electronics

MDPI

Article

Variational Mode Decomposition-Based Processing for Detection of Short-Circuited Turns in Transformers Using Vibration Signals and Machine Learning

David Camarena-Martinez [1], Jose R. Huerta-Rosales [2], Juan P. Amezquita-Sanchez [2], David Granados-Lieberman [3], Juan C. Olivares-Galvan [4] and Martin Valtierra-Rodriguez [2,*]

1 ENAP-RG, División de Ingeniería, Campus Irapuato-Salamanca, Universidad de Guanajuato (UG), Carretera Salamanca-Valle de Santiago km 3.5 + 1.8 km, Comunidad de Palo Blanco, Salamanca CP 36885, Mexico; david.camarena@enap-rg.org

2 ENAP-RG, CA-Sistemas Dinámicos y Control, Facultad de Ingeniería, Campus San Juan del Río, Universidad Autónoma de Querétaro (UAQ), Rio Moctezuma 249, San Cayetano, San Juan del Río CP 76807, Mexico; roberto.huerta@enap-rg.org (J.R.H.-R.); juan.amezquita@enap-rg.org (J.P.A.-S.)

3 ENAP-RG, CA-Fuentes Alternas y Calidad de la Energía Eléctrica, Departamento de Ingeniería Electromecánica, Tecnológico Nacional de México/ITS de Irapuato, Carretera Irapuato-Silao km 12.5, Colonia El Copal, Irapuato CP 36821, Mexico; david.granados@enap-rg.org

4 Departamento de Energía, Universidad Autónoma Metropolitana, Ciudad de México CP 02128, Mexico; jolivares@azc.uam.mx

* Correspondence: martin.valtierra@enap-rg.org

Abstract: Transformers are key elements in electrical systems. Although they are robust machines, different faults can appear due to their inherent operating conditions, e.g., the presence of different electrical and mechanical stresses. Among the different elements that compound a transformer, the winding is one of the most vulnerable parts, where the damage of turn-to-turn short circuits is one of the most studied faults since low-level damage (i.e., a low number of short-circuited turns—SCTs) can lead to the overall fault of the transformer; therefore, early fault detection has become a fundamental task. In this regard, this paper presents a machine learning-based method to diagnose SCTs in the transformer windings by using their vibrational response. In general, the vibration signals are firstly decomposed by means of the variational mode decomposition method, where a comparison with the empirical mode decomposition (EMD) method and the ensemble empirical mode decomposition (EEMD) method is also carried out. Then, entropy, energy, and kurtosis indices are obtained from each decomposition as fault indicators, where both the combination of features and the dimensionality reduction by using the principal component analysis (PCA) method are analyzed for the global effectiveness improvement and the computational burden reduction. Finally, a pattern recognition algorithm based on artificial neural networks (ANNs) is used for automatic fault detection. The obtained results show 100% effectiveness in detecting seven fault conditions, i.e., 0 (healthy), 5, 10, 15, 20, 25, and 30 SCTs.

Keywords: machine learning; short-circuited turns; interturn transformer faults; variational mode decomposition; vibration signals

check for updates

Citation: Camarena-Martinez, D.; Huerta-Rosales, J.R.; Amezquita-Sanchez, J.P.; Granados-Lieberman, D.; Olivares-Galvan, J.C.; Valtierra-Rodriguez, M. Variational Mode Decomposition-Based Processing for Detection of Short-Circuited Turns in Transformers Using Vibration Signals and Machine Learning. *Electronics* **2024**, *13*, 1215. https://doi.org/10.3390/electronics13071215

Academic Editors: Azeddine Kaddouri and Nouha Bouchiba

Received: 22 February 2024
Revised: 18 March 2024
Accepted: 21 March 2024
Published: 26 March 2024

Copyright: © 2024 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

Transformers are fundamental elements for the proper operation of power systems; in this regard, the presence of internal or external faults causes unexpected outages and large repair costs. Therefore, fault diagnosis of transformers is a very important task. During their service life, transformers could present various faults at their different components, such as windings, core, tap changer, bushings, and so on; however, one of the most common faults is the short-circuited turns (SCTs), which occurs in the transformer windings [1]. The SCT fault is generated by diverse factors such as insulation aging, electrical and mechanical

stresses, and atmospheric discharges, among others [2]. In addition, a slight SCT fault can rapidly lead to a more severe fault condition because of the sudden increase in current and temperature. Therefore, tool development is a priority task to help the personnel in charge of preventive maintenance to design the best maintenance strategies for the perfect transformer performance [3].

In recent years, diverse strategies for SCT detection have been presented in the literature [4], and are the most employed the partial discharge method, oil analysis, and frequency-response analysis; however, these methods present diverse disadvantages such as the use of special equipment in situ or laboratory tests, elevated costs, and longer analysis times, among others. On the contrary, the use of vibration signals in recent years has proved to be a non-destructive and economical method to detect faults in transformers [5,6]. However, the application of vibration analysis techniques and the selection of their evaluation criteria for the detection of transformer fault conditions are not straightforward tasks, mainly if the performance (i.e., accuracy in the diagnosis) and the complexity (i.e., computational burden) are considered. In this sense, several methodologies have been implemented for both the vibration signal analysis and the SCTs diagnosis. For instance, a set of fractal dimension algorithms (Katz, Higuchi, Box, and Sevcik) as potential fault indicators along with three classifiers (decision tree-, naïve Bayes-, and k-nearest neighbors-based methods) to perform an automatic diagnosis of SCTs is used in [7]. In [8], the analysis of vibro-acoustic signals with the wavelet transform (WT) is proposed to characterize the change of the winding mechanical condition generated by SCTs. In [9], 19 statistical time features with a support vector machine (SVM) are combined to classify several SCT conditions.

Recently, decomposition-based techniques have allowed the decomposition of the analyzed signal in diverse intrinsic mode functions (IMFs) or frequency bands with relevant information in order to be associated with the identification of faults. In this sense, three decomposition methods, i.e., empirical mode decomposition (EMD), ensemble-EMD (EEMD), and complete-EEMD (CEEMD), of analyzing current signals measured in a transformer with SCTs are evaluated in [10]. The results show that those techniques present a reliable potential for transformer diagnosis, opening the possibility of being complemented with different machine learning strategies for automatic diagnosis/classification. In [11], both the WT and the EMD methods for diagnosing internal faults from magnetization conditions in power transformers are presented. Although the decomposition techniques have provided promising results in the diagnosis of transformers, it has some limitations that can compromise their performance, e.g., IMFs can have multiple intrinsic time scales of different frequencies (i.e., mode mixing) or false modes, as well as their susceptibility to noise. To overcome some of these limitations, the variational mode decomposition (VMD) has been presented in the literature [12,13]. In general, it is characterized by adaptively decomposing the data in non-recursive intrinsic mode functions; hence, its estimated variational modes provide continuity, correlation, and noise robustness [14]. In this regard, some applications for the use of VMD for fault detection in transformers have been presented in the literature. For instance, the VMD method to detect the current transformer saturation, with light, moderate, severe and alternating current (AC) saturation conditions, is presented in [15]. In [16], a method for winding degradation assessment based on the combination of VMD and WT, using the vibration signals from the three winding of the transformer, is developed. Also, the use of VMD with an SVM and current signals to identify transformer magnetizing inrush current and fault current in power systems is presented in [17]. Finally, a method based on VMD, Choi–Williams distribution, and convolutional neural networks to detect partial discharges in power transformers is also developed [18]. In the abovementioned works, the VMD method has demonstrated to be a reliable signal processing technique to develop practical methodologies that provide an accurate diagnosis about the transformer state; in this regard, the development and application of VMD-based methodologies have to be explored in other faults, e.g., SCTs, and applied to other signals, e.g., 3-axis vibrations, in order to provide more efficient fault detection methods in terms of accuracy and complexity.

In this work, a methodology that consists of the VMD method as a signal processing technique, statistical indices as fault indicators, and artificial neural networks as pattern recognition algorithms for the diagnosis of a transformer with different SCT severities is presented. The experimental analysis is performed in a transformer capable of emulating seven SCT fault severities, i.e., 0 (healthy), 5, 10, 15, 20, 25, and 30 SCTs. At first, the acquired vibration signals are decomposed by using the VMD method, where the EMD and EEMD methods are also computed for comparison purposes. Then, entropy, energy, and kurtosis indices are employed as fault indicators, which have provided promising results in other applications [19]. The features are estimated from each IMF (36 in total, 12 for each indicator), where different feature combinations and dimensionality reduction by using the principal component analysis (PCA) method are analyzed with the purpose of improving accuracy and computational time. Finally, with the best set of input features, an artificial neural network (ANN), as a low-complex pattern recognition algorithm, is employed for automatic diagnosis.

2. Theoretical Background

This section presents a brief description of the methods and algorithms employed in this work.

