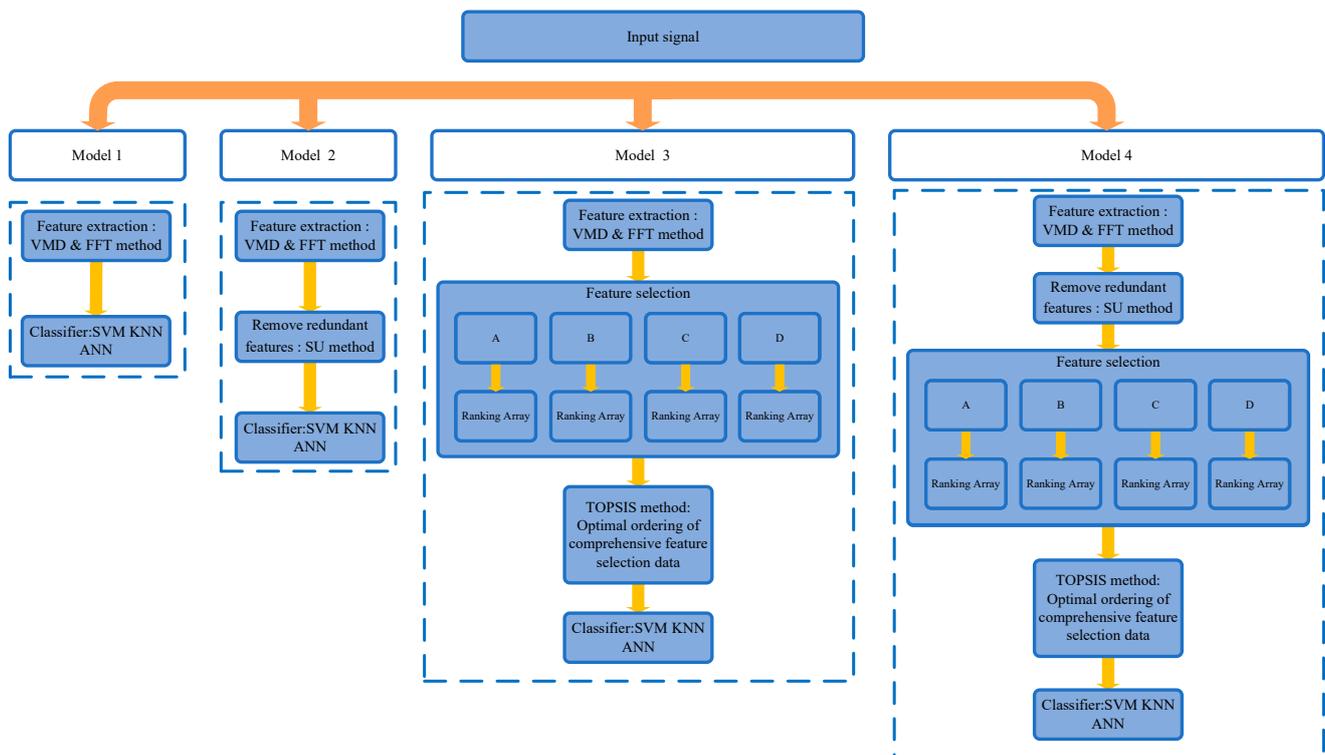


Figure 4. Propose an integrated FFS model flowchart for optimal feature selection.

**Model 1:** Using the VMD-FFT feature extraction process and presenting the result with a classifier.

**Model 2:** Using the VMD-FFT feature extraction process and SU method to remove redundant features and using the classifier to present the results.

**Model 3:** Using the VMD-FFT feature extraction process. In the feature selection part, the TOPSIS method (Figure 3) is used to obtain the ideal ranking, and the classifier is used to present the results.

**Model 4:** Using the VMD-FFT feature extraction process and using the SU method to remove redundant features. In the feature selection part, the TOPSIS method (Figure 3) is used to obtain the ideal ranking, and the classifier is used to present the results.

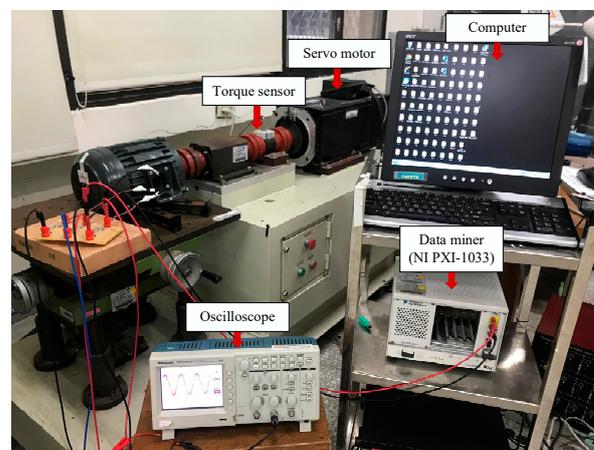
These four models are independent of each other. To test whether the SU method can effectively reduce redundant features and the combination of four different filtering methods selected in the TOPSIS method, the four training models of bearing faults constructed were compared. The next section introduces the bearing dataset and the CWRU dataset. We used these two datasets to test the results of the four models in the KNN, SVM, and ANN classifiers.

