Article

Visibility Graph Feature Model of Vibration Signals: A Novel Bearing Fault Diagnosis Approach

Zhe Zhang [1,2]**, Yong Qin** [1,2,]*****, Limin Jia** [1,2,]***** and Xin'an Chen** [1,2]

[1] State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China; zhangzhe@bjtu.edu.cn (Z.Z.); 15114217@bjtu.edu.cn (X.C.)

[2] Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University, Beijing 100044, China

***** Correspondence: yqin@bjtu.edu.cn (Y.Q.); lmjia@bjtu.edu.cn (L.J.); Tel.: +86-010-5168-4081 (L.J.)

check for
updates

Abstract: Reliable fault diagnosis of rolling bearings is an important issue for the normal operation of many rotating machines. Information about the structure dynamics is always hidden in the vibration response of the bearings, and it is often very difficult to extract them correctly due to the nonlinear/chaotic nature of the vibration signal. This paper proposes a new feature extraction model of vibration signals for bearing fault diagnosis by employing a recently-developed concept in graph theory, the visibility graph (VG). The VG approach is used to convert the vibration signals into a binary matrix. We extract 15 VG features from the binary matrix by using the network analysis and image processing methods. The three global VG features are proposed based on the complex network theory to describe the global characteristics of the binary matrix. The 12 local VG features are proposed based on the texture analysis method of images, Gaussian Markov random fields, to describe the local characteristics of the binary matrix. The feature selection algorithm is applied to select the VG feature subsets with the best performance. Experimental results are shown for the Case Western Reserve University Bearing Data. The efficiency of the visibility graph feature model is verified by the higher diagnosis accuracy compared to the statistical and wavelet package feature model. The VG features can be used to recognize the fault of rolling bearings under variable working conditions.

Keywords: rolling bearing; nonlinear vibration signals; visibility graph features; Gaussian Markov random fields

1. Introduction

Nowadays, rotating machines are used widely in industrial systems, and they have become the most critical equipment in many industrial systems. In the main components of rotating machines, the fault of rolling bearings is the most frequent reason for unexpected machine breakdown and results in economic loss [1]. Therefore, the material control in the quality design process [2,3] and the fault diagnosis of rolling bearings in the operation process have attracted considerable attention of engineers and scholars in order to reduce the system breakdown. This paper focuses on the fault diagnosis study of rolling bearings in the operation process.

Several vibration and acoustic measurement-based methods have been developed for the detection of defects in rolling element bearings. To the best of our knowledge, the vibration-based diagnosis method is the most widely employed because vibration signals contain a wealth of information about the structure dynamics [4–6]. The most common steps of fault diagnosis are: collect the vibration signals using sensors, extract the fault features through a signal processing method [7] and finally detect or recognize the fault classes through classifiers [1]. In order to get accurate diagnosis results and

develop a more efficient diagnosis method, many feature models have been developed. For example, the statistical feature model can produce a time-domain and frequency-domain statistical feature set, and these features are exclusively used to recognize faults of bearings and other machines [8,9]. Complex envelope analysis and wavelet packet analysis are used to get the envelope features and wavelet packet features [10–13]. In recent years, the image features have also been extracted from the vibration spectrum to detect the faults of bearings [1,14].

Researchers often overlook the feature classification and arbitrarily choose the number of features [15]. However, it is not certain that all the features contribute to fault diagnosis. Further, the irrelevant features increase the measurement and storage requirements, increase computing cost and may lead to the poor prediction performance of classifiers [16]. Therefore, the necessary processing step before fault diagnosis is to select good features through feature selection (FS). Usually, the relevant features are selected and the redundant features are eliminated after FS [17]. A survey on the FS can be found in [18,19].

Artificial intelligence (AI) has been developed and applied to reduce the breakdowns of manufacturing systems [20,21]. Researchers prefer to use machine learning-based classifiers to estimate the classification performance of their proposed features, even though some works only use visual inspection of peaks in frequency graphs for fault diagnosis [22]. The artificial neural network (ANN) was developed to recognize the fault classes of rolling bearings, and the time-domain features were used as the input of the ANN classifier in [23]. However, the various loading conditions make the ANN task very complicated. The support vector machine (SVM) is another classifier widely used in the fault diagnosis of rolling bearings. SVM classifies better than ANN because of the principle of risk minimization [24]. The inputs of SVM used in previous work are the intrinsic mode function envelop spectrum, time- and frequency-domain features [25,26].

In sum, feature extraction and selection are very important for fault diagnosis of rolling bearings because the classification performance depends on the quality of the dataset and largely the features used. Although many feature models have been developed, it is still very difficult to extract them correctly due to the nonlinear/chaotic nature of the vibration signal, so the number of features is growing now in order to recognize the fault or estimate the health of rolling bearings more precisely and efficiently [27].

It is well known that the vibration signals are one kind of time series data acquired by sensors. Recently, a new and simple method was introduced to convert a time series to a network, called the visibility graph (VG). It has been proven that the network structure resulting from the VG approach inherits some of the properties of the time series data [28]; for example, periodic series resulting in regular graphs, random series resulting in random graphs and fractal series converting into scale-free networks [29]. The visibility graph concept comes from the space and geometry theory; each node of the graph represents a location in a special space, and the edge between two nodes shows that the two nodes can see each other [30]. For the time series data, the nodes represent the data values in a planar coordinate system, and the edge between two nodes shows that the two data values can see each other. After the VG was proposed, researchers presented improved versions of the VG approach such as the horizontal visibility graph (HVG) [31], the limited penetrable visibility graph (LPVG) [32,33] and the parametric natural visibility graph (PNVG) [34]. The VG approach has been studied increasingly in many fields such as economic/marketing [35,36], ecology [37] and health analysis [38,39]. However, the VG approach has not been used in fault diagnosis of mechanical components such as gear and bearings. Furthermore, it has been shown that the VG approach has a noise resistance ability [40]. Therefore, we investigate whether VG can be used as an effective tool for diagnosing rolling bearing faults in this paper.

The graph analysis method has been applied in many areas including traffic systems [41], biology [42] and communication [43], except vibration signal processing. Using the VG approach, we can obtain a graph that has special topological properties. It allows us to use the complex graph approach to analyze the vibration signals. Fortunately, the features or properties of the complex graph

have been discussed in previous studies. The features to measure the hidden information in the graph include graph density, degree distribution [44], average path length [45], graph diameter [46], the clustering coefficient, etc. [47].

However, there is little research to bridge graph analysis and signal-based fault diagnosis, and the graph analysis method is also limited. In this paper, the VG approach is used successfully to transform the vibration signals into a graph or a binary matrix. The global VG features are proposed based on the graph analysis, and the local VG features are proposed based on the texture analysis method of images. Therefore, a new feature model of vibration signals is formulated in this paper.