2.1. Time-Frequency Methods

2.1.1. Variational Mode Decomposition

Proposed in [12], the VMD is characterized by being a new adaptive method with the capability of decomposing a time signal, $g(t)$, in various narrow frequency bands known as variational IMFs, u_k, calculated around a central frequency ω_k. In general, the resulting constrained variational problem is given by [12]:

$$\begin{aligned} &\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \left\| d(t) \left[\left(\delta(t) + \tfrac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &\text{Subjected to } \sum_{k=1}^{K} u_k(t) = g(t) \end{aligned} \tag{1}$$

where $\delta(t)$ is a unit impulse function, $\|\bullet\|_2^2$ represents the gradient L^2-norm of shifted signal, * is the convolutional operator, and $j^2 = -1$.

To solve the optimization problem, the abovementioned problem is changed into an unconstrained problem by including a quadratic penalty term and a Lagrangian multiplier λ as follows:

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) := \alpha \sum_{k=1}^{K} &\left\| \partial(t) \left[\left(\delta(t) + \tfrac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| g(t) - \sum_k u_k(t) \right\|_2^2 \\ &+ \left\langle \lambda(t), g(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{2}$$

where α is a parameter used to constraint the data fidelity. The solution to Equation (2) is now considered as the saddle point of the Lagrangian by updating $u_k^{n+1}, \omega_k^{n+1}$, and λ_k^{n+1} in an iterative sub-optimization called the alternating direction method of multipliers.

Figure 1 depicts the VMD algorithm procedures. A detailed explanation of the VMD method can be found in [12].

Although the VMD method has reported promising results in different areas, it is important to compare its performance with similar signal decomposition methods, such as EMD-based methods, in order to analyze and compare its performance for a specific application.

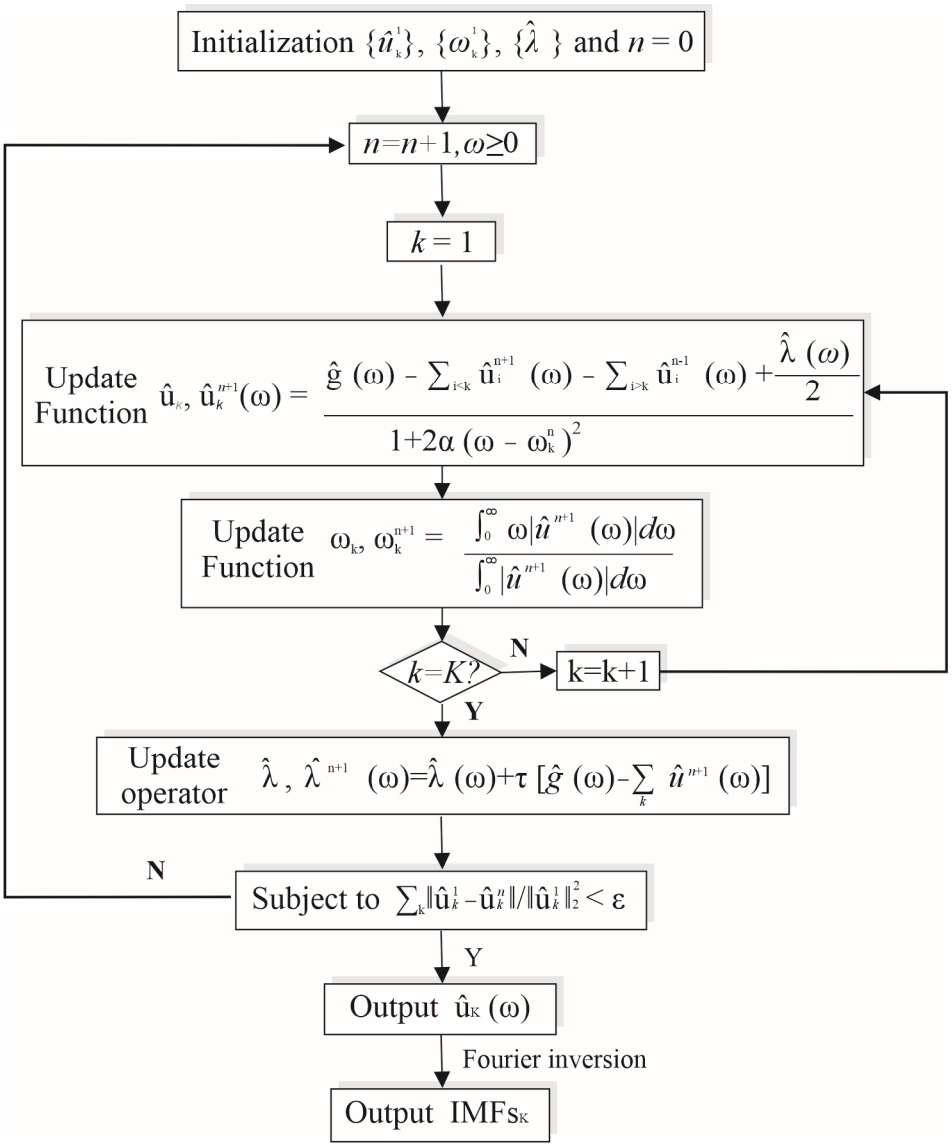


Figure 1. Flow chart for the VMD algorithm.

2.1.2. Empirical Mode Decomposition-Based Methods

EMD is an adaptive technique that is developed to deal with non-linear and non-stationary signals. It is characterized by decomposing the original time signal in a set of new signals called IMFs which must satisfy two criteria to be considered as an IMF: (1) the number of extremes and zero crossings for the entire time signal must be the same or differ by a maximum of one and (2) the estimated envelopes of a time signal, upper and lower, must satisfy an average equal to zero. The EMD process is completed when an IMF with a monotonic behavior is obtained.

In recent years, an improved version of EMD called ensemble-EMD (EEMD) has been introduced in the literature to lessen the main problem found into the EMD known as mode mixing (i.e., the presence of oscillations with different amplitudes in one or more IMFs). Basically, it consists of repeating multiple times the EMD process but adding a different white Gaussian noise to the original time signal in each trial. Then, the resulting IMFs are averaged. Figure 2 illustrates the process of both the EMD and EEMD algorithms. A detailed explanation for the EMD methods can be found in [10].

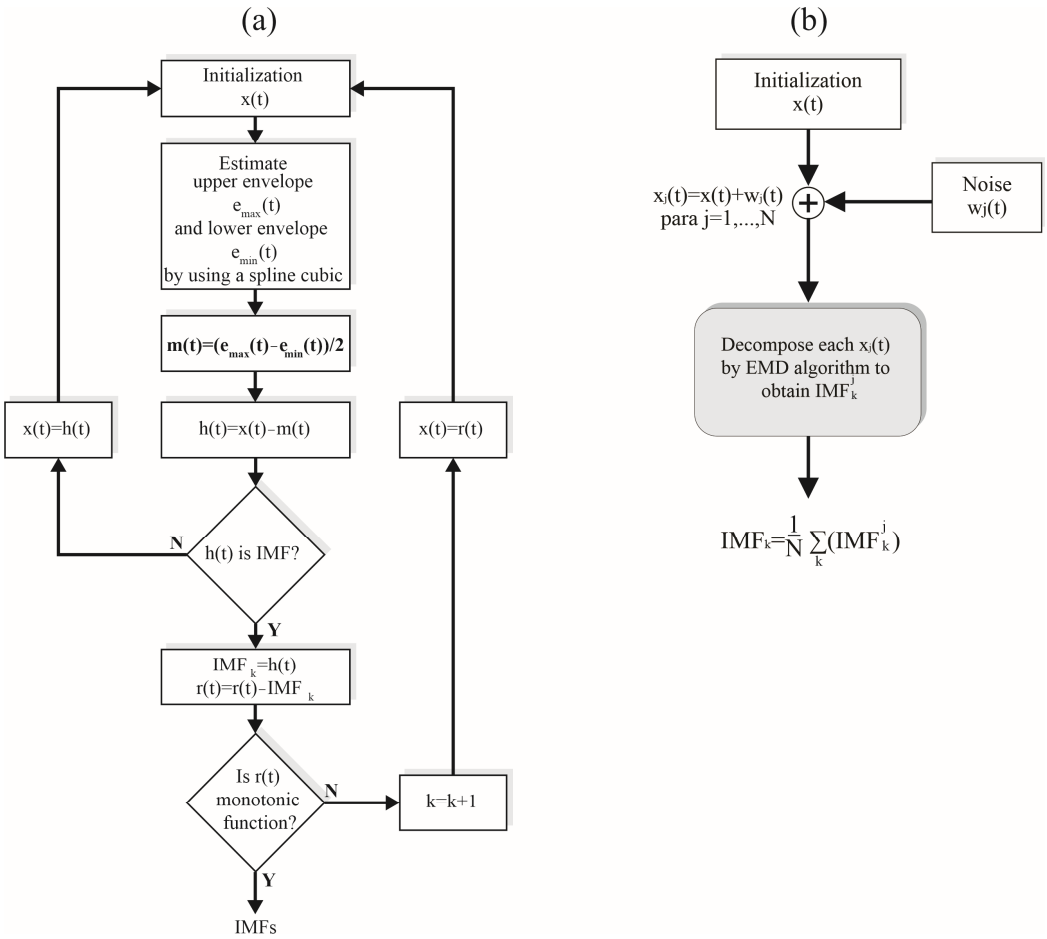


Figure 2. Flow charts for the algorithms: (**a**) EMD and (**b**) EEMD.