### 3. Hybrid Models

#### 3.1. Bearing Dataset of Current Signal Measured from an Induction Motors

This section describes the specifications of the motor used in the study and measures the motor for normal and damaged bearings, broken rotor bars, and shorted stator windings. Four current signals are used for analysis. Secondly, the equipment and methods used in the experiment and the overall process of this study are introduced, and the differences between various fault types and normal motors are preliminarily compared. Finally, the accuracy results are given by MATLAB analysis software.

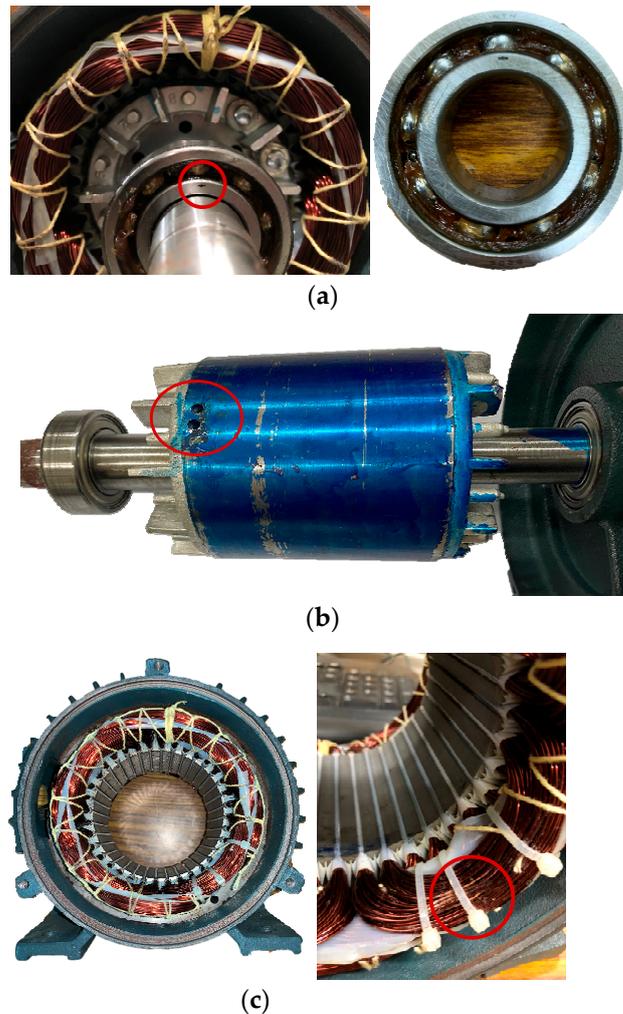
The equipment used in this study is a four-pole AC induction motor, as shown in Figure 5; its specifications are shown in Table 4, and the fault types are shown in Figure 6a–c. A signal acquisition device (NI PXI-1033), electricity meter, and computer were used for analysis, and the measurement data were then recorded. Raw current signals are obtained from experiments with common and three-faulted induction motors. Figure 7 shows the test bench hardware, a three-phase squirrel-cage induction motor of four types: normal, bearing failure, rotor drilled, and stator coil windings shorted. The three-phase squirrel-cage induction motor with bearing damage (aperture  $1.96 \text{ mm} \times 0.53 \text{ mm}$ ) is shown in Figure 6a. The three-phase squirrel-cage induction motor with the rotor hole (two holes  $\phi$  8 mm and 10 mm deep) is shown in Figure 6b. A three-phase squirrel-cage induction motor with short-circuited stator coil windings (2 coils short-circuited) is shown in Figure 6c.



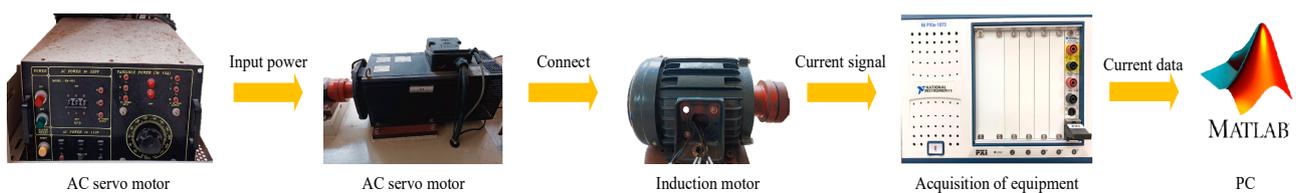
**Figure 5.** Torque sensor, servo motor, computer, oscilloscope, and data miner (NI PXI-1033).

**Table 4.** AC induction motor specifications.

Three-Phase Four-Pole Induction Motor Specifications		
Voltage 220 V/380 V	Frequency 60 Hz	Power Factor 0.8
Output 2 Hp 1.5 kW	Current 5.58 A/3.23 A	Rated speed 1764 rpm



**Figure 6.** (a) Bearing damage processing (aperture 1.96 mm × 0.53 mm). (b) Rotor drilling failure (two holes  $\phi$  8 mm depth 10 mm). (c) Short circuit between stator layers (two coils short circuit).



**Figure 7.** Signal measurement process.

### 3.2. Measurement Process in AC Induction Motor

First of all, this study measures the current signals of an AC induction motor in four conditions (normal, bearing, rotor, and stator) and obtains arbitrary phase data for the motor U, V, and W through the NI signal extractor. The data sampling time for each

measurement is 100 s, and the collection frequency is 1000 Hz. Each signal measurement is evaluated 100 times to complete the premeasurement operation.