In this paper, we present a novel nonlinear feature model that can extract the visibility graph (VG) features from the time series vibration signals. A novel signal processing method is proposed. The vibration signals of rolling bearings are converted into a binary matrix. The vibration signals are firstly analyzed from the viewpoint of graph and image processing. A new set of VG features is extracted from the vibration signals. The sequential feature selection (SFS) algorithm is applied to select the VG feature subset that can realize the best diagnosis performance. The artificial neural graph (ANN), K-nearest neighbor (KNN) [48,49] and support vector machine (SVM) are used as classifiers for the VG features. The effectiveness of the selected VG features in the fault diagnosis of the rolling bearings is verified by comparison with the statistical feature model and wavelet package feature model. Because the working condition and the fault level are also considered in the experiments, the VG features can be used to recognize the fault class and the corresponding fault level of rolling bearings under variable working conditions.

To motivate this work, the Case Western Reserve University (CWRU) Bearing Data [22] are used in the experiments. The study begins with presenting the methodological framework of VG feature selection in Section 2. According to the framework, firstly, the method to convert the data into a matrix is proposed in Section 3; secondly, the VG feature extraction and selection method is proposed in Section 4. An experiment is presented in Section 5 in order to draw some conclusions in Section 6.

2. Methodological Framework

The framework to perform the bearing fault diagnosis in this paper is illustrated in Figure 1. The sequence of signal processing steps is inspired by [50] and can be described in sequence as follows:

- Data segmentation: The signals are segmented according to the sample rate and rough shaft speed to ensure each obtained sample covers several circles of signals.
- Feature extraction: The visibility graph method is used to convert the acceleration signals into a binary matrix. From the matrix obtained, considering the feature model of the complex graph and image, we extract the feature vector within this model. We do this for all available VG feature models and produce the VG feature pool.
- Feature selection: We select the optimal number of VG features based on the diagnosis performance. Multiple classifiers can be used to estimate the performance of features.
- Results analysis: The advantage of the VG feature model is validated by comparing the performance of the VG feature model, the statistical feature model and the wavelet package feature model.
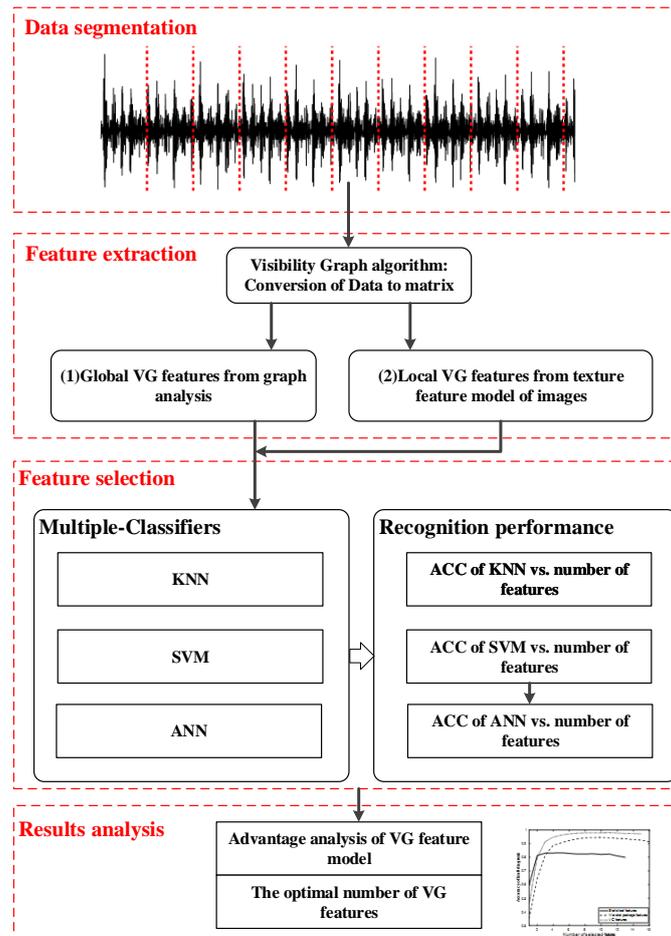
Figure 1. Framework of the VG feature model for bearing fault classification.

3. Visibility Graph Construction

We denote the vibration signals by $s(i)$, where $i = 1, 2, ..., N$ is a discrete time step indexing the time of collecting each signal. In order to process the data using our method, we have applied the visibility graph approach for mapping vibration signals in a corresponding complex graph in the experiments. The visibility graph method has been investigated in previous studies; however, the application of this approach is bounded. The visibility algorithm is a map that assigns each signal point to a node/vertex in a complex graph. Two nodes will be connected whenever one can draw a line in the time series space without intersecting any intermediate signal height. For the vibration signals, two signals $(i, s(i))$ and $(j, s(j))$ will have visibility and consequently will become two connected nodes of the associated graph, if any other data $(k, s(k))$ placed between them fulfill:

$$s(k) < s(j) + (s(i) - s(j))\frac{i - k}{j - i} \tag{1}$$

The visibility graph can reflect the characteristics of the signal envelope. If the two signals can be connected according to Equation (1), the envelop between them is concave, otherwise the envelop is convex because there are some larger signals between them. Figure 2 depicts the relationship between the envelope and visibility graph. Signal s_1 cannot be connected to signal s_2, and signal s_3 cannot be connected to signal s_4 (dashed lines) because the signals under the convex envelope curves (EC) obstruct the visibility between them. Signal s_2 can be connected to signal s_3, and signal s_4 can be connected to signal s_5 (solid lines) because the signals under the concave envelope curves cannot obstruct the visibility between them.
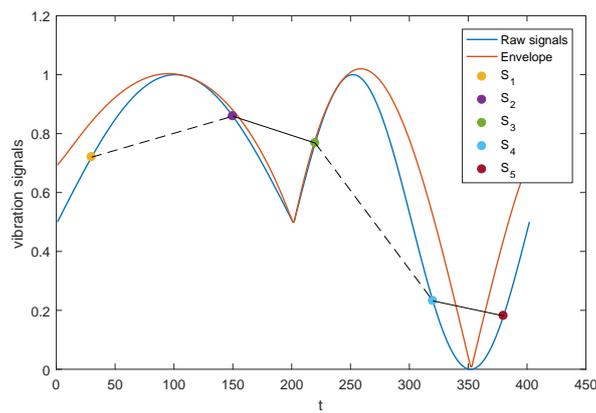
Figure 2. Relationship between the signal envelope and the visibility graph.

Therefore, we can use Jensen's inequality [51] to describe the relationship between the visibility graph and the signal envelop. Let $s(t)$ be the signal envelope connecting signals $(i, s(i))$ and $(j, s(j))$. The two signals can be connected if $s(\alpha i + (1 - \alpha)j) < \alpha s(i) + (1 - \alpha)s(j)$ for any $\alpha \in [0, 1]$. Using the VG algorithm, the acceleration signal is mapped into an ordered graph with a special spatial structure.