2.2. Time Non-Linear Features

An essential step in a diagnosis methodology is the identification of patterns or features in the analyzed signals, which must have the capability of being associated with the phenomena analyzed. In this sense, diverse non-linear features in the time domain, e.g., root mean square, variance, standard deviation, among others, have been employed for detecting consolidated damages [20], demonstrating that they could be useful tools for encountering relevant features into the time signals. Nevertheless, it is important to investigate others features, e.g., energy (ENE), kurtosis (KUR), and Shannon entropy (SEN), in order to evaluate their potential to characterize relevant information into the time signals, which can be associated with the operating condition of the transformers with SCTs.

Equations (3)–(5) mathematically describe the features studied in this work.

$$ENE = \sum_{i=1}^{N} |x_i|^2 \tag{3}$$

$$KUR = \frac{1}{N(STD)^4} \sum_{i=1}^{N} (x_i - \overline{x})^4 \tag{4}$$

$$SEN = -\sum_{i=1}^{K} p(x_i) \log_2[p(x_i)] \tag{5}$$

where x is the analyzed time signal with N samples, $\overline{x}$ represents the mean of the data analyzed, and STD corresponds to the standard deviation of the time signal. Finally, $p(x_i)$ represents the appearance probability of the data samples in the analyzed time signal.

It should be noted that each feature can provide different information; therefore, they must be analyzed and selected to avoid redundant information. In addition, a dimensionality reduction can be applied to reduce the computational board and simplify the complexity of the following stages of a machine learning scheme.

2.3. Principal Component Analysis

PCA is a single-value decomposition-based method for transforming high dimensionality data in lower dimensionality data by means of statistical procedures [21]. To perform this task, the covariance matrix **C** is firstly computed in order to obtain a matrix of size $p \times p$ as follows:

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}^T \cdot \mathbf{X} \tag{6}$$

where n refers to the number of samples, p describes the number of dimensions, and $\mathbf{X}^T$ represents the conjugate transpose of the feature matrix **X**.

Finally, the diagonal matrix with the eigenvalues of **C** is obtained by:

$$\mathbf{T} = \mathbf{V}^{-1}\mathbf{C}\mathbf{V} \tag{7}$$

where **V** contains the eigenvectors, i.e., the principal components of the reference model.

2.4. Multilayer Perceptron

An ANN is a pattern recognition algorithm that it is widely used for automatic fault diagnosis [22]. In recent years, diverse configurations of ANNs have been introduced in the literature, being the Multilayer Perceptron (MLP) one of the most employed ones due to its two main advantages: (1) easy implementation and (2) good performance for associating inputs with outputs [23]. These advantages motivate its use in this work to diagnose the transformer condition. The general structure is an MLP that consists of three layers: input, hidden, and output layer. In this ANN, data, x_i, move from the input layer to the output layer through the hidden layer in one (forward) direction. To minimize the error between the desired and obtained outputs, a backpropagation algorithm is employed for the MLP training.

To summarize the performance of an ANN, a confusion matrix can be used. It presents the number of correct and incorrect predictions according to the true and predicted classes. In addition, to quantify its performance, accuracy, precision, and recall values can be computed as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$precision = \frac{TP}{TP + FP} \tag{9}$$

$$recall = \frac{TP}{TP + FN} \tag{10}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Precision and recall help define the effectiveness of the classification, i.e., precision measures the portion of positive identifications in a classification set that was actually correct, while recall represents the proportion of actual positives that were correctly identified. In a multi-class problem, as in this work, the average values of these parameters for each class are calculated and called average accuracy, macro-precision, and macro-recall. They are calculated as follows:

$$average\ accuracy = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \tag{11}$$

$$precision_M = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FP_i}}{l} \tag{12}$$

$$recall_M = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FN_i}}{l} \tag{13}$$

where l is the number of classes and M indicates macro-averaging.

3. Methodology

Figure 3 shows the proposed methodology to diagnose SCTs in transformers from vibration signals. First, vibration signals in three axes (Vx, Vy, and Vz) are acquired from the transformer. The vibration signals are acquired while the transformer is in the steady state; also, the transformer can emulate seven fault conditions, i.e., 0 (healthy), 5, 10, 15, 20, 25, and 30 SCTs. The number indicates how many turns are short-circuited. It is worth noting that the current signal is also monitored, but it is only used to ensure that the transformer is in a steady state, i.e., when the inrush current has passed. The following step is to apply a decomposition method (VMD, EMD, and EEMD) to each vibration signal in their three axes. Then, in feature extraction, the statistic indices (ENE, KUR, and SEN) are computed from each IMF yielded by the decomposition methods (36 features); consequently, to obtain the best feature or features, combinations with one, two, and three indices are made; then, a dimensional reduction by the PCA algorithm is also carried out. Finally, an ANN is trained with each combination (or a single) of indices to obtain the diagnosis of the transformer. Then, comparing the diverse results, the best decomposition method and the best features are selected to provide the best-suited method in terms of accuracy and complexity to diagnose the transformer under different SCT fault conditions.

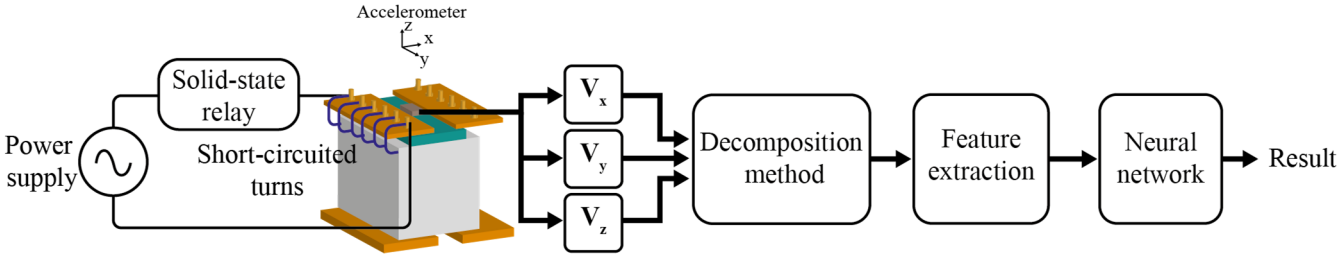


Figure 3. Proposed methodology.

The experimental setup implemented to obtain the vibration signals is shown in Figure 4. The transformer used in the experimentation is a 1.5 kVA single-phase transformer operated at 120 V at 60 Hz. It has 135 turns in its primary winding which is modified to emulate various severities of SCTs in a controlled way, i.e., 0 (healthy), 5, 10, 15, 20, 25, and 30 SCTs. To de-energize the transformer, an autotransformer is used. On the other hand, to obtain the vibration signals, a KISTLER model 8395 A triaxial accelerometer was used. It can measure until ± 10 g with a sensitivity of 400 mV/g. For the data acquisition system (DAS), an NI-USB 6211 board from National Instruments with 16-bit analog-to-digital converters configured with a sampling rate of 6000 samples/s is used. For each condition, 100 tests with a time window of 1 s are carried out. An example of the vibration signals acquired for each test and condition is shown in Figure 5. The overall methodology is implemented on a personal computer (PC) using MATLAB software 2020b.

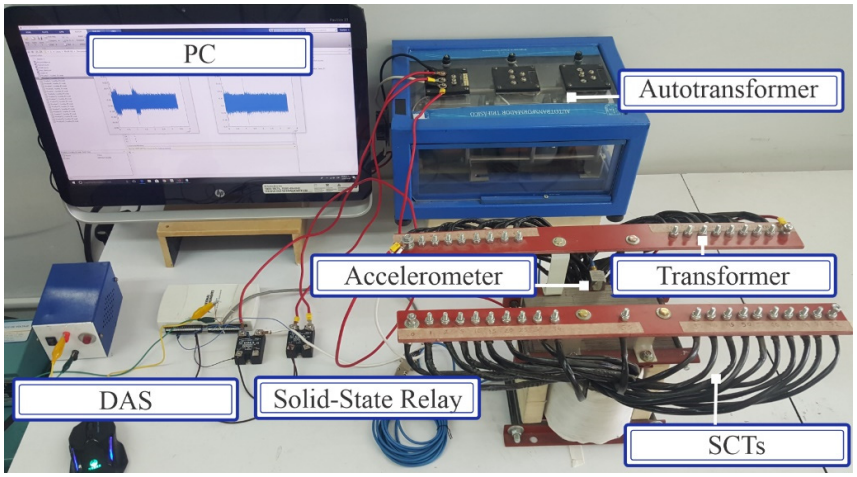


Figure 4. Experimental setup.

