

## Article

# Multi-Filter Clustering Fusion for Feature Selection in Rotating Machinery Fault Classification

Solichin Mochammad <sup>1,2</sup>, Yoojeong Noh <sup>1,\*</sup>, Young-Jin Kang <sup>3</sup>, Sunhwa Park <sup>4</sup>, Jangwoo Lee <sup>4</sup> and Simon Chin <sup>4</sup>

<sup>1</sup> School of Mechanical Engineering, Pusan National University, Busan 46241, Korea; msolichin1989@gmail.com

<sup>2</sup> Department of Mechanical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

<sup>3</sup> Research Institute of Mechanical Technology, Pusan National University, Busan 46241, Korea; zmanx@pusan.ac.kr

<sup>4</sup> H&A Research Center, LG Electronics, Changwon 51554, Korea; sunhwa1124.park@lge.com (S.P.); jonathan.lee@lge.com (J.L.); simon.chin@lge.com (S.C.)

\* Correspondence: yoonoh@pusan.ac.kr

**Abstract:** In the fault classification process, filter methods that sequentially remove unnecessary features have long been studied. However, the existing filter methods do not have guidelines on which, and how many, features are needed. This study developed a multi-filter clustering fusion (MFCF) technique, to effectively and efficiently select features. In the MFCF process, a multi-filter method combining existing filter methods is first applied for feature clustering; then, key features are automatically selected. The union of key features is utilized to find all potentially important features, and an exhaustive search is used to obtain the best combination of selected features to maximize the accuracy of the classification model. In the rotating machinery examples, fault classification models using MFCF were generated to classify normal and abnormal conditions of rotational machinery. The obtained results demonstrated that classification models using MFCF provide good accuracy, efficiency, and robustness in the fault classification of rotational machinery.

**Keywords:** feature selection; rotating machinery; fault classification; fusion; multi-filter; clustering



**Citation:** Mochammad, S.; Noh, Y.; Kang, Y.-J.; Park, S.; Lee, J.; Chin, S. Multi-Filter Clustering Fusion for Feature Selection in Rotating Machinery Fault Classification. *Sensors* **2022**, *22*, 2192. <https://doi.org/10.3390/s22062192>

Academic Editors: Giuseppe Lacidogna, Sanichiro Yoshida, Guang-Liang Feng, Jie Xu, Alessandro Grazzini and Gianfranco Piana

Received: 16 February 2022

Accepted: 9 March 2022

Published: 11 March 2022

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



**Copyright:** © 2022 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

Rotating machinery plays a crucial role in the systems and processes of industrial applications, such as manufacturing systems, transportation, home appliances, and power systems [1,2]. Since rotating machinery generally operates continuously at high speeds and with high power [3], interruption of the related processes could threaten safety and result in massive economic loss [4,5]. Therefore, fault diagnosis of rotating machinery is essential to prevent critical failures that would cause a system to shut down.

The fault diagnosis of rotating machine is performed by detecting outliers that may occur due to faults in the monitored data. Traditional fault detection methods have mainly used thresholds set based on domain knowledge. However, recently, many fault detection methods have detected faults by learning monitored data, using machine learning/deep learning technology. Classifying normal or abnormal conditions is performed with binary classification models, and multi-class classification models are used to detect different combinations of faults. These fault classification models help people make decisions and predict the occurrence of more severe failures in parts or machines in advance.

The most detectable signs of failures in rotating machinery are vibrations and noise from abnormal conditions. However, since noise generated under abnormal conditions is often difficult to distinguish from noise generated in external environments, vibration data are more frequently used to diagnose failures in rotating machinery. In particular, accelerometers are frequently used to measure vibration data, and various fault diagnostic

methods have been developed using vibration information [6,7]. Vibration data of rotating machinery measured over time are further amplified by periodic operation of the machinery at certain frequencies. Thus, the characteristics (features) of signals, in both time and frequency domains, can be important measurements to distinguish between normal and abnormal conditions of rotating machinery. Such features are used to create fault classification models that distinguish between normal and abnormal states or different failure modes.

Selection of appropriate features (key to successfully building classification models) has been studied in recent years [8,9]. When extracting features for use as input features for classification models, it is important to use features that are highly relevant to classification, and to eliminate redundant or unnecessary features [10]. This is because learning from the data to generate a model takes a long time, and the complexity of the model increases; but the accuracy can decrease as the number of unnecessary features increases. Therefore, research on effective and accurate feature selection needs to be carried out to improve the efficiency and accuracy of fault classification models [11,12].

Various feature selection methods have been proposed for fault diagnosis in rotating machinery. In particular, since the number and type of features derived from both time and frequency domains can vary, feature selection is very important for obtaining accurate classification models. Therefore, many studies have recently been conducted to optimize the combination of features with the highest classification accuracy in feature selection. An optimal combination of features has been derived using filter methods such as relief, chi-square, and information gain [13], or by using Pareto optimization [14] or a binary particle swarm optimization method [15] after using a filter method. A wrapper-based embedded method, using a support vector machine (SVM) [16], and a method of deriving optimal features using the sensitivity of features [17] or a genetic algorithm [18] have been proposed. Although these methods have shown results in improving classification accuracy, each method has been applied to specific classification problems, which is insufficient to show the universality and robustness of the proposed methods.

Some studies have proposed effective fault diagnosis methods with several components, by extracting several features based on deep learning models using multivariate sampling data [19]. However, because these methods use a backpropagation training process, they are time-consuming and can have unsatisfactory performance when dealing with high-dimensional data. In another study, a universal domain adaptation method was proposed, to enhance the generalization ability of a data-driven model for fault diagnosis [20]. The fault diagnosis results of roller bearings showed that the proposed method yielded the best performance compared with other neural network methods. However, the study assumed that the balanced data were available in the training process. Thus, it might not be applicable to unbalanced data, which often occur in real industry applications.

This study proposes a fault classification model for rotating machinery, which is combined with a notable feature selection method, as follows: (1) Multi-filter clustering fusion (MFCF) was developed to provide an adaptive threshold capable of determining the total number of relevant features through hierarchical clustering. (2) An exhaustive search of the wrapper method was used to find the best feature sets maximizing classification accuracy. (2) The performance of the proposed method was validated in four rotating machinery cases with different operating processes, fault modes, and numbers of datasets. (3) The selected features were used to train and test several classifiers, including the SVM, k-nearest neighbors (KNN), and multilayer perceptron (MLP), to ensure that the final selected features are compatible with all classifiers. Finally, the proposed method was shown to have high accuracy, robustness, efficiency, and generalizability in fault classification for rotating machinery through multi-domain feature extraction and multi-filter fusion.

## 2. Related Methods

### 2.1. Feature Selection Methods

Feature selection can be divided into wrapper, hybrid, and embedded methods. Filter methods include methods determining the ranks of features by evaluating close relationships or similarity of features, based on information theory and statistics [17]. They evaluate the relative importance of features, but there is no absolute criterion for selecting them [21], so it is difficult to distinguish between necessary and unnecessary features [22]. Users need to arbitrarily determine the number of features, or select features according to a user-specified percentage [10], making it difficult to clearly conclude that certain filter methods are superior to others [23]. Therefore, while filter methods can efficiently remove unnecessary features based on importance, there is no guideline for selecting important features.

The most commonly used filter methods include chi-square (CS), the extra trees classifier (ETC), and a correlation matrix (CM). The CS method provides a ranking of features based on an independence test of two events using  $\chi^2$  values. The ETC uses entropy values to measure the probability of the same class by aggregating the learning ensemble, and a CM measures the similarity between two features, with a final coefficient of the degree of linear correlation, as shown in Equations (1)–(3):

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

where  $O_i$  is the observed feature data, and  $E_i$  is the expected feature data,

$$\text{Entropy} (E) = - \sum_{i=1}^c p_i \log_2(p_i) \quad (2)$$

where  $c$  is the number of group labels, and  $p_i$  is the proportion of feature values associated with group  $i$ , and

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \sqrt{\sum(Y_i - \bar{Y})^2}} \quad (3)$$

where  $X_i$  and  $Y_i$  are feature observation data, and  $\bar{X}$  and  $\bar{Y}$  are the mean values of the two features,  $X$  and  $Y$ .