The obtained visibility graph can be represented by its adjacent matrix M (symmetric matrix), whose elements $m_{i,j} = 0$ or 1. If the signals s_i and s_j can be connected according to Equation (1), $m_{i,j} = 1$, otherwise, $m_{i,j} = 0$, that is:

$$m_{i,j} = \begin{cases} 1, & s(k) < s(j) + (s(i) - s(j))\frac{i-k}{j-i}, i < k < j \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Based on the obtained adjacent matrix of the visibility graph (VGAM), firstly, we can extract visibility graph features from VGAM, for example graph density and graph index complexity; however, the graph features prefer to describe the global characteristics of VGAM rather than local properties. Meanwhile, the VGAM can also be seen as binary images (Figure 3), which only include zero and one. Therefore, the local properties of VGAM can be extracted by using the image processing methods. The texture features of VGAM resulting from the GMRF method are extracted in this paper because the GMRF method performs better than other methods without considering the rotated invariant [52]. Therefore, the VG features including global and local features of VGAM can be extracted for bearing fault diagnosis.
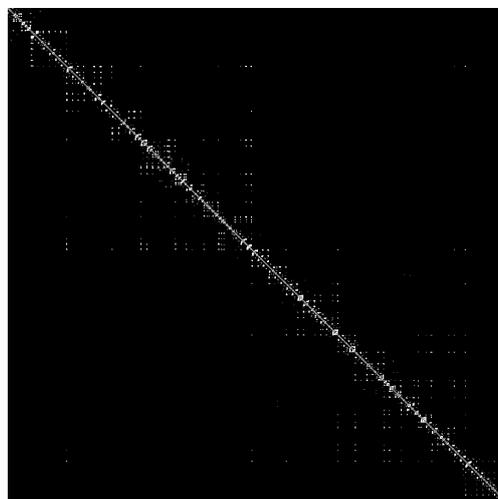

Figure 3. The binary image of the adjacent matrix of the visibility graph (VGAM).

4. VG Features' Extraction and Selection

In this section, the candidate VG features and the feature selection algorithm are described. We can construct VG associated with the temporal acceleration signal of rolling bearing vibration by using the method mentioned above. However, the graph cannot be directly analyzed numerically. Therefore, we gleaned the VG features from previous studies as much as possible. The graph density [53], degree distribution [54] and graph index complexity are selected to be global VG features, and the Gaussian Markov random fields (GMRF) are used to produce the local VG features.

4.1. Candidate VG Features

The candidate VG feature set is composed of the global VG feature subset and the local VG feature subset. The global VG features come from graph theory. The three global VG features used in this paper include the graph density, standard deviation of node degree and graph index complexity, and the 12 local VG features used in this paper refer to the 12 GMRF parameters from the five-order traditional GMRF analysis of VGAM.

4.1.1. Global VG Features

(1)   VG density:

The VG density (VGD) reflects the size of graph or its adjacent matrix. Let G denote the visibility graph transformed from vibration signals. The density G_d of G is defined as a ratio of the number of edges to the number of possible edges in the graph. Let Ddenote the VGAM, the VG density can also be represented by the VGAM, that is:

$$G_d = \frac{\sum_{i,j} m_{i,j}}{N(N-1)} \tag{3}$$

If the signal envelope is composed of much longer concave EC than convex EC, more signals can be connected, and thus, the VG density is larger, otherwise, the VG density is lower.

(2)   VG complexity:

The VG complexity (VGC) is to signify the global complexity of the complex graph structure and adjacent matrix. Let λ_{max} be the maximum eigenvalue of the adjacency matrix of VG; VGC is defined as follows [55]:

$$VGC = 4\kappa(1 - \kappa) \tag{4}$$

where:

$$\kappa = \frac{\lambda_{max} - 2cos(\pi/(N+1))}{N - 1 - 2cos(\pi/(N+1))} \tag{5}$$

in which $2cos(\pi/(N+1)) \leq \lambda_{max} \leq n - 1$ [56]. It has been proven that that the graph complexity strongly depends on the number of edges. Unlike the VG density, a medium number of edges alone already guarantees a high VGC [55]. Further, the VGC index has been also used to characterize electroencephalograms [38]. Therefore, we can use the VGC to distinguish the graph with a medium number of edges, which is determined by the distribution of convex and concave EC, as described in Section 4.1.1.

(3)   VG degree:

The VG degree is the number of connections or edges the signal i has to the other nodes. For the vibration signals, a smaller or larger value (compared with other points) makes them visible or invisible by many other points, which in turn causes their corresponding nodes to be or not to be the hot spot of the graph with many connections. The degree of signal i can be formulated as:

$$D(i) = \sum_j m_{i,j} \tag{6}$$

However, the VG degree is array data, so feature preprocessing is thus needed [57], and the mean value and standard deviation of the VG degree are suggested in this paper. Because the mean value of the VG degree is $N - 1$-times VG density G_d, only the standard deviation of the VG degree is adopted in the experiments.

4.1.2. Local VG Features

The local VG features are extracted by considering the binary matrix as the gray-level matrix of binary images. Gaussian Markov random fields (GMRF) have been shown to perform better than other models in both the classification and segmentation of textured images [52,58]; therefore, we use the GMRF to extract the local VG features. In the GMRF model, the $d_{i,j}$ in VGAM D is represented by $y(s)$.

$$y(s), s \in \Omega, \Omega = s = (i,j) : 0 \leq i, j \leq N - 1 \tag{7}$$

for the $N \times N$ VGAM D. The Markov random field models are described by the conditional probability $P(y(s)|y(\eta_s^n))$, and $\eta_s^n = s + r, r \in \eta^n$ is an n-th-order symmetric neighbor set of site s. The first- to fifth-order neighbor MRF relationships are depicted in Figure 4.

θ_1	θ_9	θ_6	θ_{10}	θ_{12}
θ_7	θ_3	θ_2	θ_4	θ_8
θ_5	θ_1	S	θ_1	θ_5
θ_8	θ_4	θ_2	θ_3	θ_7
θ_{12}	θ_{10}	θ_6	θ_9	θ_1

Figure 4. The parameters of the Markov random fields (MRF) model for the first-order neighbor to the fifth-order neighbor.