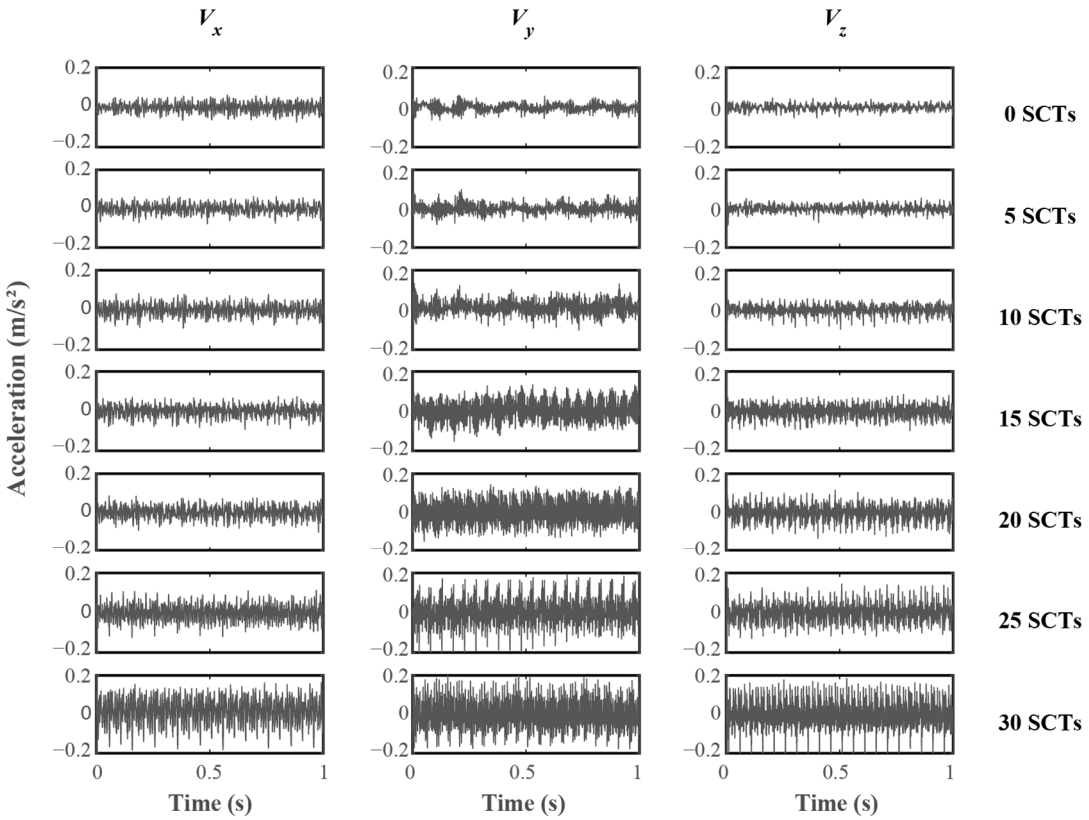


Figure 5. Vibration signals for each SCT condition.

4. Results

In this section, the results of the proposed methodology are presented. Firstly, the decompositions obtained through the VMD method for the cases of 0 and 30 SCTs are shown in Figures 6 and 7, respectively, where a comparison with the decompositions obtained through the EMD and EEMD methods is also presented. The results for other SCT conditions are not provided because of space reasons. Although the behavior of the three decomposition methods seems similar, it is not possible to decide which method is better if quantitative parameters are not taken into account. To identify the best features, 15 indicators (mean, standard deviation, square mean root, range, mode, median, root mean square, skewness, energy, kurtosis, normalized fifth moment, normalized sixth moment, shape factor, Shannon entropy, impulse factor, latitude factor and Shannon

entropy) were computed for each condition. Subsequently, the Kruskal–Wallis test was conducted to ascertain significant differences among three or more independent groups in a variable. Following the test, it was determined that the most relevant features for this study encompassed energy (ENE), kurtosis (KUR), and Shannon entropy (SEN). These features also showed a great potential to extract good information related to the transformer condition, while keeping a low computational burden. Figure 8 shows the boxplot (i.e., mean and standard deviation) of these three selected features for the IMF2 in the y-axis, where it is observed that the feature value increases according to the fault severity, i.e., the indices are sensitive to the fault severity. Despite this high sensitivity, the computed features are overlapped, which negatively affects the straightforward classification; that is, the SCT condition cannot be determined using only one feature, requiring the use of some pattern recognition algorithms.

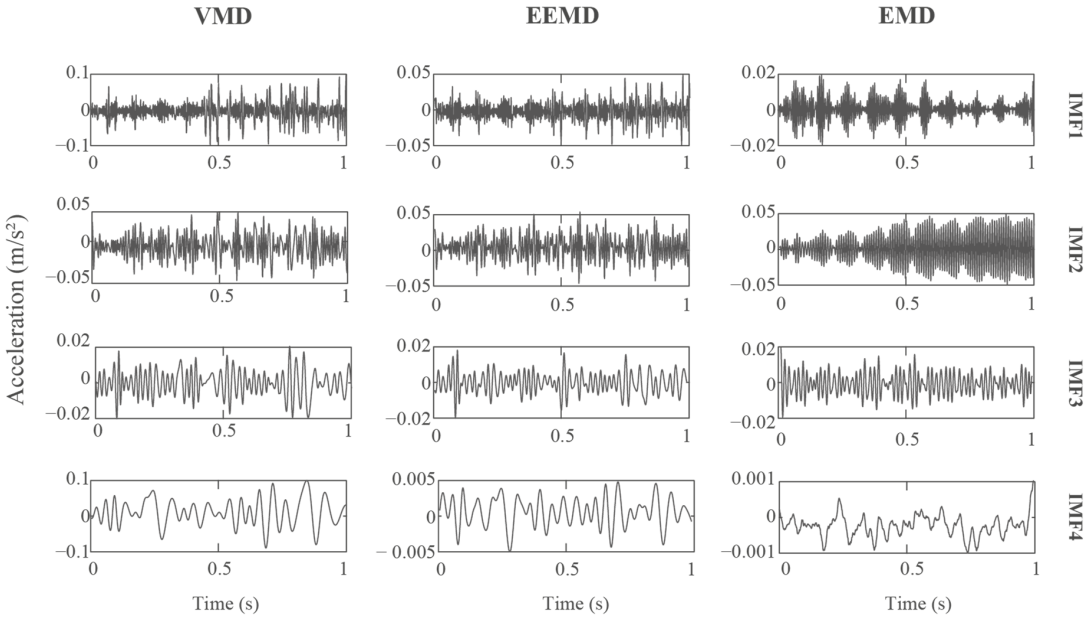


Figure 6. Decompositions of V_x for the condition of 0 SCTs.

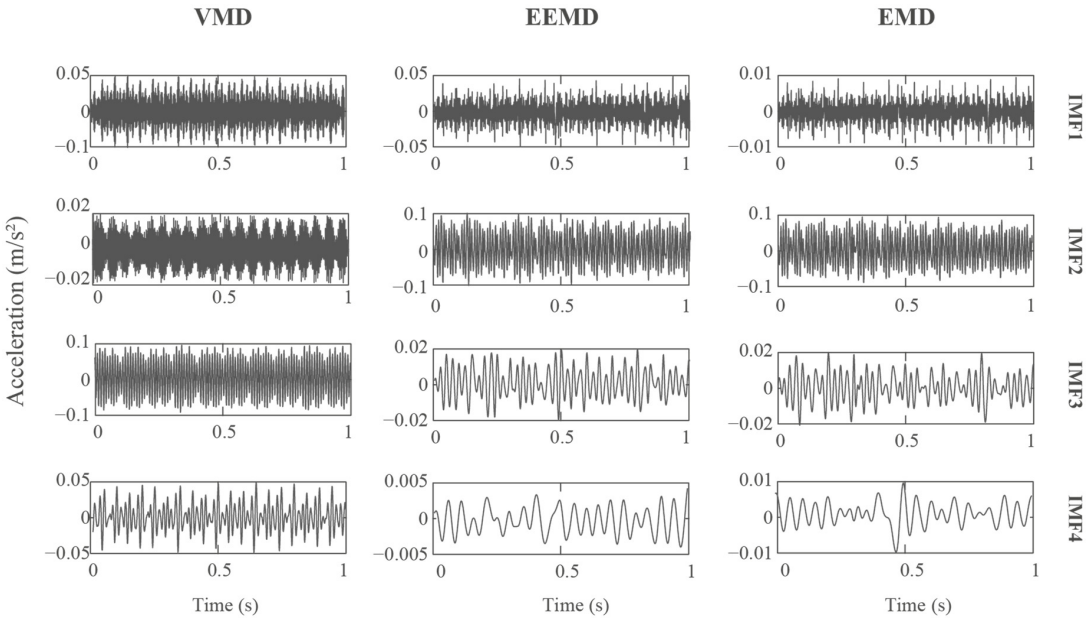


Figure 7. Decompositions of V_x for the condition of 30 SCTs.