However, each method has different measures to evaluate feature importance, so they can yield different rankings of features. Therefore, for effective and robust feature selection, the method of extracting and combining key features from each feature selection method with different characteristics becomes an important issue, in finally deriving the best feature set.

The wrapper method determines the types and number of features based on the accuracy of classification models. All possible feature combinations are used as input features in the classification model, so the feature combination with the highest classification accuracy is chosen as the final feature set [4]. The wrapper method, unlike the filter method, provides an optimal combination of features, but requires a long computational time, because it generates a classification model for every combination of features [24,25]. Among the wrapper methods, an exhaustive search enables accurate and robust feature selection by simultaneously evaluating all combinations of features, instead of gradually adding or removing features. Exhaustive search is inefficient, compared to the other methods if all existing features are used without removing unimportant features. However, if the filter method is applied first, and the number of key features derived from the filter method is small, then an exhaustive search can find optimum features more effectively and efficiently. The hybrid method is a combination of filter and wrapper methods, to improve the shortcomings in each one. For example, after unnecessary features are removed using the filter method, the wrapper method is applied to only find the best feature set from the reduced features, resulting in a significant reduction in computational time [26,27]. While

hybrid methods reduce the computational time needed with wrapper methods, they still need to select the appropriate number of features in the filter process.

## 2.2. Classifiers

Various classifiers can be applied to build classification models for the diagnostic needs of rotating machinery [28–30]. MLP with a neural network structure, the SVM with a decision boundary, and the distance-based KNN model are widely used classifiers. Since classifiers may have very different performances, depending on their characteristics, this study attempts to verify performance through a combination of the proposed feature selection methods, with the above three representative classifiers.

The SVM solves linear and nonlinear classification problems by finding hyperplanes that maximize the distance between groups, by learning from training data and determining the kernel type, such as linear, polynomial, or radial basal plane [31,32]. The SVM classifier is formulated as follows:

$$f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \quad (4)$$

where  $\mathbf{w}$  is the vector of the weight,  $b$  is the bias for optimizing the hyperplane, and  $\varphi(\mathbf{x})$  is the mapping function of the kernel. The vector of weight  $\mathbf{w}$  can be known by minimizing it:

$$\text{Minimize } \mathbf{w}^T \mathbf{w} + C \sum_i \xi_i \quad (5)$$

where  $C$  is the penalty hyperparameter, and  $\xi_i$  is a slack variable for  $i = 1, 2, \dots, N$ , with  $N$  as the number of data samples.

KNN is a type of supervised learning that can be used as a task in classification and regression. It performs classification by measuring similarity (e.g., distance functions) between data points [32]. Euclidean distance is often used as the distance metric, as follows:

$$\text{Dist}(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

where  $x_i$  and  $y_i$  are the coordinate values of the sample for  $X$  and  $Y$  as two data points, and  $n$  is the dimension of the data points. KNN attempts to find the distance between the query and all sample data. After that, it specifies the number of samples ( $k$ ) closest to the query, and then, the most frequent label is selected.

MLP is an algorithm in machine learning that works with feed-forward neural networks. It has a structure consisting of an input layer, multiple hidden layers, and an output layer. MLP is famous for being able to solve complex problems, because of its outstanding performance in building classifications [33,34]. In simple terms, the MLP output function is expressed as

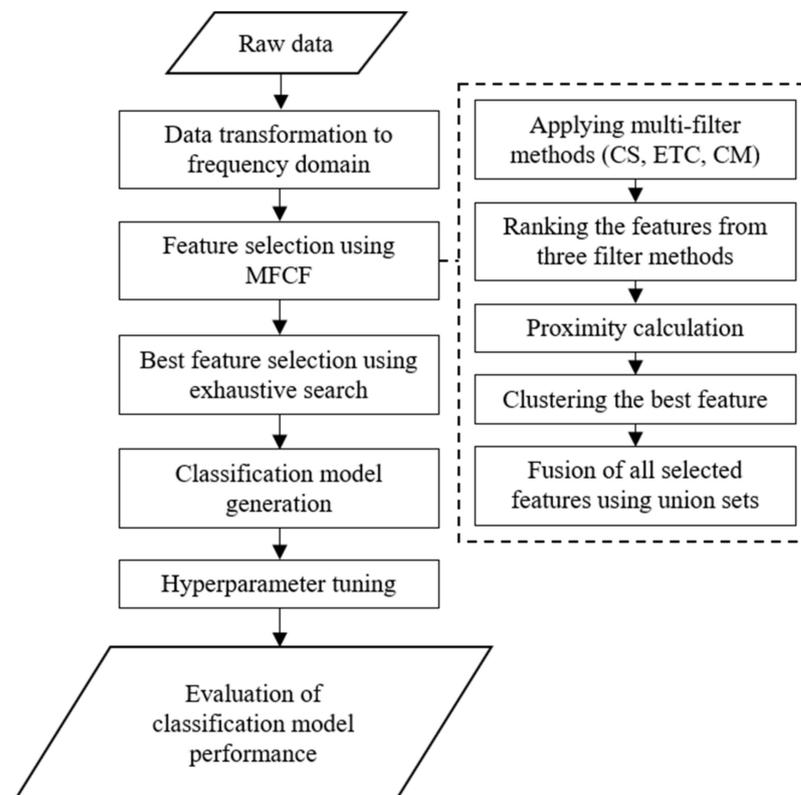
$$y = g(\mathbf{W}^T \mathbf{x} + \mathbf{b}) \quad (7)$$

where  $\mathbf{x}$  is the input variable in vector form,  $y$  is the output;  $g(\cdot)$  is the activation function of the nodes,  $\mathbf{W}$  is the weight matrix linked to the input layer and hidden layer, and  $\mathbf{b}$  is the bias vector of hidden layer nodes. Each component of the input layer, multiple hidden layers, and output layers can be assigned according to the level of complexity in the problem.

## 3. Proposed Method

Each filter method described in Section 2.1 can select different features depending on the type of features and the characteristics of the data, so it is important to systematically and effectively select the most important features that affect classification performance. The proposed MFCF feature selection focuses on how to cut off unnecessary features adaptively from the candidate feature sets and find the best feature combination in an efficient and

systematic way. For this, the raw data are first used to extract features from time and frequency domains through fast Fourier transform (FFT), as shown in Figure 1.



**Figure 1.** Flow chart of the proposed method.

MFCF is used to extract candidate feature sets using multiple filter methods and feature clustering, and an exhaustive search is used to select the optimal feature set that maximizes classification accuracy. The selected features are used to generate fault classification models (such as SVM, KNN, and MLP) where hyperparameters of the three models are optimized using a grid search. The performance of the proposed method is evaluated in terms of accuracy, efficiency, stability, and robustness. Accuracy and efficiency are evaluated using measures such as the percentage of the correct predictions and computational time, respectively. Stability is estimated from changes in both accuracy and efficiency when the method is applied to training and testing the data. Robustness is measured through variation of accuracy values, through cross validation.

### 3.1. Fusion Multi-Filter Feature Selection

Before MFCF is applied, all features from the time and frequency domains first need to be defined. In order to determine the statistical characteristics of the measured data in the time and frequency domains, 12 features were extracted from each domain, including absolute mean (abs\_mean), peak-to-peak (ptp), kurtosis (kur), skewness (skew), root mean square (rms), etc.; 25%, 50%, and 75% are the 25th, 50th, and 75th percentile values, respectively. The 24 features were evaluated with CS, ETC, and CM methods. The numbering for the 72 features is in Table 1.

**Table 1.** Feature number.

Frequency Domain Features	$F_{f\_CS}$	$F_{f\_ETC}$	$F_{f\_CM}$	Time Domain Features	$F_{t\_CS}$	$F_{t\_ETC}$	$F_{t\_CM}$
abs_mean_F	0	24	48	abs_mean_T	12	36	60
peak_m_F	1	25	49	peak_m_T	13	37	61
kur_F	2	26	50	kur_T	14	38	62
skew_F	3	27	51	skew_T	15	39	63
rms_F	4	28	52	rms_T	16	40	64
mean_F	5	29	53	mean_T	17	41	65
std_F	6	30	54	std_T	18	42	66
min_F	7	31	55	min_T	19	43	67
25%_F	8	32	56	25%_T	20	44	68
50%_F	9	33	57	50%_T	21	45	69
75%_F	10	34	58	75%_T	22	46	70
max_F	11	35	59	max_T	23	47	71

Referring to Table 1, feature numbering can be expressed as  $F_{Comb} = [F_0 F_1, \dots, F_N]$  for  $k = 0, 1, \dots, 71$ , where  $k$  is the list of feature numbers, and  $N$  is the total number. Then the term is redefined as follows:  $F_{Comb} = [F_{CS} F_{ETC} F_{CM}]$  where  $F_{CS} = \{F_{f_{CS}} F_{t_{CS}}\}$ ,  $F_{ETC} = \{F_{f_{ETC}} F_{t_{ETC}}\}$ , and  $F_{CM} = \{F_{f_{CM}} F_{t_{CM}}\}$ .