The GMRF model assumes that the $y(s)$ obey the following equation:

$$y(s) = \sum_{r \in \eta_s} \theta_r y(s + r) + e(s) \tag{8}$$

where the η_s is a neighbor set dependent on the order and type of model used and θ_r is the GMRF parameter for neighbor r and characterizes the local properties of VGAM. The parameter set θ_r satisfies:

$$\theta_r = \theta_{-r}, r \in \eta_s \tag{9}$$

$e(s)$ is a stationary Gaussian noise sequence defined by:

$$e(s) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{y(s)^2}{2\sigma^2}) \tag{10}$$

Equation (8) can be also represented by the matrix:

$$y(s) = \theta^T Q_s + e_s \tag{11}$$

where θ is a vector comprised of θ_r and Q_s is a vector defined by:

$$Q_s = [y_{s+r1} + y(s-r1), ..., y(s+rn) + y(s-rn)]^T \tag{12}$$

Then, the vector θ and variance σ can be estimated using the least-squares (LSQR) approach presented in the following equations:

$$\hat{\theta} = \left[\sum_{s\in\Omega} Q_s Q_s^T\right]^{-1} \left[\sum_{s\in\Omega} Q_s y(s)\right] \tag{13}$$

$$\hat{\sigma} = \frac{1}{L^2} \sum_{s\in\Omega} \left[y(s) - \hat{\theta}^T Q_s\right] \tag{14}$$

where L denotes the length of $y(s)$. In this paper, we adopt the fifth-order GMRF, which produces 12 GMRF parameters $\hat{\theta}$.

In summary, three global and 12 local VG features are extracted. However, it is not certain that all the VG features are necessary for fault diagnosis because some VG features may reflect similar properties of vibration signals; therefore, we should select the necessary features that greatly contribute to the fault diagnosis.

4.2. Diagnosis Performance-Based VG Feature Selection

In this section, our goal is to select the necessary features that greatly contribute to the fault class diagnosis. The SFS algorithm [59] is applied to the VG feature selection. The algorithm starts with an empty set and adds one VG feature for the first step, which gives the highest value for the objective function. From the second step onwards, the remaining VG features are added individually to the current subset, and the new subset is evaluated.

A multiple-classifier including support vector machine (SVM), k-nearest neighbor (KNN) and artificial neural network (ANN) classifier are applied to detect the bearing fault classes and produce the significance of each VG feature with regard to the updated feature subset. The diagnosis accuracy (ACC) is selected as the criteria to judge the classification performance. The accuracy (ACC) or its complement, i.e., the error rate, is the most easily understandable index. Let ξ denote the ACC, E denote the number of testing samples and E_r denote the number of testing samples that are classified correctly; ξ is equal to E_r/E.

The significance of each VG feature f_i with regard to the updated feature subset F_c is the difference between the ACC ξ obtained by using VG feature subsets F_c and $F_c \cup f_i$. The individual VG feature with the most significance is permanently included in the subset F_c in each step.

5. Experiments

5.1. Data and Experiments' Description

The two datasets used in this paper come from the Case Western Reserve University (CWRU) Bearing Data Centre and Intelligent Maintenance Systems (IMS) bearing run-to-failure dataset because the two datasets have been widely used in previous studies [4,50,60–62]. In the IMS dataset, 12 bearings in the experimental setup were tested in the same conditions of speed and load. Despite that, only four bearings were broken. The total number of records is 5394. Each record is formed by 20,480 signals. Because of the overlap between normal-degradation zones and degradation-faulting zones, the number of records is reduced to 3000. Three kinds of fault type including normal bearings, outer race fault and inner race fault are selected in the experiments. More information about the data structure can be found on the website (http://data-acoustics.com/measurements/bearing-faults/bearing-4/).

The CWRU dataset consists of time series of the vibration acceleration value (vibration signal) measured from a sensor mounted on the bearing housing at the drive end (DE), fan end (FE)

and base of the induction motor. The vibration signals of the rolling bearings were obtained from different states—(1) normal operating conditions (NM); (2) inner race fault (IR); (3) rolling ball fault (RB); (4) outer race fault (OR)—and recorded for motor loads of 0–3 horsepower (motor speeds of 1797–1720 RPM). Fault levels ranging from 0.007 inches in diameter to 0.0210 inches in diameter were introduced separately for each fault class. More information about the data structure can be found on the website (http://csegroups.case.edu/bearingdatacenterpages/download-data-file). The dataset used in the experiment is acquired by the sensor at the DE when the bearing at the DE has faults because the sensor at the FE can detect the faults at the DE with less confidence, and the sampling frequency is 12 kHz.

In the experiment, 1000 signals are considered as one sample. Tables 1 and 2 list the machine condition classes defined for the experiments. Table 1 lists machine condition classes of IMS data, while Table 2 lists machine condition classes of CWRU data. Note that the number of classes is much more extensive than that found usually in work related to the CWRU data. We tried to distinguish among the different bearing fault locations and the severity of the fault (i.e., the diameter of the artificially-drilled hole into the material). Even different loads for the same fault were considered as separate classes, which constitutes the most challenging classification problem. There are 21 classes to be distinguished, and the details about all classes are shown in Table 2.

Table 1. Class distribution and description of the Intelligent Maintenance Systems (IMS) dataset.

Class	Name	Samples	Distribution	Data Description
1	NM	100	33.33%	normal bearings
2	OR	100	33.33%	Outer race fault
3	IR	100	33.33%	Inner race fault

Table 2. Class distribution and description of the Case Western Reserve University (CWRU) dataset. RB, rolling ball.

Class	Name	Samples	Distribution	Data Description
1	NM_0	100	2.083%	NM load = 0
2	NM_1	100	2.083%	NM load = 1
3	NM_2	100	2.083%	NM load = 2
4	NM_3	100	2.083%	NM load = 3
5	IR007	400	8.333%	IR fault level = 0.007
6	IR014_0	100	2.083%	IR fault level = 0.014 load = 0
7	IR014_1	100	2.083%	IR fault level = 0.014 load = 1
8	IR014_2	100	2.083%	IR fault level = 0.014 load = 2
9	IR014_3	100	2.083%	IR fault level = 0.014 load = 3
10	IR021	400	8.333%	IR fault level = 0.021
11	OR007	400	8.333%	OR fault level = 0.007
12	OR014	400	8.333%	OR fault level = 0.014
13	OR021	400	8.333%	OR fault level = 0.021
14	RB007	400	8.333%	RB fault level = 0.007
15	RB014	400	8.333%	RB fault level = 0.014
16	RB021_0	100	2.083%	RB fault level = 0.021 load = 0
17	RB021_1	100	2.083%	RB fault level = 0.021 load = 1
18	RB021_2	100	2.083%	RB fault level = 0.021 load = 2
19	RB021_3	100	2.083%	RB fault level = 0.021 load = 3
20	IR028	400	8.333%	IR fault level = 0.028
21	RB028	400	8.333%	RB fault level = 0.028

In the experiments, we used the Sklearn [63], which is a Python package to realize the proposed classification algorithm. The statistical features including time-domain and frequency features and wavelet packet features are the typically used inputs of classifiers [50,64]; therefore, we compare the performance of the proposed VG features and 13 statistical features in this section.