### 3.3. CWRU Benchmark Dataset

The CWRU benchmark dataset provides validation of the ball bearing test data for normal and faulty bearings [42]. The test bench hardware consists of a 2 hp induction motor, load motor, and torque encoder. The CWRU benchmark dataset is unique in that each experiment carefully records the actual test conditions of the motor as well as the bearing fault states, including four different load levels (0 hp, 1 hp, 2 hp, 3 hp), three different fault locations (inner ring, outer ring, ball), three different defect diameters (0.007", 0.014", 0.021" inches), and a sampling rate of 12 kHz. The main purpose is to determine the severity and location of bearing failures. Normally, the signal is cut into 2000 data points at each level, so there are 660 samples in total except the normal signal, the signal at each level is sliced into 780 samples. The length samples for the three different defect diameters each have 2000 data points.

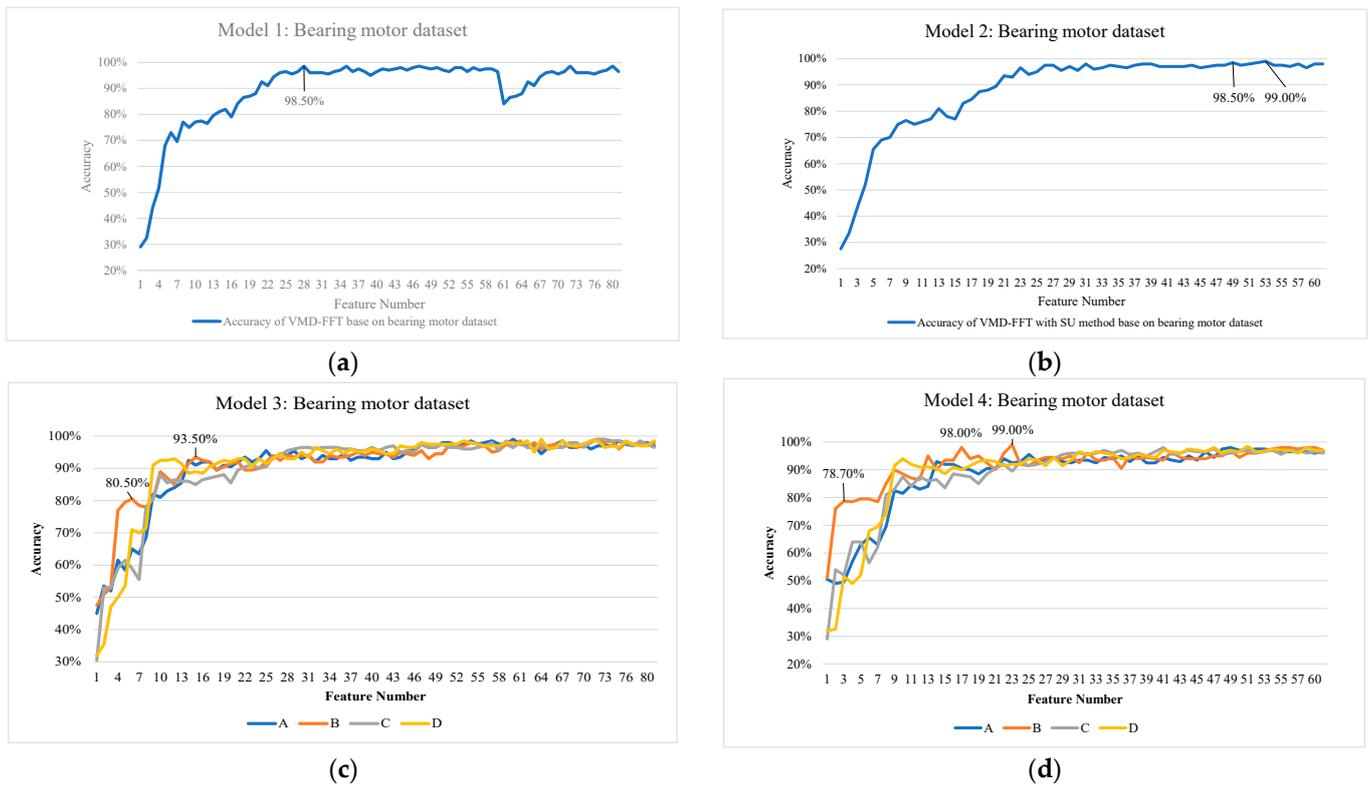
## 4. Measurement Method of the Motor Signal

In this study, the bearing dataset of the current signal measured from an induction motor and the CWRU bearing dataset are used as experimental samples for simulation. After VMF-FFT feature extraction, the comprehensive optimal SU feature removal method and TOPSIS method are used, and finally, the SVM, KNN, and ANN classifiers are used to compare the accuracy rate.