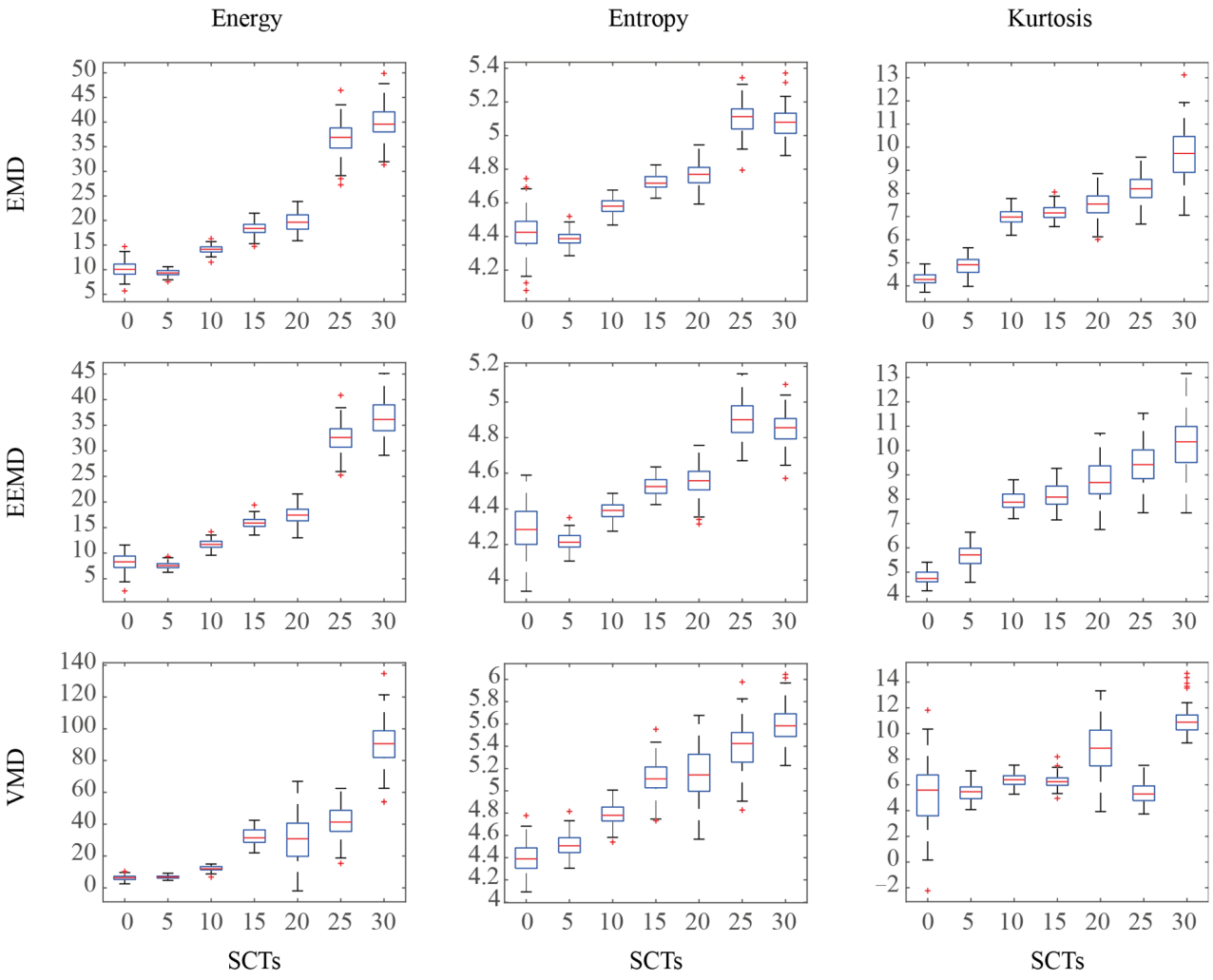


Figure 8. Boxplots for the IMF2 of the *y*-axis vibration signals. Red cross marks indicate outlier values.

The dataset arrangement for developing and testing the proposed methodology is shown in Figure 9. For the seven SCT conditions (i.e., 0, 5, 10, 15, 20, 25, and 30 SCTs), the three methods, i.e., EMD, EEMD, and VMD, are applied for comparison purposes. For each SCT condition and each decomposition method (e.g., see the VMD method for the condition of 5 SCTs in Figure 9), 100 tests are considered. In each test, there are three vibration signals (Vx, Vy, and Vz), where four IMFs are extracted from each signal. For each IMF, three statistical indices (energy, entropy, and kurtosis) are computed. Therefore, in each SCT condition, 3600 data samples (100 tests $\times$ 3 vibration signals $\times$ 4 IMFs $\times$ 3 statistical indices = 3600) are obtained. Thus, each test, as can be observed in Figure 9 with test 1 for the condition of five SCTs and the VMD method, has 36 features, i.e., 12 for each statistical index.

With the aim of evaluating the capability of the selected indices to provide information about the SCT condition, while keeping a low computational burden, various feature configurations are tested. Figure 10 shows the analyzed configurations. Initially, as shown in Figure 10a, a single statistical index (i.e., only one indicator for each decomposed signal) is considered as input of the ANN, resulting in 12 inputs. Subsequently, as shown in Figure 10b, the combination of two or three indices for each decomposed signal is analyzed, resulting in ANNs with 24 and 36 inputs, respectively. Finally, as shown in Figure 10c, dimensionality reduction using the PCA algorithm is carried out on the data resulting in the best combination of two features reducing from 24 features to 3 dimensions.

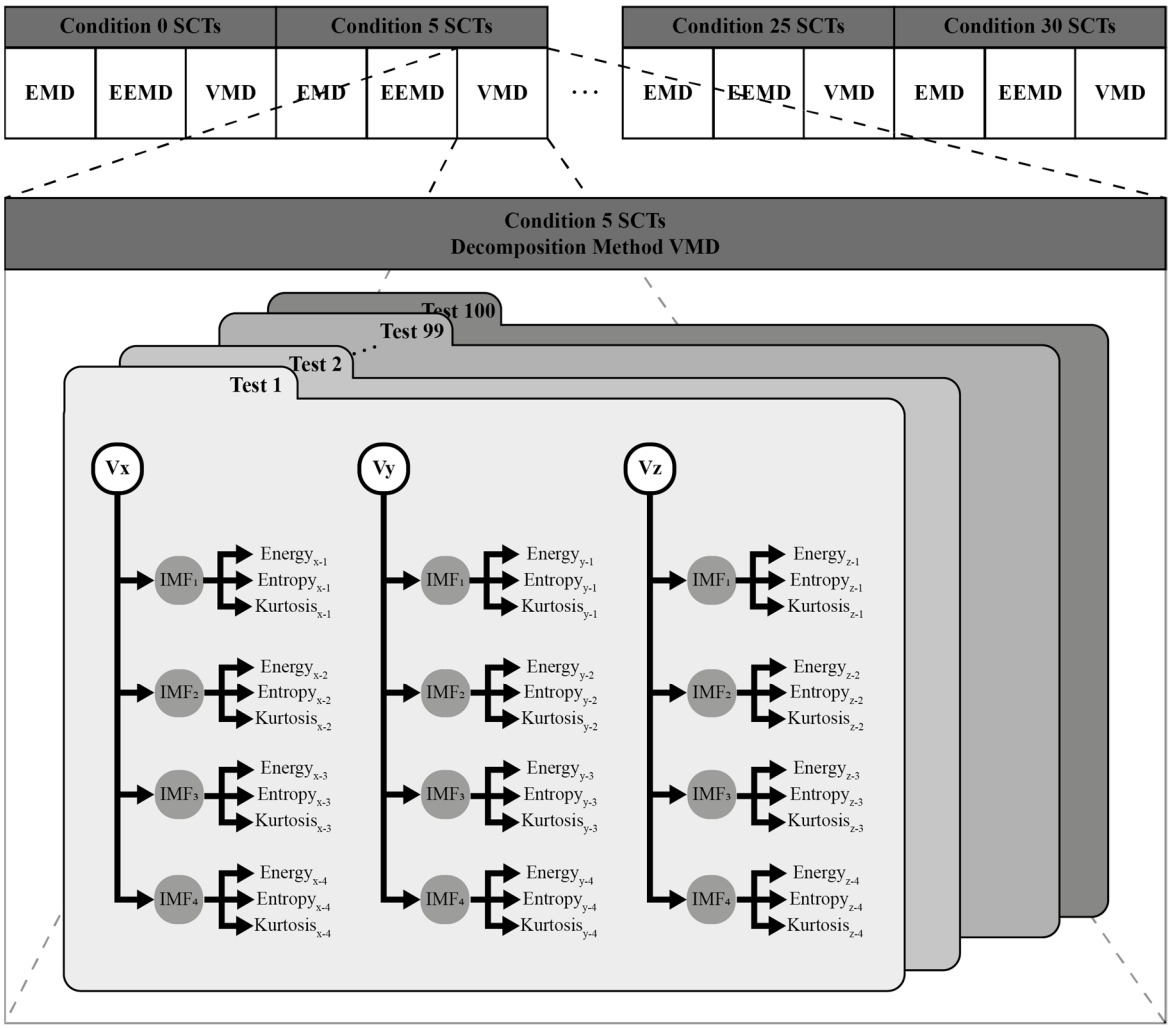


Figure 9. Overall dataset.

On the other hand, it is worth noting that in all these cases, the internal structure of the ANNs maintains eight neurons in the hidden layer and seven outputs that represent each transformer condition. To improve the performance of the training stage, a combination of two methods, i.e., simulated annealing and conjugate gradient, is used. Finally, the SoftMax function is used as the activation function. In all the cases, 80% of the data is used for training and 20% for validation.