The 13 statistical features used in this paper include root mean square (RMS), square root of the amplitude (SRA), kurtosis value (KV), skewness value (SV), peak-to-peak value (PPV), crest factor (CF), impulse factor (IF), margin factor (MF), shape factor (SF), kurtosis factor (KF), frequency center (FC), RMS frequency (RMSF) and root variance frequency (RVF). The procedure proposed in [11] is used to produce the wavelet packet features. The mother wavelet is Daubechies 4, and refining is done down to the fourth decomposition level. Sixteen (2^4) wavelet features are obtained finally. The $n-$th, $(n = 1, 2, ..., 16)$ wavelet package feature is the respective percentages of the energy of each leaf.

5.2. VGAM Construction

First, subjecting the bearing fault data to the visibility approach defined in Section 3, the VGAM, which may represent the characteristics of vibration signals, is obtained finally. The examples of VGAM converted from vibration signals are shown in Figure 5. It is clear that the temporal distribution of 0-1 (black-white) varies for different fault types. However, the VGAM is so complicated that it is necessary to extract features to describe the VGAM. The VG features for representative machine conditions in the section are calculated. Fifteen VG features are obtained finally.
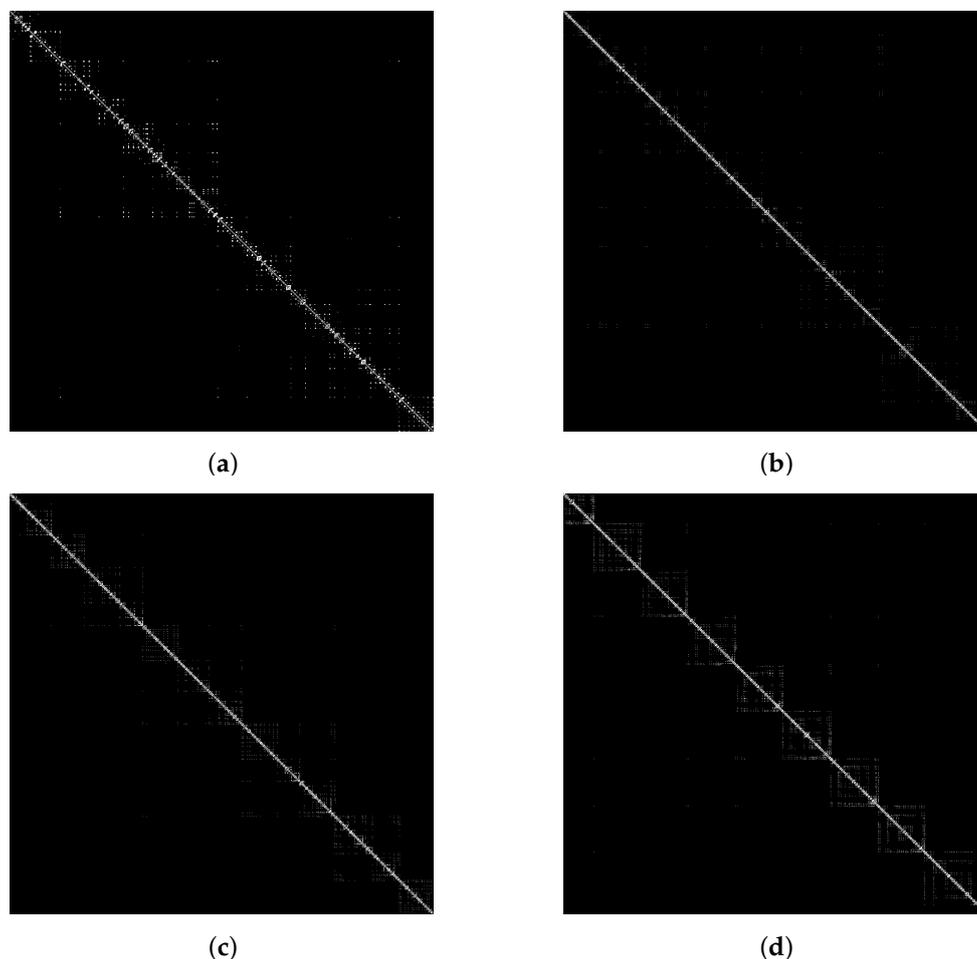

(a)

(b)

(c)

(d)

Figure 5. VGAM from vibration signals. **(a)** NM; **(b)** RB; **(c)** IR; **(d)** OR.

5.3. Results and Discussion

In this section, the optimal feature subset is solved based on the proposed VG feature selection algorithm described in Section 4.2. The linear SVM (LSVM), 1-NN and MLP were used in the experiments, and the number of hidden layers was equal to the number of features for MLP.

All classifiers always used 10-fold cross-validation to estimate the diagnosis ACC. Figures 6 and 7 show the diagnosis ACC of the three classifiers during the process of feature selection using IMS and CWRU data, respectively. Based on the obtained ACC with different number of VG features, the optimal number of features can be produced.
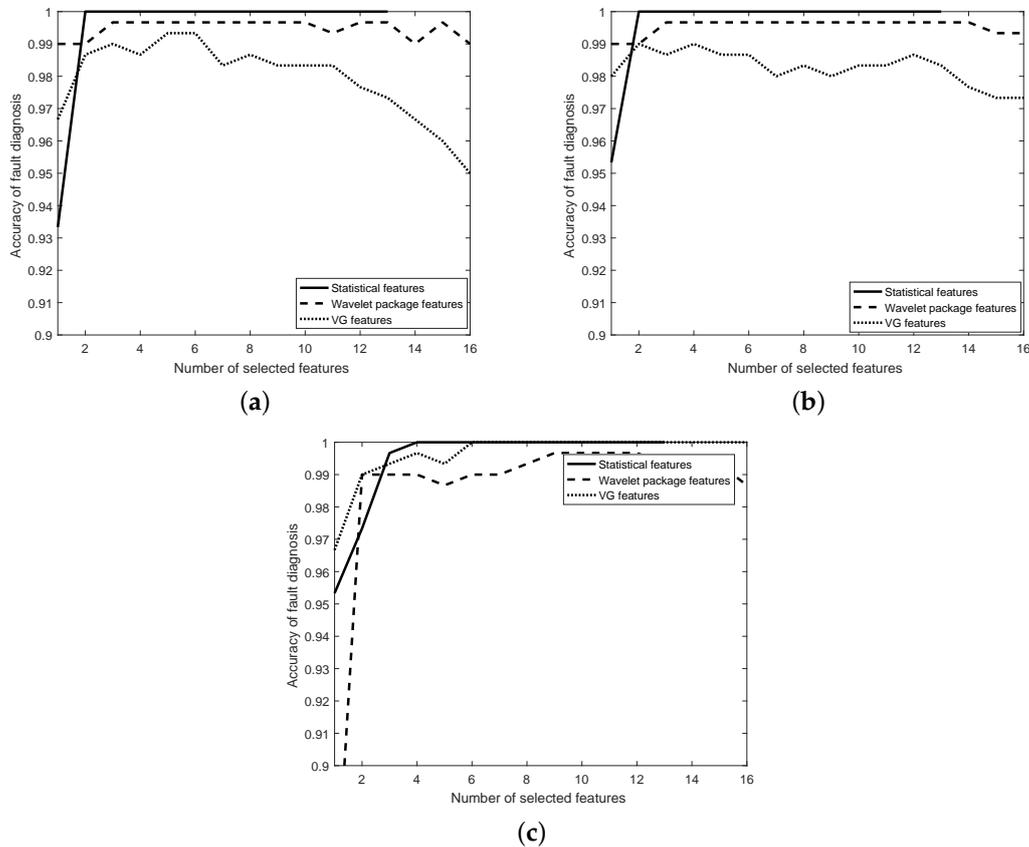

Figure 6. Estimated accuracy during feature selection for all features using IMS data. (**a**) KNN; (**b**) LSVM; (**c**) MLP.