Clustering of the 72 features should be performed to classify them into important features and unimportant features, to be used for classification based on the feature importance measures from each filter method. For this, all feature values are normalized, and the distances between two feature values are calculated using Euclidean distance for all features, as shown in Equation (8):

$$d_{ij} = d(F_i, F_j) = \sqrt{\sum_{k=0}^N (F_{i,k} - F_{j,k})^2} \quad (8)$$

where  $d_{ij}$  is the distance between feature  $i$  and feature  $j$ , and  $N$  denotes the amount of data, including all feature values. Using hierarchical clustering, the distances between features are repeatedly calculated, and features with small or large distances are combined into one of two clusters: selected features or removed features. Using the Euclidean distance between two features in Equation (8), a pairwise distance matrix to find cluster A with selected features and cluster B with removed features can be defined as follows:

$$d_{AB} = \begin{bmatrix} 0 & d_{01} & \dots & d_{0N} \\ d_{10} & 0 & \dots & \dots \\ \dots & \dots & 0 & \dots \\ d_{N0} & \dots & \dots & 0 \end{bmatrix} \quad (9)$$

where  $d_{AB}$  is a proximity matrix for measuring the distances between features. Features with short distances are clustered based on  $\text{Min} \{d(F_i, F_j)\}$ , and then, the proximity matrix is expressed as  $d_{AB} = \{d_A\} \{d_B\}$ . In other words, features with high proximity values are clustered into one group, whereas features with low proximity values are clustered into the other. This agglomerative clustering is repeated by building a new matrix, until the last matrix consists of only two large clusters, separating one important feature group and one unimportant feature group.

$d_{AB} = \{d_A\} \{d_B\}$  represents the final proximity result capable of building cluster A containing the feature set ( $d_A$ ), with high-ranking values from evaluating important features obtained from each single-filter method. On the other hand, cluster B, containing the unimportant feature set ( $d_B$ ), which is far from cluster A, is discarded.

This feature selection is unsupervised learning, in which the algorithm automatically searches for important features by using Ward's method, through error sum of squares (ESS) and calculating the loss associated with each cluster. The ESS is computed, to measure the distance between two clusters of important and unimportant features of multi-filter

scoring, which is called the linkage function. Ward's linkage function is known to be the most suitable method to quantify a good group based on the variance of the clusters. The target of the linkage search is to minimize the increment of the ESS at each step, to find the minimum information loss. This algorithm works by fusing two clusters as the mean vector, and it then calculates the ESS from each cluster, namely the selected feature cluster and the discarded feature cluster. The following equations define the ESS in Equation (10) and the linkage between clusters  $A$  and  $B$ ,  $D(A,B)$ , in Equation (11):

$$ESS(d_A) = \sum_{i=1}^{T_a} \left| Va_i - \frac{1}{T_a} \sum_{j=1}^{T_a} Va_j \right|^2 \quad (10)$$

$$D(A,B) = ESS(d_{AB}) - [ESS(d_A) + ESS(d_B)] \quad (11)$$

where  $Va$  is the value of each feature, and  $T_a$  is the number of data points in cluster  $A$ . With the same formula as Equation (10),  $ESS(d_{AB})$  and  $ESS(d_B)$  are calculated by changing the names of the variables, such as  $Va$  to  $Vb$  as a feature value in cluster  $B$ , and  $Vab$  as a feature value in the combined cluster resulting from cluster fusion. Then, cluster  $A$  consists of several features that may be the same, due to feature selection by clustering based on all members. To optimize the combination of all potential features, after constructing a union of features selected with each filter method, redundant features are removed, and features are sorted according to importance:

$$\text{cluster\_}A = [ F_{CS\_new} \quad F_{ETC\_new} \quad F_{CM\_new} ] \quad (12)$$

$$F_{CS\_new} = \{ F_{CS_p} | p = 1, 2, 3, \dots, P; F_{CS_p} \in F_{CS} \} \quad (13)$$

$$F_{ETC\_new} = \{ F_{ETC_q} | q = 1, 2, 3, \dots, Q; F_{ETC_q} \in F_{ETC} \} \quad (14)$$

$$F_{CM\_new} = \{ F_{CM_r} | r = 1, 2, 3, \dots, R; F_{CM_r} \in F_{CM} \} \quad (15)$$

where  $F_{CS\_new}$ ,  $F_{ETC\_new}$ , and  $F_{CM\_new}$  are feature sets consisting of  $P$ ,  $Q$ , and  $R$  features selected using CS, ETC, and CM filter methods, respectively. Multi-filter clustering fusion can be defined as follows:

$$F_{\text{fusion}} = F_{CS\_new} \cup F_{ETC\_new} \cup F_{CM\_new} \quad (16)$$

### 3.2. Exhaustive Search Application

The next step in the feature selection process is to derive a fusion feature set,  $F_{\text{fusion}}$ , that combines the selected features, by considering the accuracy of the classification model. The algorithm used to find the best combination from among all combinations of features is an exhaustive search used in the wrapper method. In this algorithm, the fusion feature has at least two to four features. The minimum number of features is determined so that the classification model can have at least two dimensions. Up to four features are used (considering the computational time), but a larger number of features can be used. The set of all features is defined as Equation (17), and the number of feature combinations is calculated using Equation (18):

$$Y_c = \{ y_l | l = 1, 2, 3, \dots, C; y_l \in F_{\text{fusion}} \} \quad (17)$$

$$C = \sum_{s=2}^{S_s} \frac{m!}{(m-s)!s!} \quad (18)$$

where  $Y_c$  is a set of all feature combinations,  $C$  is the number of feature set combinations,  $m$  is the length of  $F_{\text{fusion}}$ , and  $s$  is the number of subset features that are combined as a target feature set, with the maximum feature combination being  $S_s = 4$ .

The process of determining the combination of these feature sets is evaluated with various classifiers, such as an SVM, KNN, and MLP. Normal and abnormal data are labeled as binary levels 0 and 1, respectively, and the accuracy of the classifiers is calculated as follows:

$$\text{Accuracy}(y, \hat{y}) = \frac{1}{S_{\text{samples}}} \sum_{l=0}^{S_{\text{samples}}} 1(\hat{y}_l = y_l) \quad (19)$$

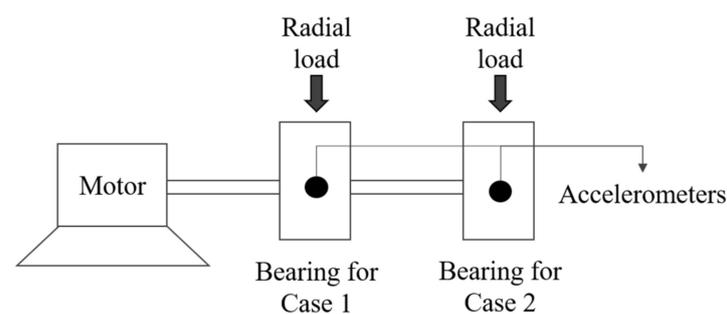
where  $y$  is the measured label values,  $\hat{y}$  is the predicted label values,  $S_{\text{samples}}$  is the amount of data, and  $1(\cdot)$  denotes an indication factor. The accuracy of the classification models using all feature combinations is tested and then sorted into a set of combinations with the highest accuracy. Equation (20) is used to obtain the best combination of features based on the highest accuracy value:

$$Y_{\text{optimum}} = \max(\text{accuracy}(Y_c)) \quad (20)$$

This study uses hyperparameter tuning to improve model accuracy, and uses 10-fold cross-validation to verify robustness. The grid search is the most representative tuning technique for computing the optimum hyperparameter value. Since it does not require much time for a small search space, and only combines a set of hyperparameters, it is simple and easy to apply. Hyperparameter optimization stops when the objective function of the hyperparameters, such as accuracy, reaches its highest value. Then, 10-fold cross-validation is performed on the generated models (based on the optimized parameters from using a grid search) and is repeated 10 times, while changing the test dataset. The accuracy of the classification model is evaluated by calculating the average accuracy for each test dataset.

