energies

MDPI

Article

Diagnosing Disk-Space Variation in Distribution Power Transformer Windings Using Group Method of Data Handling Artificial Neural Networks

Omid Elahi [1], Reza Behkam [1], Gevork B. Gharehpetian [1] and Fazel Mohammadi [2,3,*]

[1] Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran 15916-39675, Iran
[2] Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 1K3, Canada
[3] Electrical and Computer Engineering and Computer Science Department, University of New Haven, West Haven, CT 06516, USA
* Correspondence: fazel@uwindsor.ca or fazel.mohammadi@ieee.org

Abstract: Monitoring centers in the smart grid exchange the collected data by sensors and smart meters to monitor the current conditions and performance of electric power components. Distribution Power Transformers (DPTs) have a key role in maintaining the integrity of power flow in the smart grid. Online monitoring of DPTs to detect possible faults can potentially increase the reliability of modern power systems. Mechanical defects of DPTs are the major issues in their proper operation that must be detected in their early stage of occurrence. One of the most effective solutions for diagnosing mechanical defects in DPTs is Frequency Response Analysis (FRA). In this study, an appropriate condition monitoring scheme for DPTs is developed to identify even minor winding defects. Disk-Space Variation (DSV), a common DPT windings fault, is applied to the 20 kV-winding of a 1.6 MVA DPT in various locations and with different severity. Their corresponding frequency responses are then computed, and all four components of the frequency responses, i.e., amplitude, argument, and real and imaginary parts, are evaluated. Different data-driven-based indices are implemented to extract appropriate feature vectors in the preprocessing stage. Group Method of Data Handling (GMDH) Artificial Neural Networks is proposed to assist monitoring centers in interpreting FRA signatures and identifying DPT defects at primary stages. GMDH has a data-dependent structure, which gives high flexibility to modeling nonlinear characteristics of FRA test results with different data sizes. It is demonstrated that the proposed approach is capable of accurately determining the fault location and fault severity. The proposed Artificial Intelligence (AI)-based approach is used to extract essential features from frequency response traces in order to detect the position and degree of Disk-Space Variation (DSV) in the DPT windings. The experimental results verify the effectiveness of the proposed methods in determining the severity and location of DSV defects.

Keywords: smart grid; online monitoring; distribution power transformer (DPT); disk-space variation (DSV); data mining

Citation: Elahi, O.; Behkam, R.; Gharehpetian, G.B.; Mohammadi, F. Diagnosing Disk-Space Variation in Distribution Power Transformer Windings Using Group Method of Data Handling Artificial Neural Networks. *Energies* **2022**, *15*, 8885. https://doi.org/10.3390/en15238885

Academic Editor: Abu-Siada Ahmed

Received: 5 October 2022
Accepted: 21 November 2022
Published: 24 November 2022

1. Introduction

Distribution Power Transformers (DPTs) are considered as indispensable elements for the power delivery process in modern power systems. DPTs are vulnerable to various faults, including mechanical winding deformations, the second highest rate among all DPT faults. Disk-Space Variation (DSV), Radial Deformation (RD), and Axial Displacement (AD) are the three mechanical winding deformations in DPTs. Discriminating DPT defects at the preliminary stages is of great importance to maintain the integrity and reliability of modern power grids [1–4]. Operational data centers gather the measured datasets of DPTs on a daylong basis to monitor the state of health condition of DPTs. The Frequency Response Analysis (FRA) method, which is also known as the Transfer Function (TF) technique, can

be utilized in the monitoring process, providing an efficient way of condition assessment [5]. The specialized operators do not have access to a reliable and standardized code to interpret the TF test results. In this regard, enormous research studies have been conducted in recent years to deal with interpreting the TF test results. Equivalent RLC circuit modeling [5–7], utilizing Numerical Indicators (NIns) [8–11], and machine learning methods [12–20] are the three main approaches to interpret TF test results. Because of the nonlinear nature of the TF results, Artificial Intelligence (AI) techniques can result in more accurate interpretations to identify the location(s) of faults and their impacts.