With respect to the experiments using the IMS dataset, the ACC of the 1-NN and LSVM classifiers is depicted in Figure 6a,b. The statistical feature model performed better than the other two feature models. However, the proposed VG feature model had lower ACC than the other two feature models. The ACC (99.33% and 99%) was also acceptable for bearing fault diagnosis. The ACC of the MLP classifier is depicted in Figure 6c. The proposed VG feature model performed better than the wavelet package feature model because the ACC of fault diagnosis was 100%.

With respect to the the experiments using the CWRU dataset, the ACC of the 1-NN classifier is depicted in Figure 7a. The statistical feature model performed better than the other two feature models when the number of selected features was less or equal to two. However, the ACC was unacceptable for bearing fault diagnosis. The proposed VG feature model performed better than the other two feature models when the number of selected features was greater than two. The maximum ACC of 1-NN using the VG feature model (98.0%) was higher than the statistical features (83.3%) and the wavelet package features (94.5%). The corresponding optimal number of features was 11, 9 and 10 for the VG, statistical and wavelet package feature model.

The ACC of the LSVM classifier is depicted in Figure 7b. The proposed VG feature model performed better than the other two feature models when the number of selected features ranged from 1–15. The maximum ACC of LSVM using the VG feature model (98.6%) was also higher than

the statistical features (85.4%) and the wavelet package features (96.6%). The corresponding optimal number of features was 15, 13 and 15 for the VG, statistical and wavelet package feature model.

The ACC of the MLP classifier is depicted in Figure 7c. The proposed VG feature model performed better than the other two feature models when the number of selected features ranged from 1–15. The maximum ACC of LSVM using the VG feature model (98.9%) was also higher than the statistical feature model (86.7%) and the wavelet package feature model (96.4%). The corresponding optimal number of features was 11, 9 and 14 for the VG, statistical and wavelet package feature model.
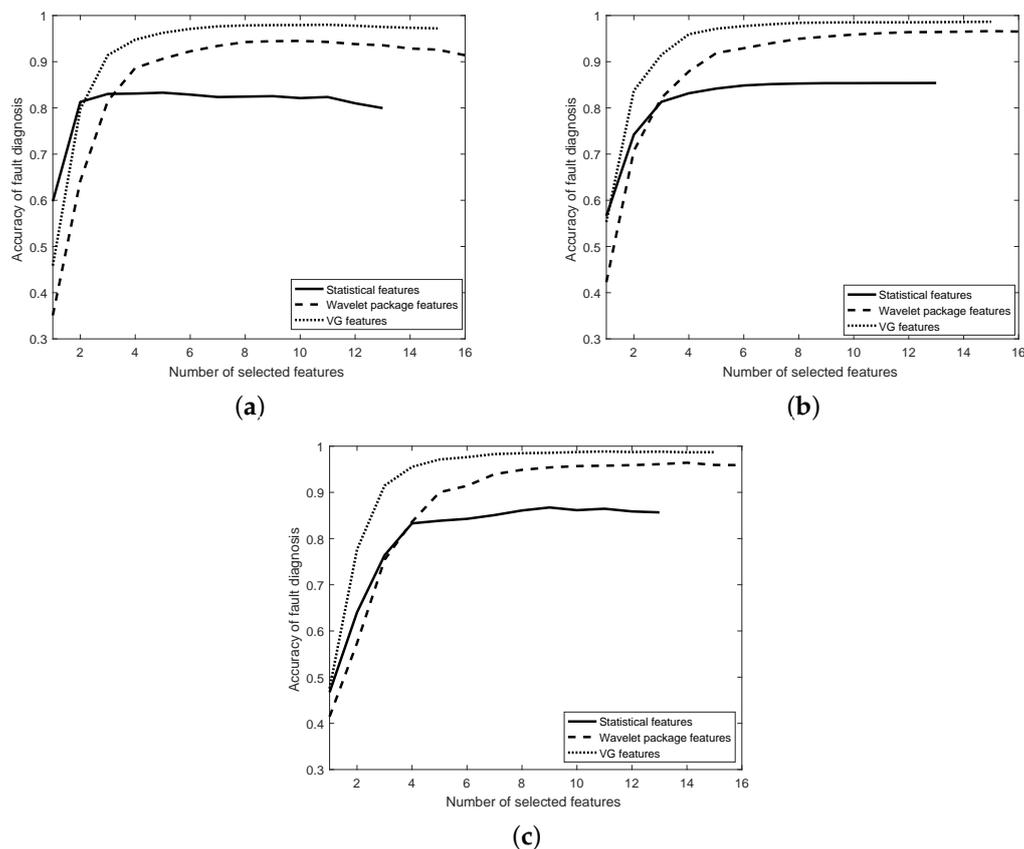

Figure 7. Estimated accuracy during feature selection for all features using the CWRU data. (**a**) KNN; (**b**) LSVM; (**c**) MLP.

All in all, the optimal VG feature subset was obtained by comparing the performance of three classifiers. The optimal VG feature subset was composed of three global VG features and eight local VG features based on the results of the 1-NN and MLP classifiers. This fact shows that both the global and local VG features should be extracted for bearing fault diagnosis. The diagnosis ACC was so high that the VG features can be considered as the input of classifiers. Because the working condition and the fault level were also considered in the experiments, the VG features can be used to recognize the fault class and the corresponding fault level of rolling bearings under the same or variable working conditions. The three classifiers using the VG feature model performed better than the statistical feature model and wavelet package feature model, which are often used as the input of classifiers. In the three classifiers, the MLP using the VG feature model performed best.

6. Conclusions and Extension

The paper proposed a framework to extract VG features from the vibration signals of rolling bearings, which are used widely in industrial systems. Firstly, the VG approach is used to convert the normalized vibration signals into a binary matrix. It is a successful attempt to analyze vibration signals even though the VG approach has been used to analyze data in other areas. The approximate

homogeneity between the VG of signals and the signal envelope is analyzed based on Jensen's inequality. The VG feature pool, which is comprised of 15 VG features, has been extracted to describe the VGAM structure.