### 4.1. CASE STUDY 1: Bearing Motor Dataset

In order to demonstrate the performance of the proposed TOPSIS method, the bearing motor dataset was used in this case to test the results. Table 5 shows the effect of applying different feature selection techniques on various classifier architectures. On the classifier side, it was used as an evaluation measure during a 10-fold cross-validation process with 30 repetitions. For a fair performance evaluation, consider different constraints that affect classification performance, such as training dataset, classifier model, and several selected features. In this regard, we should evaluate different possible combinations of the four states applied to the three classifiers. In the case of using SVM machine learning, the average accuracy rate of the VMD-FFT signal analysis method is 87.98%, and the highest accuracy rate is 98.5%, as shown in Figure 8a. After using the VMD-FFT signal analysis method combined with the SU feature selection method, the 80 features can be reduced to 60, respectively, and the average accuracy rate is 88.09%. Compared with model 1, the average accuracy rate curve is relatively stable, as shown in Figure 8b. Using the VMD-FFT signal analysis method combined with the TOPSIS feature selection method, in Feature Number 4 and Feature Number 16, the accuracy rates are 80.50% and 93.50%. It can be seen that the six-select method is better than other selection methods, as shown in Figure 8c. It is clear that using the VMD-FFT signal analysis method combined with the TOPSIS feature selection method combined with the SU feature selection method can reduce the 80 features to 60, respectively. In Feature Number 3, Feature Number 18, and Feature Number 24, the accuracy rates are 78.70%, 98.0%, and 99.0%. Compared with the results of model 3, the curve of the six-select method in model 4 tends to stabilize faster, as shown in Figure 8d. Therefore, it can be judged that this method can delete redundant and unimportant features, obtain a better feature subset, and effectively improve accuracy.

In this case study, the proposed bearing fault diagnosis model is compared with state-of-the-art models. Since the results in Table 5 show that the six-select method achieves better results in model 3 and model 4, the six-select method is the main method in model 3 and model 4 in the three classifiers in Table 5. The accuracy of the bearing motor current signal dataset in model 4 is 78.72% for the KNN classifier, 91.82% for the SVM classifier, and 99.48% for the ANN classifier. The proposed model with the ANN classifier achieves the highest accuracy in the bearing motor current signal dataset



**Figure 8.** (a) Model 1: Bearing motor dataset of the VMD-FFT method in an SVM. (b) Model 2: Bearing motor dataset of the VMD-FFT method with redundant features removed by the SU method in an SVM. (c) Model 3: Bearing motor dataset of the TOPSIS method not removing features in an SVM. (d) Model 4: Bearing motor dataset of the TOPSIS method with redundant features removed by the SU method in an SVM.

**Table 5.** Result in bearing motor current signal dataset.

Bearing Motor Current Signal Dataset	KNN			SVM			ANN		
	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)
model 1	90.50	76.75	8.47	98.50	87.98	123.95	100	99.38	1.16
model 2	90.00	77.03	6.25	98.50	88.09	124.21	100	99.41	1.04
model 3 (B)	90.50	78.63	5.77	99.00	91.68	76.83	100	99.43	1.01
model 4 (B)	91.50	78.72	4.22	99.00	91.82	69.18	100	99.48	0.93

In the case of using SVM learning, the average recognition rate of the VMD-FFT signal analysis method is 87.98%, and the highest accuracy rate is 98.5%, as shown in Figure 8a. After using the VMD-FFT signal analysis method combined with the SU feature selection method, the 80 features can be reduced to 60, respectively, and the average accuracy rate is 88.09%. Compared with model 1, the average recognition rate curve is relatively stable, as shown in Figure 8b. Using the VMD-FFT signal analysis method combined with the TOPSIS feature selection method, in Feature Number 4 and Feature Number 16, the accuracy rates are 80.50% and 93.50%. It can be seen that the six-select method is better than other selection methods, as shown in Figure 8c. Using the VMD-FFT signal analysis method combined with the TOPSIS feature selection method combined with the SU feature selection method can reduce the 80 features to 60, respectively. In Feature Number 3, Feature Number 18, and Feature Number 24, the accuracy rates are 78.70%, 98.0%, and 99.0%. Compared with the results of model 3, the curve of the six-select method in model 4 tends to stabilize faster,

as shown in Figure 8d. Therefore, it can be judged that this method can delete redundant and unimportant features, obtain a better feature subset, and effectively improve accuracy.

In this case study, the proposed bearing fault diagnosis model is compared with state-of-the-art models. Since the results in Table 5 show that the six-select method achieves better results in model 3 and model 4, the six-select method is the main method in model 3 and model 4 in the three classifiers in Table 5. The accuracy of the bearing motor current signal dataset in model 4 is 78.72% for the KNN classifier, 91.82% for the SVM classifier, and 99.48% for the ANN classifier. The proposed model with the ANN classifier achieves the highest accuracy in the bearing motor current signal dataset. Therefore, the proposed bearing fault diagnosis model has better capability and can be applied to the practical task of fault diagnosis.

In this case study, the average running time of each method under different classifiers is shown in Table 5. The proposed method performs the best under each classifier. Model 4 had the shortest average running time of 4.22 s. for the KNN, 69.18 s. for the SVM, and 0.93 s. for the ANN. The SVM still has the longest average operation time and is significantly longer than the proposed method. In this case study, the proposed bearing fault diagnosis model is validated. Therefore, ANN classifiers are more suitable than those of the KNN and the SVM.

Based on the above results, in addition to showing the accuracy of each model, the performance of the three classifiers is shown, and type B of model 3 and model 4 are determined at the same time, which is an ideal solution for the feature selection method in TOPSIS.

4.2. CASE STUDY 2: CWRU Benchmark Dataset

In this case study, the proposed bearing fault diagnosis models are compared. Table 6 shows the classification of the four models of the CWRU 0 Hp benchmark dataset. The proposed model among the six-selected feature selection methods in model 4 achieves 97.52% accuracy in the KNN classifier. The accuracy of the SVM classifier is 98.60%, and the accuracy of the ANN classifier is 99.62%. In this case study, the proposed model with the ANN classifier achieves the highest accuracy, which is also higher than that of the proposed model with KNN and SVM classifiers.

Table 6. Result in CWRU bearing load 0 Hp dataset.