The next subsections present the classification results for the cases previously described. Section 4.1 presents the results of using a unique feature extracted from the IMFs of each decomposition method; later, in Section 4.2, the results for the combination of two or three features from the IMFs of each decomposition method are presented. Finally, Section 4.3 shows the results of the PCA-based dimensionality reduction.

4.1. Results for a Single Index

When a single statistical index is used, there are 12 features as inputs for the ANN (see Figure 10a). Table 1 reports the number of errors and average accuracy obtained in the classification for each decomposition method. The average accuracy is calculated using Equation (11), taking into account the number of positive ($TP + TN$) and negative ($FP + FN$) results across the seven classes that represent the conditions studied. This table, shaded in gray, shows that both VMD and EEMD methods coupled with a single index, i.e., energy, present good enough results to evaluate the transformer condition since they have only one error (of 140 possible cases). Although all the methods have an average accuracy greater

than 0.9, the EMD method has the worst performance, reaching more than 10 errors with kurtosis and entropy indices.

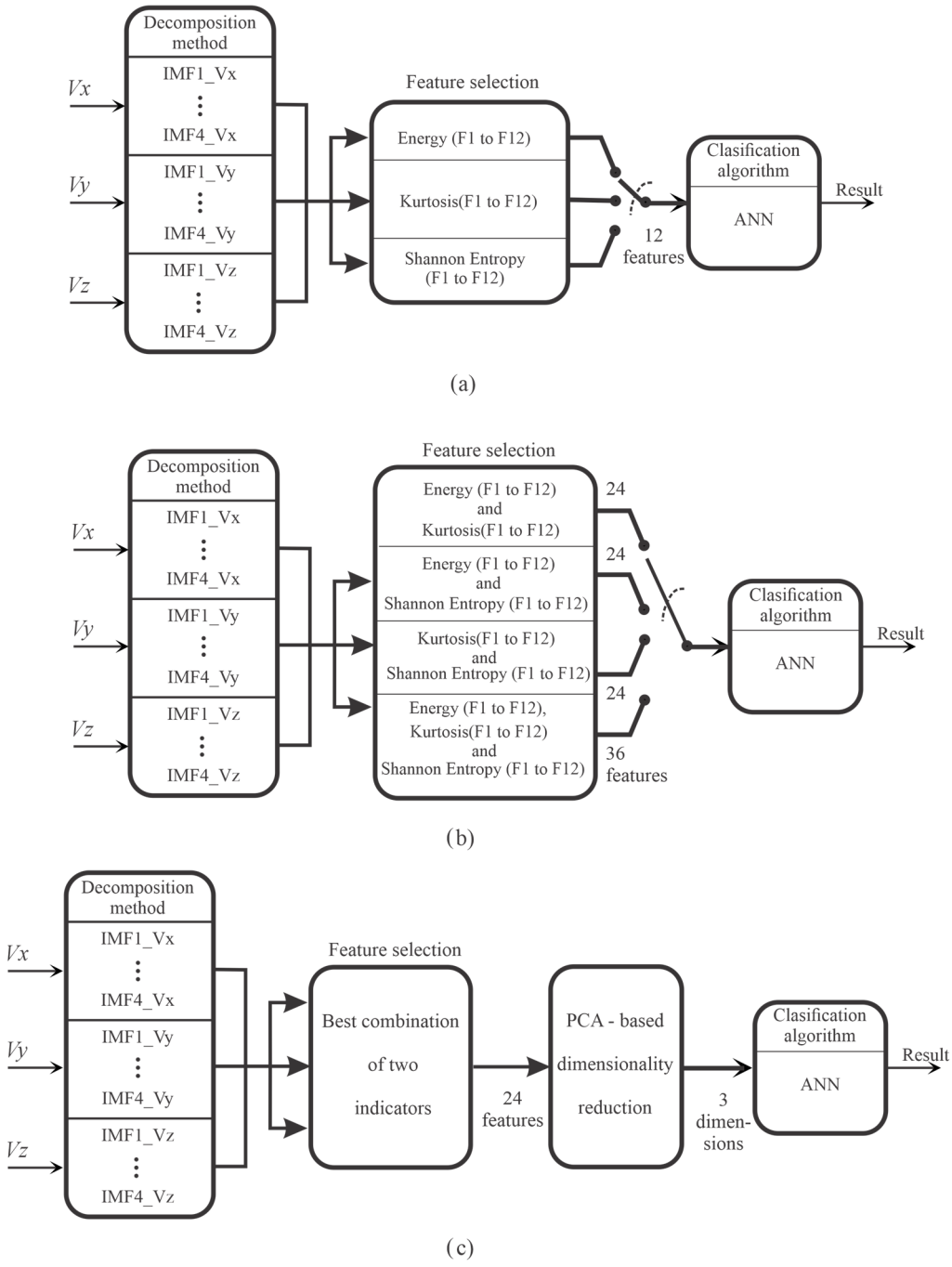


Figure 10. Different data configurations for a: (**a**) single indicator, (**b**) combination of two or three indicators, and (**c**) PCA-based dimensionality reduction.

On the other hand, Table 2 shows the macro-recall and macro-precision values for the three decomposition methods, where EEMD and VMD coupled with energy index (shaded in gray) show the best results as well. Although EMD results have also high values in their metrics, they are not satisfactory because they incorrectly classify the cases of a healthy condition.

Table 1. Validation results: average accuracy.

Feature	EMD		EEMD		VMD	
	Errors	Average Accuracy	Errors	Average Accuracy	Errors	Average Accuracy
Energy	3	0.971	1	0.9928	1	0.9928
Entropy	13	0.9071	4	0.9714	9	0.9357
Kurtosis	12	0.9143	2	0.9857	3	0.9789

Color mark highlights the best results.

Table 2. Summary of validation results (macro-recall and macro-precision) for a single feature.

Feature	EMD		EEMD		VMD	
	Recall_M	Precision_M	Recall_M	Precision_M	Recall_M	Precision_M
Energy	0.971	0.9731	0.9928	0.9932	0.9928	0.9932
Entropy	0.9071	0.9087	0.9714	0.9731	0.9357	0.9386
Kurtosis	0.9143	0.9151	0.9857	0.9860	0.9789	0.9785

Color mark highlights the best results.

In order not to clutter this document with many confusion matrices, only the confusion matrices with the best results are shown. Tables 3 and 4 show the confusion matrices for EEMD and VMD in combination with the energy index, providing only one single error. As can be observed, only one data point is misclassified (highlighted in red) in both tables. In Table 3, a condition of 15 SCTs is classified as 10 SCTs, which is somewhat similar in terms of severity. On the other hand, although the misclassified data point (5 SCTs) is further from its correct class (25 SCTs), it still indicates a damage condition.

Table 3. Confusion matrix for EEMD–energy.

	EEMD-Energy						
	0	**5**	**10**	**15**	**20**	**25**	**30**
0	20	0	0	0	0	0	0
5	0	20	0	0	0	0	0
10	0	0	20	0	0	0	0
15	0	0	1	19	0	0	0
20	0	0	0	0	20	0	0
25	0	0	0	0	0	20	0
30	0	0	0	0	0	0	20

Table 4. Confusion matrix for VMD–energy.

	VMD-Energy						
	0	**5**	**10**	**15**	**20**	**25**	**30**
0	20	0	0	0	0	0	0
5	0	20	0	0	0	0	0
10	0	0	20	0	0	0	0
15	0	0	0	20	0	0	0
20	0	0	0	0	20	0	0
25	0	1	0	0	0	19	0
30	0	0	0	0	0	0	20

4.2. Results for the Combination of Multiple Indices

Table 5 shows the results for the combination of two and three indices, where higher accuracy values than the ones obtained in the previous section, are obtained for the VMD and EEMD methods (shaded in gray). As can be observed in Table 5, five combinations of features provide one of accuracy, indicating that all the tests are correctly classified. Table 6 shows the confusion matrix obtained for the five combinations, i.e., the results are the same for all those combinations. Despite these excellent results, it is important to remember that the number of inputs in the ANN increases; that is, if two or three features are used, the number of inputs is 24 or 36, respectively, which can represent an increase in the computational burden.

Table 5. Summary of validation results for different feature combinations.

Features	EMD		EEMD		VMD	
	Errors	Average Accuracy	Errors	Average Accuracy	Errors	Average Accuracy
Energy-entropy	4	0.9714	3	0.9785	1	0.9928
Kurtosis–entropy	13	0.9071	0	1	5	0.9642
Kurtosis–Energy	17	0.8785	0	1	0	1
Kurtosis–Energy—Entropy	20	0.8571	0	1	0	1

Color mark highlights the best results.

Table 6. Confusion matrix for the cases: (i) EEMD–energy–kurtosis, (ii) EEMD–entropy–kurtosis, (iii) EEMD–energy–entropy–kurtosis, (iv) VMD–energy–kurtosis, and (v) EEMD–energy–entropy–kurtosis.