In [6], a high-frequency three-winding model of an actual DPT is considered, which has repetitive RLC sections and mutual inductances. The TF signature of the DPT is measured, and model parameters are estimated to have the same TF signature. The motivation of this method is that the TF of a DPT reflects its physical characteristics. Parameter adjustments are made based on the individual impacts on the simulated TF. In [7], different mechanical and electrical faults affecting the TF traces are checked. The defects are simulated with the 3D finite integration analysis. The amplitude part of the TF and four connection schemes of the TF test are considered. The impacts of the mechanical faults on different frequency sub-bands are investigated. The standard deviation of differences between intact and faulty signatures is used to extract features with the windowing method. The minimum numerical value of the extracted features in each sub-band is used for sensitivity analysis. In [8], the occurrence, type, and severity of common DPT faults, including AD, RD, DSV, core deformation, and Short Circuit (SC), are investigated. The amplitude part of the TFs of the faults is measured at different severity. The Interval Maximum of the Global Maximum (IMGM) index and Grasshopper Optimization Algorithm (GOA)-based K-means are applied to detect the type of faults. Finally, fault severity is defined based on different indices' variations. In [9], the type of faults in the DPT windings is detected. FRA regression analysis, based on the measured TF, defines the condition of the DPT considering the Correlation Coefficient (CC) index. The output of the regression analysis is a single number that is compared with the coefficient of determination, and accordingly, the type of fault is detected. Features are extracted from low, medium, and high-frequency sub-bands of the TF signatures using the CC and Absolute Sum of Logarithmic Error (ASLE) indices. The defining boundary for NIns output for detection applications is a major issue, and NIns can be used along with the other methods to assist in precise fault detection. Applying NIns to the entire frequency range of the TF results can lead to a wrong FRA interpretation. In [10], a moving window approach is proposed to extract features from the data in the window using four NIns named CC, Spectrum Deviation (SD), Minimum Squared Error (MSE), and ASLE. Different widths of the window are selected for scanning the TF traces and extracting features. It is found that the window's width affects the sensitivity of the extracted features with respect to deformations. Hence, the proper width of the window can be defined. However, in [10], the suitable window width is needed to be defined for other TF results again. Besides, the value of fault detection criteria must be defined, possibly inaccurate. By employing the windowing method, the dimension of the extracted features is high, and if feature reduction techniques reduce such a dimension, it may be easier to interpret. However, there is no standard to define the window width size. Three appropriate NIns, called Cross-Correlation Factor (CCF), Sum of Errors (SE), and Lin's Concordance Coefficient (LCC), are introduced in [11] as monotonic, and the most reliable indices among all of the NIns are applied to interpret TF test results. In [12], an Empirical Wavelet Transform (EWT) is applied to extract a set of Empirical Wavelet Functions (EWFs) from vibration signals. The Kernel Extreme Learning Machine (KELM) optimized with the Salp Swarm Algorithm (SSA) is used to determine a DPT's abnormal functions and the type of the identified fault. Using SSA led to a longer training time, but it is necessary for effective tuning of the parameters. Moreover, training high-dimensional features is time-consuming, mainly when an optimization algorithm is used. A Multi-layer Perceptron (MLP) Neural Networks is trained with the obtained features related to LCC index in [13]. The proposed MLP using TF results is employed to discriminate AD and DSV faults. In [14], TF traces are

analyzed for online monitoring of DPTs to detect SC, RD, and DSV faults. The DSV fault detection accuracy is lower compared with SC and RD. The Support Vector Machine (SVM) classifier using the extracted features by the MSE indicator could detect fault types with an accuracy of 83.3%. A DPT is simulated with the lumped parameter-based model in [15]. In this model, six faults with different levels are simulated, and corresponding frequency responses are achieved. Image processing methods (grayscale, image enhancement, image sampling, and projection) are utilized to extract high-dimensional features. The Decision Tree (DT) model and a fully-connected Artificial Neural Networks (ANN) are trained with such features. Training a model with high-dimensional features is also time-consuming [16]. Feature extraction and dimension reduction methods can be applied to have fewer features to be considered as the main inputs of the models [17,18]. The adaptive frequency slicing algorithm is implemented in [19] to define frequency sub-bands for the TF results of DPTs. The Comparative Standard Deviation (CSD), SD, CCF, LCC, and SE are the NIns utilized to extract features from sub-bands. In addition, only the amplitude part of the TF is used to detect the existence of winding deformations, but it cannot determine the location and severity of mechanical winding deformations. In [20], an unsupervised isometric mapping approach is employed to identify SC faults. In all aforementioned research studies, only the TF's amplitude part is considered, and fault types are detected using different AI-based approaches. However, determining both location and severity of winding deformations requires more investigation.