#### 4. Case Studies

From four examples of failures or faults that occur in different types of rotating machinery, data were collected from experiments. Cases 1 and 2 contain experimental bearing data from NASA repositories collected from the Intelligent Maintenance System Center (IMS) [35]. Cases 3 and 4 are vibration data collected from air conditioning compressors. The experiment settings in Examples 1 through 4 are shown in Figures 2–5, and details are described in Sections 4.1 and 4.2.



**Figure 2.** Schematic of experimental setup in cases 1 and 2.

##### 4.1. Data Collection

Cases 1 and 2 are rotating machinery problems that occurred with Rexnord ZA-2115 double row bearings installed on a shaft, as illustrated in Figure 2. The rotation speed of the shaft remained constant at 2000 rpm under a radial load of 6000 lbs. The bearings operated while being lubricated (so it is considered non-dry), and failure occurred after more than 100 million revolutions.

The bearing vibrations were measured using an accelerometer, recording 20,480 points at a sampling rate of 20 kHz. The data for cases 1 and 2 were recorded at 10-min intervals, and were measured 2155 times over five weeks. Cases 1 and 2 had different causes of

bearing failure; case 1 had defects in the inner race, and case 2 had defects in the roller. Figure 2 shows an experiment schematic for cases 1 and 2.

Figures 3 and 4 are vibration data for cases 1 and 2, measured over a five-week period. The data points along the X-axis are the time index at 10-min intervals. Figure 3 shows that the failure in case 1 occurred around the end of operations, at the 1789th measurement of the 2155 measurements. Therefore, data from 1789 measurements can be classified as normal, and 366 can be classified as abnormal. On the other hand, Figure 4 shows that the failure in case 2 occurred earlier than in case 1, during the 1434th measurement. Therefore, in case 2, there were 1434 normal measurements, and 471 abnormal measurements. The normal and abnormal data were divided based on the history of the vibrations, as shown in Figures 3 and 4, and based on the threshold for kurtosis (a feature mainly used in fault classification). In addition, data segmentation was verified through an operation indicating the time when bearing abnormalities occurred [36,37].

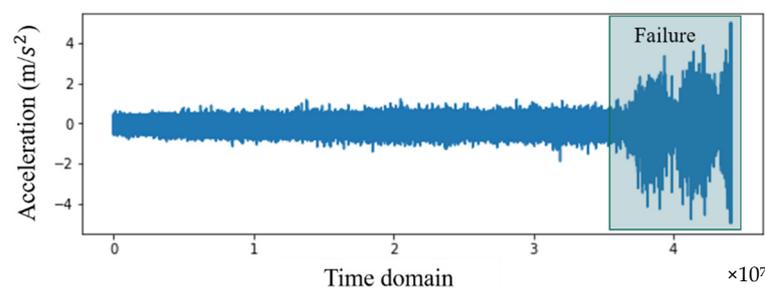


Figure 3. Raw data for case 1.

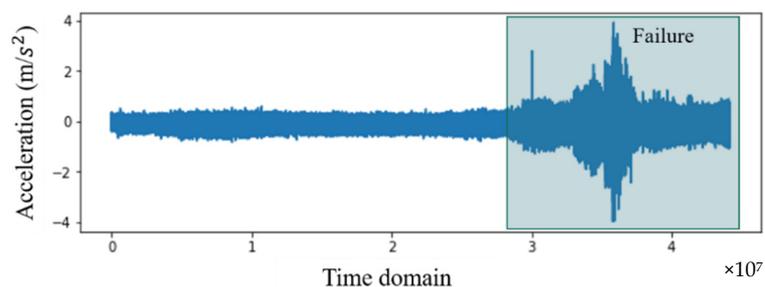


Figure 4. Raw data for case 2.

In cases 3 and 4, vibration data collected from an air conditioner compressor were used to apply the proposed method to fault data with various characteristics. The faults in these two cases were caused by two different failure modes (a mechanical defect, and lack of refrigerant inside the compressor). The machine used in cases 3 and 4 was a twin rotary compressor with low vibration and a 180° phase difference when rotating the shaft. Figure 5 shows a schematic for cases 3 and 4.

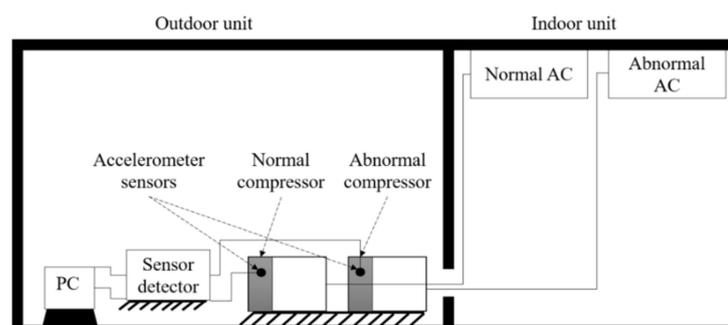


Figure 5. Schematic of the experiment setup in case 3 and case 4 [38].

The experiment was conducted in two different rooms, an outdoor unit and an indoor unit, to simulate actual conditions for using air conditioners. In cases 3 and 4, an accelerometer measured the vibrations in the compressors shown in Figure 5. Details of the experiment variables in cases 3 and 4 are shown in Tables 2 and 3, respectively.

**Table 2.** Experiment variables in case 3.

Conditions	EEV	Fan Speed (rev/min)	Frequency (Hz)
Cooling	60, 120, 180, 240, 300, 360	350, 500, 700	20, 30, 40, 50
Heating	60, 120, 180, 240, 300, 360	350, 500, 700	20, 30, 40, 50

**Table 3.** Experiment variables in case 4.

Conditions	Refrigerant (%)	Frequency (Hz)
Normal	100	30~90
Abnormal	50~90	30~90

In case 3, both normal and abnormal compressors operated, consisting of six electric expansion valve (EEV) variables  $\times$  three fan-speed variables  $\times$  four frequency variables  $\times$  two conditions = 144 measurements. To improve the accuracy of the classification model, the data were partitioned into 50 intervals of the compressor cycle, increasing to 7200 measurements in total [39]. Case 4 collected normal and abnormal data at seven frequencies, where each variable was repeated three times. A state in which the refrigerant is charged at the 100% level is considered normal, and a state in which the refrigerant level is 50–90% is considered abnormal. This indicates that the refrigerant charge gradually declined, due to continuous operation of the air conditioner. Therefore, there are 39 normal measurements and 216 abnormal measurements.