CWRU Bearing Load 0 Hp Dataset Data 2: 2000 × 660	KNN			SVM			ANN		
	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)
Model 1	99.09	96.73	9.89	99.69	97.34	287.2	100	99.54	1.33
Model 2	99.09	97.03	6.71	99.84	97.35	276.3	100	99.61	1.17
Model 3	99.39	97.19	9.05	99.84	98.45	177.6	100	99.62	1.26
Model 4	99.39	97.52	8.61	99.84	98.60	176.7	100	99.62	1.21

In this case study, Table 7 shows the classification of the four models of the CWRU 1 Hp benchmark dataset. The proposed model among the six-selected feature selection methods in model 4 achieves 98.41% accuracy in the KNN classifier. The accuracy of the SVM classifier is 98.45%, and the accuracy of the ANN classifier is 99.54%. In this case study, the proposed model with the ANN classifier achieves the highest accuracy, which is also higher than the proposed model with KNN and SVM classifiers.

In this case study, Table 8 shows the classification of the four models of the CWRU 2 Hp benchmark dataset. The proposed model among the six-selected feature selection methods in model 4 achieves 98.66% accuracy in the KNN classifier. The accuracy of the SVM classifier is 98.67%, and the accuracy of the ANN classifier is 99.55%. In this case study, the proposed model with the ANN classifier achieves the highest accuracy, which is also higher than the proposed model with KNN and SVM classifiers. In this case study,

Table 9 shows the classification of the four models of the CWRU 3 Hp benchmark dataset. Among the six feature selection methods selected in model 4, the proposed model achieved 99.73% accuracy in the ANN classifier. In this case study, the proposed model with the ANN classifier achieves the highest accuracy. To confirm the credibility of the results, this study compares the results of ANN classifiers in other papers in the CWRU data. In Zhiqiang Zhang, Funa Zhou, and Sijie Li’s paper [43], there are a total of six classifiers used for comparison, DNN, MDNN, MCNN, MRFNN, CMRFNN, G-CMRFNN, respectively. The method with the highest accuracy is G-CMRFNN, and the average accuracy of the classifiers is 98.20%. The average accuracy of the ANN classifier in this study is 99.62% for the CWRU 0 Hp model 4, 99.54% for the CWRU 1 Hp model 4, 99.55% for the CWRU 2 Hp model 4, and 99.73% for the CWRU 3 Hp model 4, as shown in Table 10. It can be seen from this that the model 4 method is superior in performance.

Table 7. Result in CWRU bearing load 1 Hp dataset.

CWRU Bearing Load 1 Hp Dataset Data 2: 2000 × 780	KNN			SVM			ANN		
	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)
Model 1	98.33	95.44	12.09	99.35	97.34	287.4	99.49	98.92	1.55
Model 2	98.46	95.93	7.91	99.35	97.47	277.5	99.87	99.33	1.26
Model 3	99.61	98.22	11.60	99.84	97.38	270.3	100	99.45	1.29
Model 4	99.74	98.41	7.73	99.84	98.45	260.4	100	99.54	1.24

Table 8. Result in CWRU bearing load 2 Hp dataset.

CWRU Bearing Load 2 Hp Dataset Data 3: 2000 × 780	KNN			SVM			ANN		
	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)
Model 1	99.23	96.31	11.58	99.74	97.21	324.3	100	99.22	1.53
Model 2	99.74	96.54	7.59	99.23	97.75	333.2	100	99.34	1.43
Model 3	99.35	98.57	11.78	99.87	98.13	176.3	100	99.50	1.35
Model 4	99.35	98.66	8.03	99.94	98.67	166.3	100	99.55	1.37

Table 9. Result in CWRU bearing load 3 Hp dataset.

CWRU Bearing load 3 Hp Dataset Data 4: 2000 × 780	KNN			SVM			ANN		
	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)	Best Accuracy (%)	Avg (%)	Time (s)
Model 1	99.74	97.74	11.76	99.74	97.95	286.3	100	99.66	1.58
Model 2	99.74	97.38	7.53	99.87	98.07	279.2	100	99.69	1.33
Model 3	99.48	98.31	11.31	99.87	98.89	106.8	100	99.70	1.31
Model 4	99.53	98.13	7.71	99.87	98.91	100.9	100	99.73	1.27

In the paper written by Laohu Yuan, Dongshan Lian, Xue Kang, Yuanqiang Chen, and Kejia Zhai [44], there are a total of six classifiers are used for comparison, namely PNN-SFAM, BPNN, CNN-HMM, DAFD, DGNN, and CNN-SVM. The best one is the CNN-SVM model with an average accuracy of 98.75%. In addition to the accuracy results of the ANN, this study also compares the accuracy results of the SVM. The average accuracy of CWRU 0 Hp model 4 of the SVM classifier is 98.60%, and the average accuracy of CWRU 1 Hp model 4 is 98.45%. The average accuracy of the CWRU 2 Hp model 4 is 98.67%, and

the average accuracy of the CWRU 3 Hp model 4 is 98.91%, as shown in Table 11. The performance of this study in SVM classification is not the best, but it is evenly matched.