Secondly, the sequential feature selection (SFS) method is applied to select the optimal VG feature subset. The classifiers including LSVM, 1-NN and MLP are applied to validate the VG feature model based on the CWRU data. The results show that the VG feature model can be used to analyze the vibration signals because the selected VG features can be taken as the input of the classifiers in order to recognize the fault precisely. The VG feature model can be used to recognize the fault class of rolling bearings under the variable working conditions, which means that our model can predict the localization of the fault and its severity without knowing the load and rotation speed clearly.

Finally, the advantage of the VG feature model is verified by comparison with the statistical feature model and the wavelet package feature model, and the VG feature model performs better than the other two feature models. While the results generated using the proposed method were based on limited test data, they offer useful insights into fault diagnosis. The VG features are novel and can be selected by engineers to detect bearing faults.

In the future study, we will validate the VG features by using vibration signals of other rolling bearings. The proposed method may also serve as an analytical method to detect the faults of other machines such as gear boxes using the combination of the VG feature model and other features in the future. Considering the real-time condition monitoring of rolling bearings, a resampling method of vibration signals will be proposed to reduce the computing complexity of the visibility graph in our future works.

Author Contributions: Conceptualization, Z.Z. and L.J.; methodology, Y.Q.; validation, X.C.

Conflicts of Interest: The authors declared that they have no conflicts of interest in this work.

References

1. Cheng, Y.; Zhou, B.; Lu, C.; Yang, C. Fault Diagnosis for Rolling Bearings under Variable Conditions Based on Visual Cognition. *Materials* **2017**, *10*, 582. [CrossRef] [PubMed]
2. López de Lacalle, L.N.; Rodriguez, A.; Lamikiz, A.; Celaya, A.; Alberdi, R. Five-axis machining and burnishing of complex parts for the improvement of surface roughness. *Mater. Manuf. Process.* **2011**, *26*, 997–1003. [CrossRef]
3. Fernández-Abia, A.I.; Barreiro, J.; de Lacalle, L.N.L.; Martínez-Pellitero, S. Behavior of austenitic stainless steels at high speed turning using specific force coefficients. *Int. J. Adv. Manuf. Technol.* **2012**, *62*, 505–515. [CrossRef]
4. Boudiaf, A.; Moussaoui, A.; Dahane, A.; Atoui, I. A comparative study of various methods of bearing faults diagnosis using the case Western Reserve University data. *J. Fail. Anal. Prev.* **2016**, *16*, 271–284. [CrossRef]
5. Ghafari, S.; Golnaraghi, F.; Ismail, F. Effect of localized faults on chaotic vibration of rolling element bearings. *Nonlinear Dyn.* **2008**, *53*, 287–301. [CrossRef]
6. Saruhan, H.; Saridemir, S.; Qicek, A.; Uygur, I. Vibration analysis of rolling element bearings defects. *J. Appl. Res. Technol.* **2014**, *12*, 384–395. [CrossRef]
7. Rai, A.; Upadhyay, S. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [CrossRef]
8. Mahamad, A.K.; Hiyama, T. Fault classification based artificial intelligent methods of induction motor bearing. *Int. J. Innov. Comput. Inf. Control* **2011**, *7*, 5477–5494.
9. Wu, S.D.; Wu, C.W.; Wu, T.Y.; Wang, C.C. Multi-scale analysis based ball bearing defect diagnostics using Mahalanobis distance and support vector machine. *Entropy* **2013**, *15*, 416–433. [CrossRef]
10. Peter, W.T.; Peng, Y.; Yam, R. Wavelet analysis and envelope detection for rolling element bearing fault diagnosis-their effectiveness and flexibilities. *J. Vib. Acoust.* **2001**, *123*, 303–310.

11. Eren, L.; Devaney, M.J. Bearing damage detection via wavelet packet decomposition of the stator current. *IEEE Trans. Instrum. Meas.* **2004**, *53*, 431–436. [CrossRef]

12. Wang, Y.; Xu, G.; Liang, L.; Jiang, K. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **2015**, *54*, 259–276. [CrossRef]

13. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [CrossRef]

14. Amar, M.; Gondal, I.; Wilson, C. Vibration Spectrum Imaging: A Novel Bearing Fault Classification Approach. *IEEE Trans. Ind. Electron.* **2015**, *62*, 494–502. [CrossRef]

15. Sugumaran, V.; Ramachandran, K. Effect of number of features on classification of roller bearing faults using SVM and PSVM. *Expert Syst. Appl.* **2011**, *38*, 4088–4096. [CrossRef]

16. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.

17. Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.* **2004**, *5*, 1205–1224.

18. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]

19. Lazar, C.; Taminau, J.; Meganck, S.; Steenhoff, D.; Coletta, A.; Molter, C.; Schaetzen, V.D.; Duque, R.; Bersini, H.; Nowe, A. A Survey on Filter Techniques for Feature Selection in Gene Expression Microarray Analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2012**, *9*, 1106–1119. [CrossRef] [PubMed]

20. Urbikain, G.; Alvarez, A.; López de Lacalle, L.N.; Arsuaga, M.; Alonso, M.A.; Veiga, F. A reliable turning process by the early use of a deep simulation model at several manufacturing stages. *Machines* **2017**, *5*, 15. [CrossRef]

21. Bustillo, A.; Urbikain, G.; Perez, J.M.; Pereira, O.M.; de Lacalle, L.N.L. Smart optimization of a friction-drilling process based on boosting ensembles. *J. Manuf. Syst.* **2018**, *48*, 108–121. [CrossRef]

22. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [CrossRef]

23. Samanta, B.; Al-Balushi, K. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Signal Process.* **2003**, *17*, 317–328. [CrossRef]

24. Sugumaran, V.; Muralidharan, V.; Ramachandran, K. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mech. Syst. Signal Process.* **2007**, *21*, 930–942. [CrossRef]

25. Yang, Y.; Yu, D.; Cheng, J. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement* **2007**, *40*, 943–950. [CrossRef]

26. Abbasion, S.; Rafsanjani, A.; Farshidianfar, A.; Irani, N. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mech. Syst. Signal Process.* **2007**, *21*, 2933–2945. [CrossRef]

27. Djemili, R.; Bourouba, H.; Korba, M.C.A. Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals. *Biocybern. Biomed. Eng.* **2016**, *36*, 285–291. [CrossRef]

28. Lacasa, L.; Luque, B.; Ballesteros, F.; Luque, J.; Nuno, J.C. From time series to complex networks: The visibility graph. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4972–4975. [CrossRef] [PubMed]

29. Ni, X.H.; Jiang, Z.Q.; Zhou, W.X. Degree distributions of the visibility graphs mapped from fractional Brownian motions and multifractal random walks. *Phys. Lett. A* **2009**, *373*, 3822–3826. [CrossRef]

30. Turner, A.; Doxa, M.; O'Sullivan, D.; Penn, A. From isovists to visibility graphs: A methodology for the analysis of architectural space. *Environ. Plan. B Plan. Des.* **2001**, *28*, 103–121. [CrossRef]

31. Luque, B.; Lacasa, L.; Ballesteros, F.; Luque, J. Horizontal visibility graphs: Exact results for random time series. *Phys. Rev. E* **2009**, *80*, 046103. [CrossRef] [PubMed]

32. Zhou, T.T.; Jin, N.D.; Gao, Z.K.; Luo, Y.B. Limited penetrable visibility graph for establishing complex network from time series. *Wuli Xuebao* **2012**, *61*, 355–367.