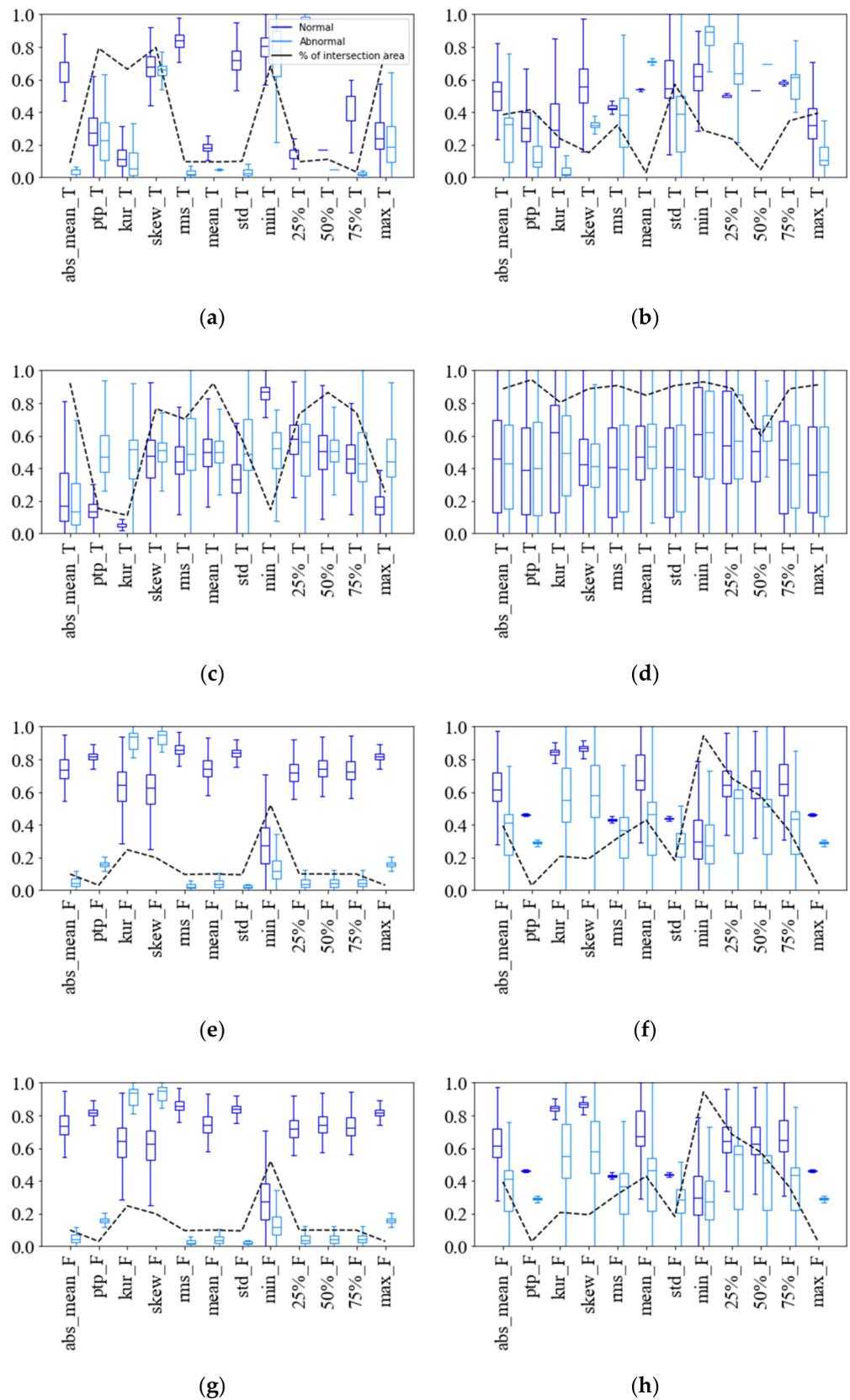
#### 4.2. Feature Extraction and Selection

The data in this study constitute the time domain and the frequency domain. The data measured in the time domain are acceleration. The data in the frequency domain were obtained by transforming time domain data using fast Fourier transform (FFT), as follows:

$$f_k = \sum_{r=0}^L x_r e^{-2\pi j \frac{kr}{L}} \quad (21)$$

where  $L$  is the length of the data sequence of  $x_r$  as input time domain, and  $k = 0, 1, \dots, L$ . As described in Table 1, 12 features were extracted from each domain, for a total of 24 features. Thus, two input feature matrices were formed into time-domain and frequency-domain combinations:  $F_t = [F1_t \dots F12_t]$ ,  $F_f = [F1_f \dots F12_f]$ , and  $F = [F_t \ F_f]$ .

After normalizing using a min–max scaler method with the data in both domains, box plots of the data for each feature can be obtained for the four cases, as shown in Figure 6, where the intersection refers to the overlapped normal and abnormal data distributions. The intersection of normal and abnormal data distributions is used to show whether each feature sufficiently distinguishes between normal and abnormal data. The smaller the intersection area, the more easily the corresponding feature classifies normal and abnormal data; the larger the intersection area, the more difficult the classification is. Thus, the intersection area can be used as an initial estimate of whether the feature is easy or difficult to classify into normal and abnormal conditions.

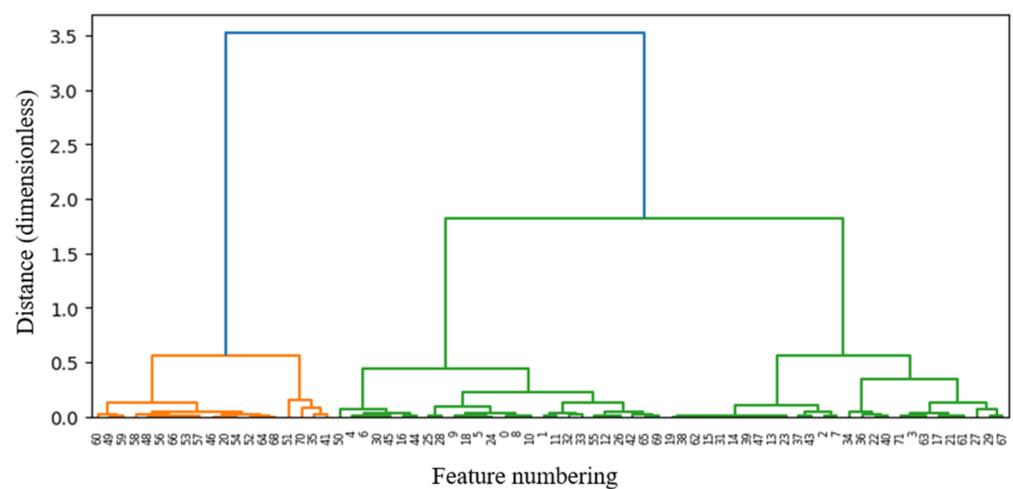


**Figure 6.** Distribution and intersection areas of normalized feature values. (a) Time domain features for case 1; (b) Time domain features for case 2; (c) Time domain features for case 3; (d) Time domain features for case 4; (e) Frequency domain features for case 1; (f) Frequency domain features for case 2; (g) Frequency domain features for case 3; (h) Frequency domain features for case 4.

As seen in Figure 6, the intersection areas vary greatly, depending on the feature types and cases. Some features have a small intersection area, which means they can clearly distinguish between normal and abnormal data, while others are not useful for fault classification. Case 1 has the smallest intersection area, indicating that failure classification is easiest. However, a large number of key features can be selected, so the number of features needs to be reduced to improve fault classification accuracy and decrease the computational time. On the other hand, the distributions of normal and abnormal data for most features are not clearly distinguished in case 3, and their intersection areas are close to 1 in both time and frequency domains, making it very difficult to derive important features. Case 4 shows that some features in the time domain are valid, but most features in the frequency domain are invalid. The results from case 4 confirm that using multiple domains rather than a single domain helps improve the accuracy of fault classification. In summary, each case had a different number and type of features extracted, due to different causes of failure and the different data characteristics. Therefore, it is necessary to correctly select the type and number of features suitable in each case.

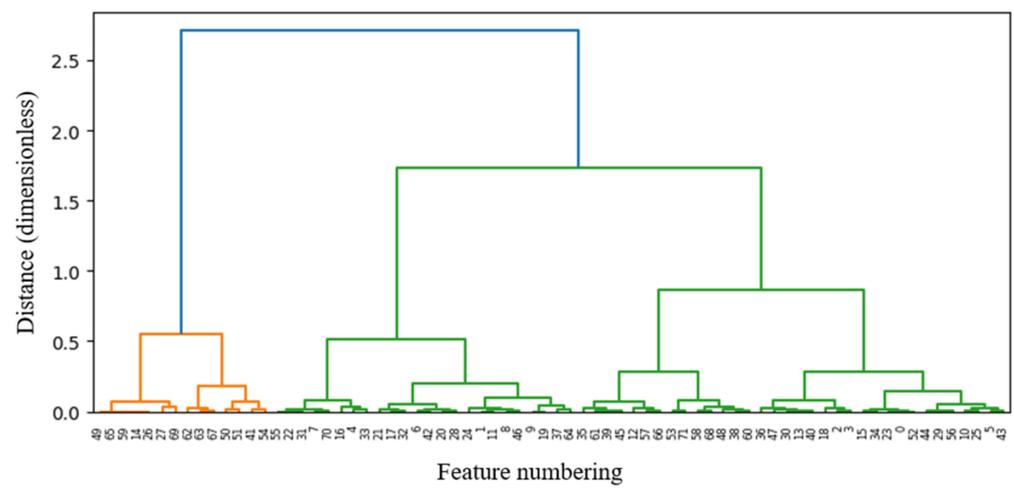
Figure 7 shows a dendrogram of the results from hierarchical clustering obtained for all features by applying MFCF in cases 1 through 4. The dendrogram represents the hierarchical relationship between the clusters, where the X-axis represents feature numbering (see Table 1) listed by importance, and the Y-axis represents the proximity of the Euclidean distance between two features. The features can mainly be clustered into two groups (orange lines and green lines). The orange lines include the main features with high importance and proximity, and the green lines include features that need to be deleted based on the three filtering methods. Cases 1 through 4 have data with different characteristics, so the types and numbers of selected features are different in all cases.