**Table 10.** Comparison of neural network.

CWRU Dataset	Model	Average Accuracy (%)
Compare model	G-CMRFNN [43]	98.20
	CNN-SVM [44]	98.7593.70
	PDC-LR-HCNN [22]	
Proposed model	0 Hp model 4	99.62
	1 Hp model 4	99.54
	2 Hp model 4	99.55
	3 Hp model 4	99.73

**Table 11.** Comparison of neural support vector machine.

CWRU Dataset	Model	Average Accuracy (%)
Compare model	CNN-SVM [44]	98.75
	0 Hp model 4	98.60
Proposed model	1 Hp model 4	98.45
	2 Hp model 4	98.67
	3 Hp model 4	98.91

The average accuracy of CWRU 3 Hp model 4 is 0.16% higher than that of the proposed CNN-SVM method. In this study, the average accuracy of the ANN classifier is 99.62%, the CWRU 0 Hp model 4, CWRU 1 Hp model 4 is 99.54%, the CWRU 2 Hp model 4 is 99.55%, CWRU 3 Hp model 4 is 99.73%, as shown in Table 10. By comparing the results with other methods, it is not difficult to see that the method in this study has achieved a high diagnostic accuracy, which further proves the effectiveness of the method.

In the paper written by Shaohui Ning and Kangning Du [22], there are a total of four classifiers to compare in this paper, namely, traditional CNN, PDC-CNN, LR-CNN, PDC-LR-CNN, and the best of the four methods is the PDC-LR-HCNN method, with an average accuracy of 93.70%. The average accuracy of the ANN classifiers in this study are as follows: the CWRU 0 Hp model 4 is 99.62%, the CWRU 1 Hp model 4 is 99.54%, the CWRU 2 Hp model 4 is 99.55%, and the CWRU 3 Hp model 4 is 99.73%, as shown in Table 10. The method research proposed in this paper can not only diagnose bearing faults quickly but also maintain the accuracy of the diagnosis, which is obviously of great significance to the actual fault diagnosis.

### 5. Discussion

The following two points can be summarized based on the above data results:

Improve the combination of TOPSIS selection methods: Sort by detecting the distance between the evaluation object and the optimal solution and the worst solution; if the evaluation object is the closest to the optimal solution and at the same time farthest from the most cracked, it is the best; otherwise, it is the worst. According to the assumption, the ideal solution is the optimal solution, and its various attributes have reached the best value among the alternative solutions. If the distance between the optimal solution and the most cracked solution is too large, it leads to judgment errors in the model, such as the line segment selected type A and the line segment selected type B in Figure 8d, which show that it is not choosing more reference weights to increase the accuracy rate but choosing reference weights suitable for the data that can reduce the interval between the ideal solution and the optimal solution, thereby improving the accuracy rate.

Improve the application of TOPSIS in multiobjective decision-making problems: Reduce redundant features through VMD-FFT and SU methods, as shown in Table 5. Compared with model 3, model 4 reduces 25% of the features, but in the recognition rate and

running time, it is much better than model 3. Model 4 overcomes the shortcomings of poor objectivity and many assumptions and provides a more effective method for the selection of the optimal sequence

Apart from advantages, the proposed model still has flaws that need to be noted.

**Model multiplicity:** As mentioned in the feature selection method process, this study uses the filter method. This method does not consider which model to use in the future for learning. When selecting, it only evaluates the correlation between variables and predicted values and excludes the most irrelevant variables. Due to the relationship between variables not being considered, this study uses the TOPSIS method for aggregation. Compared to the wrapper and embedded methods, filter methods are simpler. Furthermore, when there are large datasets, testing tends to amplify into significant small differences in distributions that are not important.

**Combination of filter methods in TOPSIS:** This study proposes a combination of four different filter methods, as shown in Table 2. Due to the wide variety of filter methods, after excluding unsuitable types, six common filter methods were selected in this study, and only six combination methods have certain limitations on the weight of the optimal ranking, using other types of feature selection. The type of model needs further study.

## 6. Conclusions

Early detection of potential motor failures remains an important issue in operations and maintenance procedures. Therefore, this study proposes a motor bearing fault detection model. The symmetric uncertainty method and interpretability contained in the bearing vibration signal can effectively remove redundant features, remove irrelevant data, improve the accuracy of the learning model, reduce computational complexity, and improve the understandability of the model results. In order to overcome the shortcomings of the TOPSIS model in the multi-index decision-making process, the matrix between the original data sample and the ideal plan is used as a new decision matrix, and the ideal solution method is used to sort the plans. It overcomes the shortcomings of the traditional ideal solution, which is only based on the original data and is difficult to mine the inherent laws of the data, and provides a new idea for the decision-making problem under the condition of limited samples. At the same time, the method of combination weighting is proposed, which overcomes the shortcomings of traditional subjective and objective weighting methods. Results from measured bearing vibration data show that the proposed model 4 method outperforms traditional frequency-domain methods, feature selection methods, and other state-of-the-art filtering methods. Different FFS methods combined with different datasets in the TOPSIS method may result in different feature sets with different discriminants, so when given the wrong weights, the discrimination results may not be as expected. The results show that the proposed model 4 method helps researchers select more stable features from feature selection by integrating FFS methods. The proposed method has the advantages of stability and classification performance. However, it suffers from computational complexity issues compared to the FFS method. Still, the model 4 method has lower computational complexity compared to the filter selection method. Using the evaluation model established in this paper, through the application in the bearing fault evaluation, it shows that the result is reasonable, the calculation is simple, and it has a good application prospect.

The motor bearing fault diagnosis model proposed in this study has a high recognition rate in different datasets, but the feature selection method still needs to rely on a large amount of manual data processing in the adjustment of the TOPSIS method, which can be optimized in the future so that it can be more weighted and efficient. Future research should try to combine wrapping and embedded methods to apply the model more widely.