Confusion Matrix							
	0	**5**	**10**	**15**	**20**	**25**	**30**
0	20	0	0	0	0	0	0
5	0	20	0	0	0	0	0
10	0	0	20	0	0	0	0
15	0	0	0	20	0	0	0
20	0	0	0	0	20	0	0
25	0	0	0	0	0	20	0
30	0	0	0	0	0	0	20

In Table 7, the results of macro-precision and macro-recall can be observed. For the three decomposition methods, the results of 1 in accuracy also generate values of 1 in macro-precision and macro-recall, i.e., a perfect classification. Observing the EMD method metrics, it is possible to identify reliable results; however, as with the results obtained in the previous section, some healthy cases (0 SCTs) are misclassified, i.e., the number of true positives is affected.

Table 7. Summary of validation results (macro-recall and macro-precision) for the combination of features.

Features	EMD		EEMD		VMD	
	$\mathbf{Recall_M}$	$\mathbf{Precision_M}$	$\mathbf{Recall_M}$	$\mathbf{Precision_M}$	$\mathbf{Recall_M}$	$\mathbf{Precision_M}$
Energy–Entropy	0.9714	0.9725	0.9785	0.9802	0.9928	0.9932
Kurtosis–Entropy	0.9071	0.9337	1	1	0.9642	0.9693
Kurtosis–Energy	0.8785	0.93436	1	1	1	1
Kurtosis–Energy–Entropy	0.8571	0.9285	1	1	1	1

Color mark highlights the best results.

If the EEMD and VMD methods are only considered, the integration of kurtosis and energy with the former methods could be considered the best option to evaluate the transformer condition from its decomposed vibration signals because they show the best results in metrics and, at the same time, allows considering a smaller number of inputs than the combination of the three characteristics, somehow representing a lower complexity.

4.3. PCA Results

In order to reduce the number of inputs in the ANN and, consequently, its complexity, the PCA method is applied, where only the kurtosis and energy indices (i.e., indices that provided perfect accuracy values) are considered. By considering these two indices, i.e., 24 features, the PCA is applied. Results indicate that three components explain 97% of the variance of the dataset; therefore, the number of entries is reduced from 24 to 3. Table 8 shows the obtained results. As can be observed, the number of errors increases for EMD and EEMD data; on the contrary, the linear transformation of the VMD data did not affect the ANN performance since a 100% accuracy is obtained. As can be observed, the results for the macro-precision and macro-recall metrics are 1 for the VMD method.

Table 8. Summary of results for the PCA-based dimensionality reduction case.

Number of Entries	EMD		EEMD		VMD	
	Errors	**Average Accuracy**	**Errors**	**Average Accuracy**	**Errors**	**Average Accuracy**
3	18	0.8714	10	0.9286	0	1
	$recall_M$	$precision_M$	$recall_M$	$precision_M$	$recall_M$	$precision_M$
	0.8714	0.8718	0.9286	0.9294	1	1

Table 9 shows a comparison of the computing time and accuracy obtained for each combination of methods tested in this work by considering the best results only; that is, the cases using: (i) a decomposition method and the energy index, (ii) a decomposition method with the combination of energy and kurtosis indices, and, finally, (iii) a decomposition method with the energy and kurtosis indices along with the dimensionality reduction using PCA (3 dimensions). From this table, the best configuration corresponds to VMD with PCA by reducing the dimensionality of the energy and kurtosis indices.

Table 9. Computing times.

Methodology		Computing Time (s)	Accuracy (%)
EMD	Energy	4.05	97.85
	Energy-Kurtosis	4.17	90.71
	PCA	4.06	94.28
EEMD	Energy	39.69	99.28
	Energy-Kurtosis	39.83	100.00
	PCA	39.33	90.71
VMD	Energy	31.46	99.28
	Energy-Kurtosis	32.02	100.00
	PCA	31.25	100.00

Figure 11 shows the results presented in Table 9 as a graph in order to compare the strengths of the analyzed methods in terms of computing time and accuracy. In this regard, it can be observed that the EMD method presents the smallest computing times; however, it has the worst performance and, consequently, the lowest accuracy values. On the contrary, EEMD has a 100% accuracy when the combination of energy and kurtosis is employed, but it has the longest computing time due to its computational complexity, i.e., it requires many iterations of the EMD process to perform an adequate decomposition, making it a technique that requires high resources and computing time. On the other hand, the VMD method has the best performance if a balance between accuracy and computing time is required.

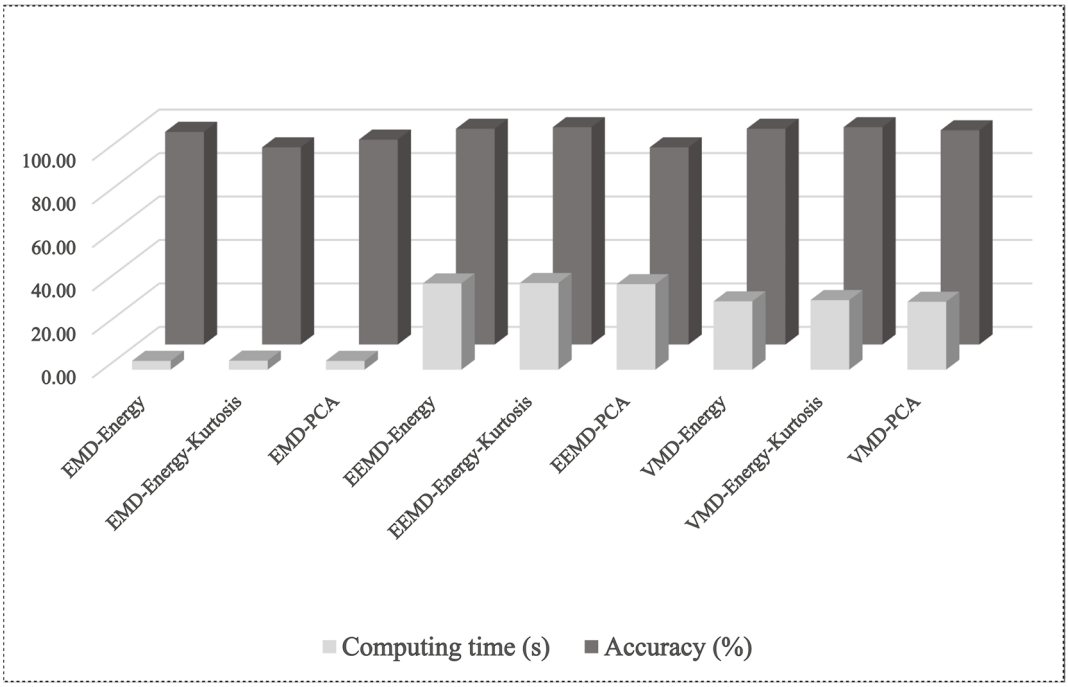


Figure 11. Computing times and accuracy.

5. Discussion

Table 10 shows a comparative analysis between the methodology proposed in this work and different transformer fault detection methodologies documented in the literature for detecting winding faults in a transformer. The comparison includes signal processing techniques, automatic diagnosis algorithms, operation states, measured variables, fault types and their severities.

As can be observed, the diagnosis of SCTs primarily involves the analysis of two types of signals: current [10,18] and vibrations [16,24,25]. Although promising results have been presented in the analysis of current signals, some of the methods used in those studies, such as CEEMD [10] and CNNs [18], are characterized by a high computational load, which can compromise online monitoring of the transformer's condition. In fact, in this work, the EEMD method, similar to the CEEMD method and the NDM method [24], exhibited the longest computing time, which is not a feature of online condition monitoring systems. It is also well known that CNNs require significantly more computing time than classic neural networks, such as the one used in this study; yet, their research applications in this field in future work are not discarded to exploit their capability to process high dimensional data without the need of manual feature engineering tasks. On the other hand, regarding studies that use vibration signals, it is important to note that some of them, such as [16,25], address few classes or severity levels in winding deformations or looseness, unlike the proposal that considers up to six severity levels. This number of analyzed classes and its high accuracy rate is due to the exhaustive analysis of data through the machine learning scheme presented in this work. From this analysis, the proposal becomes a promising solution for the detection of SCTs in transformer windings, both in terms of accuracy and computation time. Although the accuracy values are included in Table 10, it is worth noting that these values are not compared because the analyzed data, the computer equipment, and the methods used are totally different, avoiding a fair comparison. Yet, a qualitative analysis, such as the one carried out in this section, is possible thanks to the information reported in the literature regarding, for instance, the complexity of methods such as CEEMD, NMD, and CNN.

Table 10. Comparison of the proposed methodology against other methodologies reported in the literature for the detection of winding faults.

Reference	Signal Processing Method	Analyzed State	Measured Variable	Detected Fault	Severity Levels	Automatic Diagnosis Algorithm	Accuracy
Proposal	VMD, energy, and kurtosis indices	Steady	Vibrations	SCTs	6	ANN	100%
[10]	CEEMD, Shannon entropy, RMS, and energy indices	Transient and steady	Current	SCTs	6	Not reported	Not reported
[16]	VMD and WT	Transient and steady	Vibrations	Winding deformation	3	SVM	Maximum 96%
[18]	VMD and Choi–Williams distribution	Not reported	Current	Partial discharges	1	Convolutional neural networks	99.5%
[24]	NMD and HT-based RMS	Transient and steady	Vibrations	SCTs	6	Fuzzy logic system	90%
[25]	FFT—total harmonic distortion	Steady	Vibrations	Winding loosening	Incipient	Not reported	Not reported

Although the proposed methodology shows promising results for diagnosing single-phase transformers, scaling it for the analysis of three-phase transformers or other specific operating conditions (e.g., aged transformers) will require recalibration, where the current/actual condition will be taken as the reference point. In this regard, further research will be required, first in a controlled environment and then in real conditions. For three-phase transformers, vibration signals from at least each phase should be acquired to obtain a vibration signature that helps to apply a pattern recognition when a failure or another unusual condition appears, where the type and structure of the transformer will be taken into account for the sensors' location.