In order to remove the same features at the clustering stage, the total number of features selected before sorting via fusion of the three filter methods are 19, 14, 9, and 10 for Cases 1, 2, 3, and 4, respectively. As expected from the results in Figure 6, case 1 contains the largest number of features classified as main features. On the other hand, cases 3 and 4 have a smaller number of main features for classification, so the number of selected features is less than in cases 1 and 2. Hierarchical clustering allows users to easily derive valid features, by dividing all features into necessary and unnecessary sets. However, the number of clustered features is still large, so it needs to be further reduced. Details of the feature reduction process at each stage (multi-filter clustering, fusion, and the proposed method) are shown in Table 4.

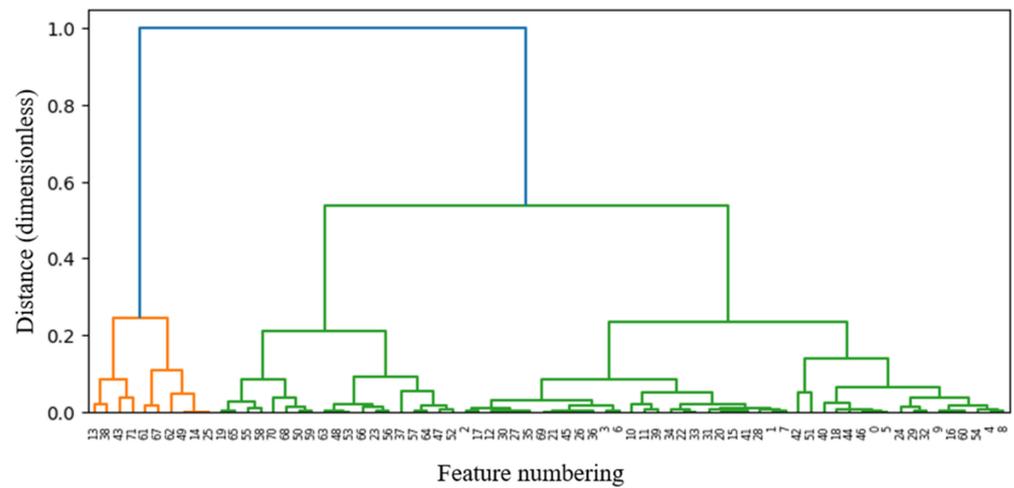


(a)

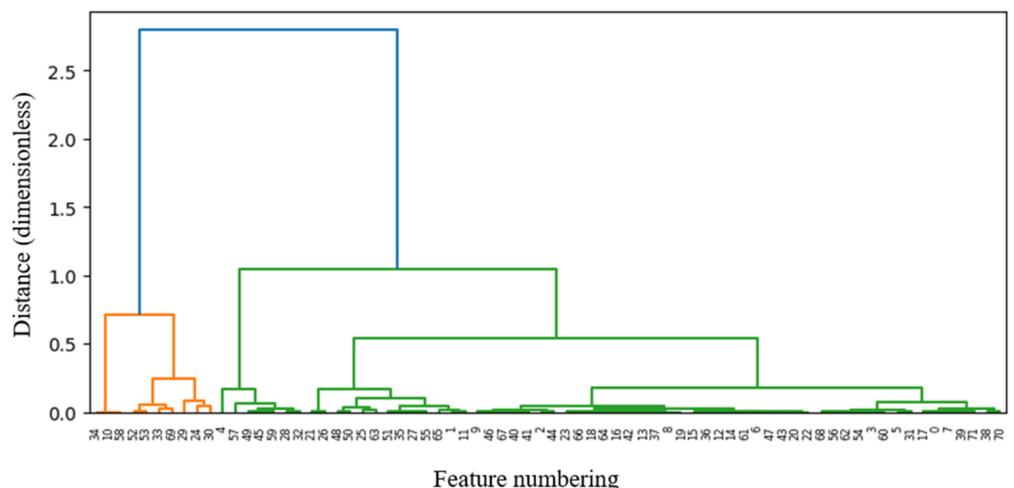
Figure 7. Cont.



(b)



(c)



(d)

Figure 7. Multi-filter clustering. (a) case 1; (b); case 2; (c) case 3; (d) case 4.

**Table 4.** Selected features at each stage.

Cases	Stages	Features	No. of Features
Case 1	Multi-filter clustering	{25%_T} ∪ {max_F, mean_T} ∪ {ptp_F, 75%_T, abs_mean_F, rms_F, mean_F, std_F, 25%_F, 50%_F, 75%_F, max_F, abs_mean_T, rms_T, std_T, 25%_T, 75%_T, skew_F}	19
	Fusion	25%_T, max_F, mean_T, ptp_F, 75%_T, abs_mean_F, rms_F, mean_F, std_F, 25%_F, 50%_F, 75%_F, abs_mean_T, rms_T, std_T, 75%_T, skew_F	17
	Final set	SVM: rms_T, 75%_T, KNN: mean_T, 75%_T, mean_F, MLP: mean_T, 75%_T, ptp_F	2,3,3
Case 2	Multi-filter clustering	{kur_T} ∪ {kur_F, skew_F, mean_T} ∪ {kur_F, skew_F, mean_T, ptp_F, std_F, max_F, kur_T, skew_T, min_T, 50%_T}	14
	Fusion	kur_T, kur_F, skew_F, mean_T, ptp_F, std_F, max_F, skew_T, min_T, 50%_T	10
	Final set	SVM: skew_F, std_F, kur_T, skew_T, KNN: kur_T, skew_F, std_F, skew_T, MLP: kur_F, max_F, skew_T, mean_T	4,4,4
Case 3	Multi-filter clustering	{ptp_T, kur_T} ∪ {kur_T, min_T} ∪ {ptp_F, ptp_T, kur_T, min_T, max_T}	9
	Fusion	ptp_T, kur_T, min_T, ptp_F, max_T	5
	Final set	SVM: ptp_T, kur_T, ptp_F, min_T, KNN: ptp_T, kur_T, ptp_F, min_T, MLP: kur_T, ptp_F, min_T, max_T	4,4,4
Case 4	Multi-filter clustering	{75%_F} ∪ {75%_F, 50%_F, rms_F, abs_mean_F, std_F} ∪ {75%_F, rms_F, mean_F, 50%_T}	10
	Fusion	75%_F, 50%_F, rms_F, abs_mean_F, mean_F, std_F, 50%_T	7
	Final set	SVM: abs_mean_F, rms_F, mean_F, KNN: abs_mean_F, mean_F, std_F, MLP: 75%_F, abs_mean_F, mean_F, std_F	3,3,4

Table 4 shows the features selected in each case with different subsets at each MFCF stage, where the three numbers in the last column of the final set indicate the number of features used in SVM, KNN, and MLP, respectively. Since data in each case are measured from different rotatory machines with different failure modes, each case has different numbers and types of important features extracted from the different domains. For example, in case 1, mean\_T was used as the input feature in the three classifiers. The acceleration time series data have a sine or cosine curve with almost the same amplitude, so mean\_T tends to have a constant value. However, since the abnormal data differ from the average values of acceleration of normal data, mean\_T may be an important feature for fault classification. In case 2, skew\_T was used as an input feature, because skewness measures the asymmetry of the probability density function of the vibration signals. In case 3, kur\_T, ptp\_F, and min\_T were used as common input features for the three classifiers, where kur\_T and ptp\_F indicate the degree of flatness of the probability density function near the center and the peak value of the signals, respectively. They are often used to measure the strength of signals, due to failure of rotating machinery. Furthermore, min\_T shows that the normal compressor condition had a low minimum value for acceleration response, compared to the minimum value under abnormal compressor conditions. In case 3, the intersection areas for many of the features are high, i.e., there are few important features except those three features shown in Figure 6, and they were used as input features in the fault classification models. Accordingly, the proposed method can be used more effectively in a problem that is difficult to classify. The most frequently selected features in case 4 are mean\_F and abs\_mean\_F. The amplitude of the vibration signal of rotatory machinery is particularly useful for distinguishing between a normal state and an abnormal state in the frequency domain. Therefore, mean\_F and ABS\_mean\_F functions were selected as common main

features. After the training process, to build the classifiers for the finally selected features in all cases through the exhaustive search, two to four feature combinations were derived.