**Author Contributions:** Conceptualization, C.-Y.L., T.-A.L. and C.-Y.C.; Software, C.-Y.L., T.-A.L. and C.-Y.C.; Resources, C.-Y.L. and C.-Y.C.; Data curation, T.-A.L.; Writing—original draft, C.-Y.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

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

## References

1. Hoang, D.T.; Kang, H.; Kang, J. A motor current signal-based bearing fault diagnosis using deep learning and information fusion. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 3325–3333. [\[CrossRef\]](#)
2. Singh, S.; Kumar, N. Detection of bearing faults in mechanical systems using stator current monitoring. *IEEE Trans. Ind. Informat.* **2017**, *13*, 1341–1349. [\[CrossRef\]](#)
3. Chuan, L.; de Oliveira, J.V.; Cerrada, M.; Cabrera, D.; Sánchez, R.V.; Zurita, G. A systematic review of fuzzy formalisms for bearing fault diagnosis. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 1362–1382.
4. Dolenc, B.; Bošković, P.; Juričić, D. Distributed bearing fault diagnosis based on vibration analysis. *Mech. Syst. Signal Process.* **2016**, *66–67*, 521–532. [\[CrossRef\]](#)
5. Attoui, I.; Fergani, N.; Boutasseta, N.; Oudjani, B.; Deliou, A. A new time-frequency method for identification and classification of ball bearing faults. *J. Sound Vib.* **2017**, *397*, 241–265. [\[CrossRef\]](#)
6. Zhao, H.; Yang, X.; Chen, B.; Chen, H.; Deng, W. Bearing fault diagnosis using transfer learning and optimized deep belief network. *Meas. Sci. Technol.* **2022**, *33*, 065009. [\[CrossRef\]](#)
7. Wu, Y.; Jiang, B.; Wang, Y. Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains. *ISA Trans.* **2020**, *99*, 488–495. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Kaya, Y.; Kuncan, F.; Ertuğ, H.M. A new automatic bearing fault size diagnosis using time-frequency images of CWT and deep transfer learning method. *Turk. J. Electr. Eng. Comput. Sci.* **2020**, *30*, 1851–1867. [\[CrossRef\]](#)
9. Zhang, C.H.Y.; Yuan, L.; Xiang, S. Analog circuit incipient fault diagnosis method using DBN based features extraction. *IEEE Access* **2018**, *6*, 23053–23064. [\[CrossRef\]](#)
10. Chen, Y.Q.; Fink, O.; Sansavini, G. Combined fault location, and classification for power transmission lines fault diagnosis with integrated feature extraction. *IEEE Trans. Ind. Electron.* **2018**, *65*, 561–569. [\[CrossRef\]](#)
11. Zhu, J.; Ge, Z.; Song, Z. Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data. *IEEE Trans. Ind. Inf.* **2017**, *13*, 1877–1885. [\[CrossRef\]](#)
12. Zhao, H.; Zheng, J.; Xu, J.; Deng, W. Fault diagnosis method based on principal component analysis and broad learning system. *IEEE Access* **2019**, *7*, 99263–99272. [\[CrossRef\]](#)
13. Wang, Y.; Zhu, Y.; Wang, Q.; Tang, Y.; Duan, F.; Yang, Y. Complex fault source identification method for high-voltage trip-offs of wind farms based on SU-MRMR and PSO-SVM. *IEEE Access* **2020**, *8*, 130379–130391. [\[CrossRef\]](#)
14. Ye, X.; Hu, Y.; Shen, J.; Feng, R.; Zhai, G. An improved empirical mode decomposition based on adaptive weighted rational quartic spline for rolling bearing fault diagnosis. *IEEE Access* **2020**, *8*, 123813–123827. [\[CrossRef\]](#)
15. Rehman, N.U.; Aftab, H. Multivariate variational mode decomposition. *IEEE Trans. Signal Process.* **2019**, *67*, 6039–6052. [\[CrossRef\]](#)
16. Reddy, B.S.; Chatterji, B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **1996**, *5*, 1266–1271. [\[CrossRef\]](#)
17. Wang, Y.; Zheng, L.; Gao, Y.; Li, S. Vibration Signal Extraction Based on FFT and Least Square Method. *IEEE Access* **2020**, *8*, 224092–224107. [\[CrossRef\]](#)
18. Bayram, S.; Kaplan, K.; Kuncan, M.; Ertuğ, H.M. The effect of bearings faults to coefficients obtained by using wavelet transform. In Proceedings of the 2014 22nd Signal Processing and Communications Applications Conference (SIU), Trabzon, Turkey, 23–25 April 2014; pp. 991–994.
19. Chine, W.; Mellit, A.; Lughy, V.; Malek, A.; Sulligoi, G.; Pavan, A.M. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renew. Energy* **2016**, *90*, 501–512. [\[CrossRef\]](#)
20. Luo, A.; An, F.; Zhang, X.; Mattausch, H.J. A hardware-efficient recognition accelerator using Haar-like feature and SVM classifier. *IEEE Access* **2019**, *7*, 14472–14487. [\[CrossRef\]](#)
21. Cui, M.L.; Wang, Y.Q.; Lin, X.S.; Zhong, M.Y. Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine. *IEEE Sens. J.* **2021**, *21*, 4927–4937. [\[CrossRef\]](#)
22. Wang, Q.; Wang, S.; Wei, B.; Chen, W.; Zhang, Y. Weighted K-NN classification method of bearings fault diagnosis with multi-dimensional sensitive features. *IEEE Access* **2021**, *9*, 45428–45440. [\[CrossRef\]](#)
23. Hong Lan, L.T.; Tuan, T.M.; Ngan, T.T.; Son, L.H.; Giang, N.L.; Nhu Ngoc, V.T.; Hai, P.V. A New Complex Fuzzy Inference System with Fuzzy Knowledge Graph and Extensions in Decision Making. *IEEE Access* **2020**, *8*, 164899–164921. [\[CrossRef\]](#)
24. Roy, S.S.; Dey, S.; Chatterjee, S. Autocorrelation aided random forest classifier based on bearing fault detection framework. *IEEE Sens. J.* **2020**, *20*, 10792–10800. [\[CrossRef\]](#)
25. Russell, S.J.; Norvig, P. Contributing writers. In *Artificial Intelligence: A Modern Approach*, 2nd ed.; John, F.C., Jitendra, M.M., Douglas, D.E., Eds.; Prentice Hall: Englewood Cliffs, NJ, USA, 2003.
26. Zaccane, G.; Karim, R.; Menshawy, A. *Deep Learning with TensorFlow: Explore Neural Networks with Python*; Packt: Birmingham, UK, 2017.
27. Ketkar, N. *Deep Learning with Python: A Hands-On Introduction*; Apress: New York, NY, USA, 2017; pp. 195–208.