6. Conclusions

The detection of SCT faults has attracted the interest of different researchers around the world since a low-severity SCT condition can quickly lead to more severe damages, including the complete shutdown of the transformer. In this regard, this work proposes a new methodology based on the VMD method, energy and kurtosis indices, PCA method, and an ANN, for automatic detection of different numbers of SCTs in transformers. This methodology is also compared with different traditional decomposition methods (i.e., EMD and EEMD methods) and different combinations of two and three indices. Results demonstrate the effectiveness of the proposal since 100% of the analyzed cases are correctly detected and classified.

Regarding the signal processing, the VMD method presented the best results in terms of accuracy and computing time. In addition, the analysis of different sets of features allowed us to observe that the best results are obtained when the energy with kurtosis indices are taken into account; in addition, a dimensionality reduction of up to three features by using PCA was possible without affecting the accuracy. Thus, the ANN with these three inputs (i.e., three entries obtained from the data dimensionality reduction by using PCA) presented the best results.

In future works, different opportunities for research will be explored, e.g., (i) the SCT detection when the transformer is on a loaded condition (linear and non-linear loads), (ii) the detection of other faults on one-phase and three-phase transformers, (iii) the application of optimization algorithms to determine the settings of the different methods that compound the proposed method, (iv) the implementation of nonlinear and manifold-based algorithms for feature selection such as multi-dimensional scaling (MDS), kernel PCA (KPCA), isometric mapping (ISOMAP), kernel Fisher discriminant analysis (KFDA) Laplacian Eigen maps (LE) and local linear embedding (LLE) [26,27], and (v) the hardware implementation of the proposed method into a low-end processor for online condition monitoring.

Author Contributions: Conceptualization and methodology, D.C.-M. and M.V.-R.; software, formal analysis, resources, and data curation, D.C.-M., J.R.H.-R. and M.V.-R.; writing—review and editing, all authors; supervision, project administration, and funding acquisition D.C.-M., D.G.-L., J.P.A.-S., J.C.O.-G. and M.V.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are not publicly available due to they are part of an ongoing study.

Acknowledgments: The authors would like to thank the "Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT)—México" and the "Sistema Nacional de Investigadoras e Investigadores (SNII)–CONAHCYT–México" for their support in this research.

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

References

1. Bigdeli, M.; Azizian, D.; Gharehpetian, G.B. Detection of probability of occurrence, type, and severity of faults in transformer using frequency response analysis based numerical indices. *Measurement* **2021**, *168*, 108322. [CrossRef]
2. Ehsanifar, A.; Allahbakhshi, M.; Tajdinian, M.; Dehghani, M.; Montazeri, Z.; Malik, O.P.; Guerrero, J.M. Transformer inter-turn winding fault detection based on no-load active power loss and reactive power. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 107034. [CrossRef]
3. Abbasi, A.R. Fault detection and diagnosis in power transformers: A comprehensive review and classification of publications and methods. *Electr. Power Syst. Res.* **2022**, *209*, 107990. [CrossRef]
4. Soni, R.; Mehta, B. Review on asset management of power transformer by diagnosing incipient faults and faults identification using various testing methodologies. *Eng. Fail. Anal.* **2021**, *128*, 105634. [CrossRef]
5. Tang, H.F.; Chen, J.; Dong, G.M. Dynamic linear models-based time series decomposition and its application on bearing fault diagnosis. *J. Vib. Control* **2015**, *21*, 975–988. [CrossRef]

6. Bagheri, M.; Zollanvari, A.; Nezhivenko, S. Transformer fault condition prognosis using vibration signals over cloud environment. *IEEE Access* **2018**, *6*, 9862–9874. [CrossRef]
7. Valtierra-Rodriguez, M. Fractal dimension and data mining for detection of short-circuited turns in transformers from vibration signals. *Meas. Sci. Technol.* **2019**, *31*, 025902. [CrossRef]
8. Shi, Y.; Ji, S.; Zhang FRen, F.; Zhu, L.; Lv, L. Multi-frequency acoustic signal under short-circuit transient and its application on the condition monitoring of transformer winding. *IEEE Trans. Power Deliv.* **2019**, *34*, 1666–1673. [CrossRef]
9. Huerta-Rosales, J.R.; Granados-Lieberman, D.; Garcia-Perez, A.; Camarena-Martinez, D.; Amezquita-Sanchez, J.P.; Valtierra-Rodriguez, M. Short-circuited turn fault diagnosis in transformers by using vibration signals, statistical time features, and support vector machines on FPGA. *Sensors* **2021**, *21*, 3598. [CrossRef]
10. Mejia-Barron, A.; Valtierra-Rodriguez, M.; Granados-Lieberman, D.; Olivares-Galvan, J.C.; Escarela-Perez, R. The application of EMD-based methods for diagnosis of winding faults in a transformer using transient and steady state currents. *Measurement* **2018**, *117*, 371–379. [CrossRef]
11. Sahebi, A.; Samet, H.; Ghanbari, T. Identifying internal fault from magnetizing conditions in power transformer using the cascaded implementation of wavelet transform and empirical mode decomposition. *Int. Trans. Electr. Energy Syst.* **2018**, *28*, e2485. [CrossRef]
12. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [CrossRef]
13. Zhao, Q.; Han, T.; Jiang, D.; Yin, K. Application of Variational Mode Decomposition to Feature Isolation and Diagnosis in a Wind Turbine. *J. Vib. Eng. Technol.* **2019**, *7*, 639–646. [CrossRef]
14. Liu, W.; Cao, S.; Chen, Y. Applications of variational mode decomposition in seismic time-frequency analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *81*, V365–V378. [CrossRef]
15. Abdoos, A.A. Detection of current transformer saturation based on variational mode decomposition analysis. *IET Gener. Transm. Distrib.* **2016**, *10*, 2658–2669. [CrossRef]
16. Hong, K.; Wang, L.; Xu, S. A variational mode decomposition approach for degradation assessment of power transformer windings. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 1221–1229. [CrossRef]
17. Meng, J.; Jingyi, D. Research on Magnetizing Inrush Current and Fault Identification of Transformer Based on VMD-SVM. In Proceedings of the 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China, 6–8 November 2020; Volume 1, pp. 172–178. [CrossRef]
18. Gao, A.; Zhu, Y.; Cai, W.; Yi, Z. Pattern Recognition of Partial Discharge Based on VMD-CWD Spectrum and Optimized CNN With Cross-Layer Feature Fusion. *IEEE Access* **2020**, *8*, 151296–151306. [CrossRef]
19. Wang, Z.; Zhou, J.; Wang, J.; Du, W.; Wang, J.; Han, X.; He, G. A novel fault diagnosis method of gearbox based on maximum kurtosis spectral entropy deconvolution. *IEEE Access* **2019**, *7*, 29520–29532. [CrossRef]
20. Caesarendra, W.; Tjahjowidodo, T. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines* **2017**, *5*, 21. [CrossRef]
21. Brunton, S.L.; Kutz, J.N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2022. [CrossRef]
22. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Umar, A.M.; Linus, O.U.; Kiru, M.U. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access* **2019**, *7*, 158820–158846. [CrossRef]
23. Bataineh, A.; Batayneh, W.; Harahsheh, T.; Hijazi, K.; Alrayes, A.; Olimat, M.; Bataineh, A. Early Detection of Cardiac Diseases from Electrocardiogram Using Artificial Intelligence Techniques. *Int. Rev. Model. Simul.* **2021**, *14*, 2021. [CrossRef]
24. Huerta-Rosales, J.R.; Granados-Lieberman, D.; Amezquita-Sanchez, J.P.; Camarena-Martinez, D.; Valtierra-Rodriguez, M. Vibration signal processing-based detection of short-circuited turns in transformers: A nonlinear mode decomposition approach. *Mathematics* **2020**, *8*, 575. [CrossRef]
25. Bagheri, M.; Nezhivenko, S.; Naderi, M.S.; Zollanvari, A. A New Vibration Analysis Approach for Transformer Fault Prognosis over Cloud Environment. *Int. J. Electr. Power Energy Syst.* **2018**, *100*, 104–116. [CrossRef]
26. Van der Maaten, L.J.P.; Postma, E.O.; van den Herik, H.J. Dimensionality Reduction: A Comparative Review. *J. Mach. Learn. Res.* **2009**, *10*, 13.
27. Anowar, F.; Sadaoui, S.; Selim, B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). In *Computer Science Review*; Elsevier Ireland Ltd.: Dublin, Ireland, 2021; Volume 40. [CrossRef]

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