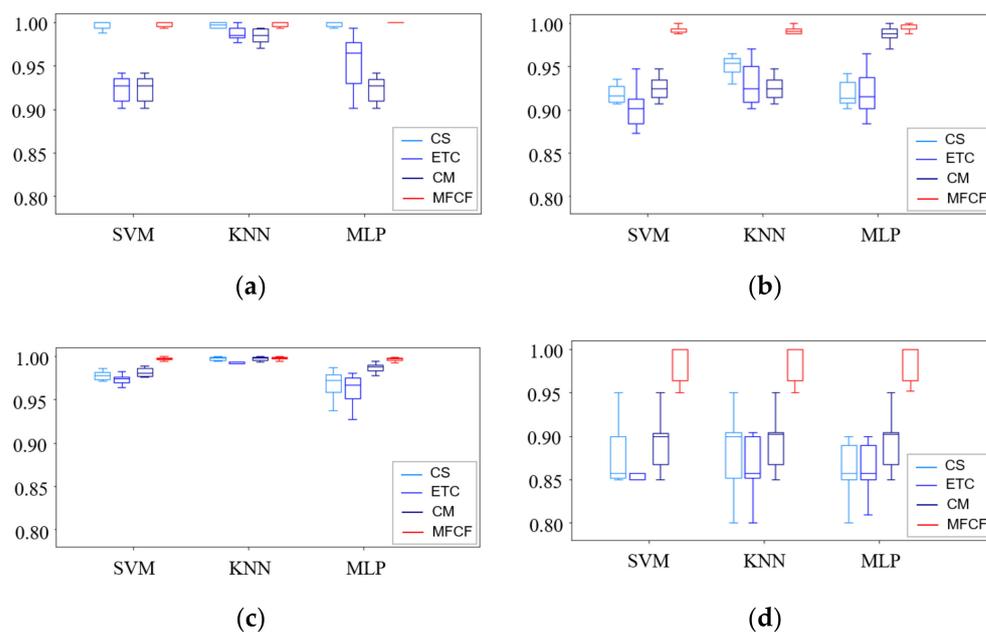
Using the finally selected features, fault classification models were generated using SVM, KNN, and MLP. Table 5 shows the accuracy and calculation times of the three single-filter methods and the proposed MFCF. The proposed method was compared with the three single-filter methods using the top three features, because it can present a good comparison by selecting the most useful features from the top three, and it can control the computational time. The accuracy of the proposed method was 1.0 (100%) for all classifiers in cases 1 and 4, and cases 2 and 3 had an average accuracy of 0.99 for all classifiers. In terms of efficiency, the proposed method consumed the least computational time compared to the others, because the exhaustive search was only performed for the selected features through MFCF, and the randomness in feature selection was low. Conversely, CS returned the lowest accuracy and required the longest running time, even though it is not much different from the other feature selection methods. In particular, the CS method yielded the lowest accuracy in case 4 (with the smallest number of samples), because it depends on sample size, and its time consumption was drastically different from the other methods, owing to the selection of classifiers. KNN was the most computationally expensive classifier. This was due to the complexity of the algorithm that stores the training data, as well as the number of iterations needed to calculate the distance between feature values. Thus, the proposed method was the most efficient, and yet had it the highest classification accuracy.

**Table 5.** Accuracy and execution times with the test data.

Methods	Case 1			Case 2			Case 3			Case 4			Avg.	
	SVM	KNN	MLP	SVM	KNN	MLP	SVM	KNN	MLP	SVM	KNN	MLP		
Accuracy	CS	0.93	0.99	0.93	0.91	0.95	0.92	0.97	0.99	0.96	0.76	0.94	0.95	0.93
	ETC	0.98	0.99	0.96	0.88	0.95	0.93	0.98	0.99	0.97	0.86	0.99	0.96	0.95
	CM	0.93	0.98	0.93	0.93	0.97	0.98	0.98	0.99	0.98	0.94	0.98	0.96	0.96
	MFCF	1.0	1.0	1.0	0.99	1.0	0.99	0.99	0.99	0.99	1.0	1.0	1.0	0.99
Efficiency (sec.)	CS	3.52	114.9	54.2	6.75	104.2	47.62	90.2	659.5	80.94	0.4	30.3	11.4	100.3
	ETC	3.01	115.4	23.9	5.23	103.6	52.7	84.1	678.9	70.89	0.4	30.8	13.4	98.5
	CM	3.5	118.6	43.3	4.99	100.9	36.6	86.5	680.8	69.27	0.4	30.5	9.1	98.7
	MFCF	2.4	113.2	13.8	4.02	100.1	39.6	83.0	655.3	81.63	0.4	30.0	10.0	94.4

These high-accuracy and low-computational times are highly advantageous for machine learning, especially when diagnosing failures in rotating machinery with many classification difficulties. The performance of the proposed method was validated by testing several cases with different characteristics, such as the number of datasets, the types of failures, the types of experimental objects, and the variables in the data collection, as described in the previous section.

To validate the classification model, 10-fold cross-validation was carried out to determine the general applicability of the proposed method. Figure 8 shows box plots for the accuracy results from CS, ETC, CM, and the proposed MFCF. Comparing each method, CS generally had low accuracy, high variability, and varying results, depending on the classifier type. CM tended to be similar to the results from CS, indicating that accuracy varies according to the classifier. ETC often had a higher accuracy than the other filter methods, but still had a lower accuracy and higher variability than the proposed method. On the other hand, the proposed method had little variability in the results, although accuracy was close to 1.0 in cases 1 to 3, where classification is easy regardless of the classifier type. However, in case 4, we can see that the lack of data resulted in lower classification accuracy and higher variability than in the other cases, but it still showed the best accuracy in comparison with the other methods.



**Figure 8.** Ten-fold cross validation. (a) case 1; (b) case 2; (c) case 3; (d) case 4.

## 5. Conclusions

This study developed a hierarchical clustering method using multiple filters (called MFCF), to extract key features from time and frequency domains, and to maximize classification accuracy by optimizing the number and type of features using an exhaustive-search-based wrapper method. MFCF enables robust, accurate, and efficient fault classification, regardless of the type of failure classification model, especially in the fault classification of rotatory machinery, including complex failure modes and different data characteristics. To validate the proposed method, vibration data from rotating machinery with four different failure modes were used, and cross-validation results confirmed that it had the best classification performance, compared to the other filter methods. Although the proposed method in this study was used for the problem of classifying normal measurements and those with abnormalities, it will be applied in the future to problems including multi-classification and multi-domain features, to verify its general applicability to broad engineering applications. In addition, this study obtained vibration signals using only accelerometer sensors, but the proposed method will be applied to extract features of data collected using various sensors, such as chemical and temperature sensors in the future.