28. Kuncan, M. An intelligent approach for bearing fault diagnosis: Combination of 1D-LBP and GRA. *IEEE Access* **2020**, *8*, 137517–137529. [[CrossRef](#)]
29. Kaplan, K.; Bayram, S.; Kuncan, M.; Ertunç, H.M. Feature extraction of ball bearings in time-space and estimation of fault size with method of ANN. In Proceedings of the 16th International Conference on Mechatronics, Brno, Czech Republic, 3–5 December 2014; pp. 295–300.
30. Yang, H.; Li, X.; Zhang, W. Interpretability of deep convolutional neural networks on rolling bearing fault diagnosis. *Meas. Sci. Technol.* **2022**, *33*, 055005. [[CrossRef](#)]
31. Lai, Y.J.; Liu, T.Y.; Hwang, C.L. TOPSIS for MODM. *Eur. J. Oper. Res.* **1994**, *76*, 486–500. [[CrossRef](#)]
32. Sun, Y.; Ma, L.; Qin, N.; Zhang, M.; Lv, Q. Analog filter circuits feature selection using MRMR and SVM. In Proceedings of the 14th International Conference on Control, Automation and Systems (ICCAS 2014), Gyeonggi-do, Republic of Korea, 22–25 October 2014; pp. 1543–1547.
33. Li, X.; Zheng, Z.; Wu, L.; Li, R.; Huang, J.; Hu, X.; Guo, P. A stratified method for large-scale power system transient stability assessment based on maximum relevance minimum redundancy arithmetic. *IEEE Access* **2019**, *7*, 61414–61432. [[CrossRef](#)]
34. Feng, N.; Zhang, Y.; Zeng, Q.; Tong, M.; Joines, W.T.; Wang, G.P. Direct-splitting-based CN-FDTD for modeling 2D material nanostructure problems. *IEEE Open J. Antennas Propag.* **2020**, *1*, 309–319. [[CrossRef](#)]
35. Mohr, J.H.M. *Utility of Piotroski F-Score for Predicting Growth-Stock Returns*; MFIE Capital Working Paper: Kalken, Belgium, 2012.
36. Yu, L.; Liu, H. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 856–863.
37. Cekik, R.; Uysal, A.K. A novel filter feature selection method using the rough set for short text data. *Expert Syst. Appl.* **2020**, *160*, 113691. [[CrossRef](#)]
38. Saghapour, E.; Kermani, S.; Sehhati, M. A novel feature ranking method for prediction of cancer stages using proteomics data. *PLoS ONE* **2017**, *12*, e0184203. [[CrossRef](#)]
39. Ding, H.; Luo, L.; Lin, H. Prediction of Cell Wall Lytic Enzymes Using Chou’s Amphiphilic Pseudo Amino Acid Composition. *Protein Pept. Lett.* **2009**, *16*, 351–355. [[CrossRef](#)]
40. Nanni, L.; Brahnam, S.; Lumini, A. Prediction of protein structure classes by incorporating different protein descriptors into general Chou’s pseudo amino acid composition. *J. Theor. Biol.* **2014**, *360*, 109–116. [[CrossRef](#)] [[PubMed](#)]
41. Chou, K.C. Some Remarks on Protein Attribute Prediction and Pseudo Amino Acid Composition (50th Anniversary Year Review). *J. Theor. Biol.* **2011**, *273*, 236–247. [[CrossRef](#)] [[PubMed](#)]
42. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–131. [[CrossRef](#)]
43. Zhang, Z.; Zhou, F.; Li, S. A Cross Working Condition Multiscale Recursive Feature Fusion Method for Fault Diagnosis of Rolling Bearing in Multiple Working Conditions. *IEEE Access* **2022**, *10*, 78502–78518. [[CrossRef](#)]
44. Yuan, L.; Lian, D.; Kang, X.; Chen, Y.; Zhai, K. Rolling bearing fault diagnosis based on convolutional neural network and support vector machine. *IEEE Access* **2020**, *8*, 137395–137406. [[CrossRef](#)]

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