33. Gao, Z.K.; Cai, Q.; Yang, Y.X.; Dang, W.D.; Zhang, S.S. Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series. *Sci. Rep.* **2016**, *6*, 35622. [CrossRef] [PubMed]

34. Bezsudnov, I.; Snarskii, A. From the time series to the complex networks: The parametric natural visibility graph. *Phys. A Stat. Mech. Appl.* **2014**, *414*, 53–60. [CrossRef]

35. Zhang, B.; Wang, J.; Fang, W. Volatility behavior of visibility graph EMD financial time series from Ising interacting system. *Phys. A Stat. Mech. Appl.* **2015**, *432*, 301–314. [CrossRef]
36. Rong, L.; Shang, P. Topological entropy and geometric entropy and their application to the horizontal visibility graph for financial time series. *Nonlinear Dyn.* **2018**, *92*, 41–58. [CrossRef]
37. Braga, A.; Alves, L.; Costa, L.; Ribeiro, A.; de Jesus, M.; Tateishi, A.; Ribeiro, H. Characterization of river flow fluctuations via horizontal visibility graphs. *Phys. A Stat. Mech. Appl.* **2016**, *444*, 1003–1011. [CrossRef]
38. Ahmadlou, M.; Adeli, H.; Adeli, A. New diagnostic EEG markers of the Alzheimer disease using visibility graph. *J. Neural Transm.* **2010**, *117*, 1099–1109. [CrossRef] [PubMed]
39. Gao, Z.K.; Cai, Q.; Yang, Y.X.; Dong, N.; Zhang, S.S. Visibility graph from adaptive optimal kernel time-frequency representation for classification of epileptiform EEG. *Int. J. Neural Syst.* **2017**, *27*, 1750005. [CrossRef] [PubMed]
40. Zeng, M.; Ma, W.; Meng, Q.; Sun, B.; Wu, Z.; Lu, J. Noise resistance ability analysis of the visibility graph and the limited penetrable visibility graph. In Proceedings of the 2016 12th World Congress on Intelligent Control and Automation (WCICA), Guilin, China, 12–15 June 2016; pp. 2648–2653.
41. Derrible, S.; Kennedy, C. The complexity and robustness of metro networks. *Phys. A Stat. Mech. Appl.* **2010**, *389*, 3678–3691. [CrossRef]
42. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **2010**, *52*, 1059–1069. [CrossRef] [PubMed]
43. Opsahl, T.; Agneessens, F.; Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Netw.* **2010**, *32*, 245–251. [CrossRef]
44. Dorogovtsev, S.N.; Mendes, J.F.; Samukhin, A.N. Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E* **2001**, *63*, 062101. [CrossRef] [PubMed]
45. Fronczak, A.; Fronczak, P.; Hołyst, J.A. Average path length in random networks. *Phys. Rev. E* **2004**, *70*, 056110. [CrossRef] [PubMed]
46. Peleg, D.; Roditty, L.; Tal, E. Distributed algorithms for network diameter and girth. In Proceedings of the International Colloquium on Automata, Languages, and Programming, Warwick, UK, 9–13 July 2012; pp. 660–672.
47. Soffer, S.N.; Vazquez, A. Network clustering coefficient without degree-correlation biases. *Phys. Rev. E* **2005**, *71*, 057101. [CrossRef] [PubMed]
48. Moosavian, A.; Ahmadi, H.; Tabatabaeefar, A.; Khazaee, M. Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing. *Shock Vib.* **2013**, *20*, 263–272. [CrossRef]
49. Lei, Y.; Zuo, M.J. Gear crack level identification based on weighted K nearest neighbor classification algorithm. *Mech. Syst. Signal Process.* **2009**, *23*, 1535–1547. [CrossRef]
50. Rauber, T.W.; de Assis Boldt, F.; Varejão, F.M. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2015**, *62*, 637–646. [CrossRef]
51. Needham, T. A visual explanation of Jensen's inequality. *Am. Math. Mon.* **1993**, *100*, 768–771. [CrossRef]
52. Porter, R.; Canagarajah, N. Robust rotation-invariant texture classification: Wavelet, Gabor filter and GMRF based schemes. *IEE Proc. Vis. Image Signal Process.* **1997**, *144*, 180–188. [CrossRef]
53. Intanagonwiwat, C.; Estrin, D.; Govindan, R.; Heidemann, J. Impact of network density on data aggregation in wireless sensor networks. In Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops, Vienna, Austria, 2–5 July 2002; pp. 457–458.
54. Strogatz, S.H. Exploring complex networks. *Nature* **2001**, *410*, 268. [CrossRef] [PubMed]
55. Kim, J.; Wilhelm, T. What is a complex graph? *Phys. A Stat. Mech. Appl.* **2008**, *387*, 2637–2652. [CrossRef]
56. Cvetković, D.M.; Doob, M.; Sachs, H. *Spectra of Graphs: Theory and Application*; Academic Press: New York, NY, USA, 1980; Volume 87.
57. Yang, J.; Yang, J.Y.; Zhang, D.; Lu, J.F. Feature fusion: Parallel strategy vs. serial strategy. *Pattern Recognit.* **2003**, *36*, 1369–1381. [CrossRef]
58. Dharmagunawardhana, C.; Mahmoodi, S.; Bennett, M.; Niranjan, M. Gaussian Markov random field based improved texture descriptor for image segmentation. *Image Vis. Comput.* **2014**, *32*, 884–895. [CrossRef]
59. Pudil, P.; Novovičová, J.; Kittler, J. Floating search methods in feature selection. *Pattern Recognit. Lett.* **1994**, *15*, 1119–1125. [CrossRef]

60. Zhang, L.; Xiong, G.; Liu, H.; Zou, H.; Guo, W. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Syst. Appl.* **2010**, *37*, 6077–6085. [CrossRef]

61. Hong, H.; Liang, M. Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform. *J. Sound Vib.* **2009**, *320*, 452–468. [CrossRef]

62. Ali, J.B.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **2015**, *89*, 16–27.

63. Haroon, D. Classification. In *Python Machine Learning Case Studies*; Springer:Berlin/Heidelberg, Germany, 2017; pp. 161–196.

64. Xia, Z.; Xia, S.; Ling, W.; Cai, S. Spectral Regression Based Fault Feature Extraction for Bearing Accelerometer Sensor Signals. *Sensors* **2012**, *12*, 13694–13719. [CrossRef] [PubMed]