**Author Contributions:** Conceptualization, S.M. and Y.N.; data acquisition, S.M. and S.P.; methodology, S.M.; coding, S.M.; validation, S.M.; formal analysis, S.M. and Y.-J.K.; investigation, S.M., resources, J.L. and S.C.; project administration, Y.N.; funding acquisition, Y.N.; writing-original draft preparation, S.M. and Y.N.; writing-review and editing, Y.N. and Y.-J.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. 2018R1D1A1A02086093) and the Korea government (MSIT) (2020R1A5A8018822 and 2021R1A2C1013557), and LG electronics.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Cheng, J.; Yang, Y.; Hu, N.; Cheng, Z.; Cheng, J. A noise reduction method based on adaptive weighted symplectic geometry decomposition and its application in early gear fault diagnosis. *Mech. Syst. Signal Process.* **2021**, *149*, 107351. [[CrossRef](#)]
2. Li, X.; Yang, Y.; Hu, Z.; Cheng, J.; Cheng, J. Discriminative manifold random vector functional link neural network for rolling bearing fault diagnosis. *Knowl.-Based Syst.* **2021**, *211*, 106507. [[CrossRef](#)]
3. Fuli, Y.L.; Jia, W.M.; Qi, Y. Centrifugal compressor fault diagnosis based on qualitative simulation and thermal parameters. *Mech. Syst. Signal Process.* **2016**, *8*, 259–273. [[CrossRef](#)]
4. Li, X.; Yang, Y.; Pan, H.; Cheng, J.; Cheng, J. A novel deep stacking least squares support vector machine for rolling bearing fault diagnosis. *Comput. Ind.* **2019**, *110*, 36–47. [[CrossRef](#)]
5. Yang, B.; Lei, Y.; Jia, F.; Xing, S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mech. Syst. Signal Process.* **2019**, *122*, 692–706. [[CrossRef](#)]
6. Li, X.; Yang, Y.; Shao, H.; Zhong, X.; Cheng, J.; Cheng, J. Symplectic weighted sparse support matrix machine for gear fault diagnosis. *Measurement* **2021**, *168*, 108392. [[CrossRef](#)]
7. Gao, Y.; Yu, D. Total variation on horizontal visibility graph and its application to rolling bearing fault diagnosis. *Mech. Mach. Theory* **2020**, *147*, 103768. [[CrossRef](#)]
8. Youwei, W.; Lizhou, F. Hybrid feature selection using component co-occurrence based feature relevance measurement. *Expert Syst. Appl.* **2018**, *102*, 83–99.
9. Uysal, A.K.; Gunal, S. A novel probabilistic feature selection method for text classification. *Knowl.-Based Syst.* **2021**, *36*, 226–235. [[CrossRef](#)]
10. Dimitrios, E.; Avi, A. An evaluation of feature selection methods for environmental data. *Ecol. Inform. J.* **2021**, *61*, 101224.
11. Abdelhamid, N.; Thabtah, F.; Abdel-Jaber, H. Phishing detection: A recent intelligent machine learning comparison based on models content and features. In Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics: Security and Big Data (ISI), Beijing, China, 22–24 July 2017; pp. 72–77.
12. Kamalov, F.; Thabtah, F. A feature selection method based on ranked vector scores of features for classification. *Ann. Data Sci.* **2017**, *4*, 483–502. [[CrossRef](#)]
13. Sánchez, R.; Lucero, P.; Vasquez, R.; Cerrada, M.; Macancela, J.; Cebrera, D. Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN. *J. Intell. Fuzzy Syst.* **2018**, *34*, 3463–3473. [[CrossRef](#)]
14. Ziani, R.; Mahgoun, H.; Fedala, S.; Felkaoui, A. Feature selection scheme based on Pareto method for gearbox fault diagnosis. *Signal Process. Appl. Rotating Mach. Diagn.* **2017**, *12*, 1–15.
15. Zhang, X.; Zhang, Q.; Chen, M.; Sun, Y.; Qin, X.; Li, H. A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method. *Neurocomputing* **2018**, *275*, 2426–2439. [[CrossRef](#)]
16. Hui, K.; Ooi, C.; Lim, M.; Leong, M.; Al-Obaidi, S. An improved wrapper-based feature selection method for machinery fault diagnosis. *PLoS ONE* **2017**, *12*, e0189143. [[CrossRef](#)]
17. Lu, N.; Zhang, G.; Xiao, Z.; Malik, O.P. Feature extraction based on adaptive multiwavelets and LTSA for rotating machinery fault diagnosis. *Shock. Vib.* **2019**, *2019*, 1201084. [[CrossRef](#)]
18. Cerrada, M.; Sánchez, M.R.V.; Cabrera, D.; Zurita, G.; Li, C. Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal. *Sensors* **2015**, *15*, 23903–23926. [[CrossRef](#)]
19. Huang, K.; Wu, S.; Li, F.; Yang, C.; Gui, W. Fault diagnosis of hydraulic systems based on deep learning model with multirate data samples. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–13. [[CrossRef](#)]
20. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial learning. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7957–7967. [[CrossRef](#)]
21. Kang, L.C.; Choon, L.T.; Kok, S.W.; Kelvin, S.C.Y.; Wei, K.T. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. *Inf. Sci.* **2019**, *484*, 153–166.
22. Amir, E.; Adel, K.; David, Y. Data-driven fault detection and diagnosis for packaged rooftop units using statistical machine learning classification methods. *Energy Build.* **2020**, *225*, 110318.
23. Andrea, B.; Xudong, S.; Bernd, B.; Jörg, R.; Michel, L. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data Anal.* **2020**, *143*, 106839.
24. Liu, C.; Jiang, D.; Yang, W. Global geometric similarity scheme for feature selection in fault diagnosis. *Expert Syst. Appl.* **2014**, *41*, 3585–3595. [[CrossRef](#)]
25. Cortizo, J.C.; Giraldez, I. Multi criteria wrapper improvements to naive bayes learning. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Burgos, Spain, 20–23 September 2006; pp. 419–427.
26. Wang, Y.; Feng, L.; Li, Y. Two-step based feature selection method for filtering redundant information. *J. Intell. Fuzzy Syst.* **2017**, *33*, 2059–2073. [[CrossRef](#)]
27. Zorarpaci, E.O.; Selma, A. A hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Syst. Appl.* **2016**, *62*, 91–103. [[CrossRef](#)]
28. Samanta, B. Gear fault detection using artificial neural network and support vector machines with genetic algorithms. *Mech. Syst. Signal Process.* **2004**, *12*, 625–644. [[CrossRef](#)]
29. Rafiee, J.; Arvani, F.; Harifi, A.; Sadeghi, M.H. Intelligent condition monitoring of a gearbox using artificial neural network. *Mech. Syst. Signal Process.* **2007**, *21*, 1746–1754. [[CrossRef](#)]

30. Wu, J.D.; Hsu, C.C. Fault gear identification using vibration signal with discrete wavelet transform technique and fuzzy-logic inference. *Expert Syst. Appl.* **2009**, *36*, 3785–3794. [[CrossRef](#)]
31. Dong, X.; Wang, C.; Si, W. ECG beat classification via deterministic learning. *Neurocomputing* **2017**, *240*, 112. [[CrossRef](#)]
32. Jha, C.K.; Kolekar, M.H. Cardiac arrhythmia classification using tunable Q-wavelet transform based features and support vector machine classifier. *Biomed. Signal Process. Control* **2020**, *59*, 101875. [[CrossRef](#)]
33. Sun, K.; Wu, X.; Xue, J.; Ma, F. Development of a new multi-layer perceptron based soft sensor for SO<sub>2</sub> emissions in power plant. *J. Process Control* **2019**, *84*, 182191. [[CrossRef](#)]
34. Wang, Z.Y.; Lu, C.; Zhou, B. Fault diagnosis for rotary machinery with selective ensemble neural networks. *Mech. Syst. Signal Process.* **2018**, *113*, 112–130. [[CrossRef](#)]
35. IMS Bearings Dataset. 2021. Available online: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/> (accessed on 1 January 2021).
36. Ali, J.B.; Saidi, L.; Mouelihi, A.; Chebel-Morello, B.; Fnaiech, F. Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations. *Eng. App. Artif. Intell.* **2015**, *42*, 67–81.
37. Duong, B.P.; Khan, S.A.; Shon, D.; Im, K.; Park, J.; Lim, D.S.; Jang, B.; Kim, J.M. A reliable health indicator for fault prognosis of bearings. *Sensors* **2018**, *18*, 3740. [[CrossRef](#)]
38. Mochammad, S.; Kang, Y.-J.; Noh, Y.; Park, S.; Ahn, B. Stable hybrid feature selection method for compressor fault diagnosis. *IEEE Access* **2021**, *9*, 97415–97429. [[CrossRef](#)]
39. Kim, S.; Noh, Y.; Kang, Y.-J.; Park, S.; Ahn, B. Fault classification model based on time domain feature extraction of vibration data. *J. Comp. Struct. Eng. Inst. Korea* **2021**, *34*, 25–33. [[CrossRef](#)]