In this paper, an AI-based method using the Group Method of Data Handling (GMDH) algorithm is proposed to determine DSV faults' location and severities in DPTs. The proposed method can be utilized in power systems' data centers to identify defected signals related to DPTs; therefore, the monitoring and maintenance of the apparatus can be efficiently performed. In addition, complicated computations and non-standardized lookup tables using NIns to determine only the existence of faults can be replaced with the proposed method to have more accurate automatic decisions; not only to detect the occurrence of faults but also to determine their location and severity.

The significant contributions of this paper can be determined as follows:

- A new AI-based interpreter of the TF test results has been proposed. A GMDH artificial neural network has been employed to determine the severity and location of DSV faults. The results of classification using GMDH have been compared to the results of the MLP neural network. In order to assess the performance of the intelligent classifiers, a well-known method called k-fold cross validation has been utilized;
- At the feature extraction stage, ten appropriate NIns used to extract feature groups to feed the proposed intelligent fault detectors;
- Sensitivity analysis considering all TF parts (imaginary, real, magnitude, and phase) has been carried out.

The structure of the rest of the study can be described as follows: Section 2 describes the experimental study and process of collecting TF test results. Detailed information about the proposed intelligent fault identifier is explained in Section 3. The results of classifying the location and severity of DSV deformations are explored in Section 4. Sensitivity analysis of TF components considering various NIns is investigated in Section 5. Finally, Section 6 concludes this paper.

2. Experimental Study and Data Preparation

2.1. Experimental Setup

The descriptions related to how the DSV defects are applied to the DPT windings and detailed information about the practical setup utilized in this paper are provided in this section. Windings of one phase of a three-phase DPT are investigated. The nominal power capacity and voltage of the understudy DPT are 1.6 MVA and 20 kV/400 V, respectively. A total number of 38 individual disks, including 782 turns, exists in high voltage windings, and two layers containing 27 turns are in low voltage windings. The total height of the high voltage windings is 494 mm; therefore, displacements with an amount of 5 mm are roughly

1% of the full height. The radius of 176.5 mm and 118.5 mm are related to the outside and inside parts of the high voltage windings, respectively. Similarly, the outer and inner sides of the low voltage windings are 106 mm and 93 mm, respectively. The core's radius is 90 mm. The dimension measurements associated with high voltage and low voltage windings are 8.5 × 2.12 mm and 11.8 × 3.35 mm, respectively. A cylindrical aluminum represents the DPT core's equipotential outer surface in the case study. Furthermore, at frequencies above 5 kHz, the cylinder is considered in order to indicate the capacitive interaction between the core and windings because of the fact that the windings' behavior can only be reliably represented by the core's external surface. The TF is obtained via impedance measurement conducted on the primary high voltage windings while leaving the secondary windings disconnected. The impedance analyzer, called WAYN KERR 6500B, is employed to collect the results. While executing the DSV test, the space between Double Disks (DDs) altered from 8 to 40 mm in six various severities, and for each of which, the related TF is collected and measured. The DSV deformation is implemented on the whole upper section of the high voltage winding, and these deformations are localized at ten different locations. Therefore, 6 × 10 specific cases existed, and the measurements are performed to have access to input impedance for each case. The whole practical device is demonstrated in Figure 1a. The High Voltage (HV) and Low Voltage (LV) windings employed in this study are depicted. Figure 1b–d respectively represent the low, medium, and high severities of applied DSV defects to DDs in the high voltage windings.

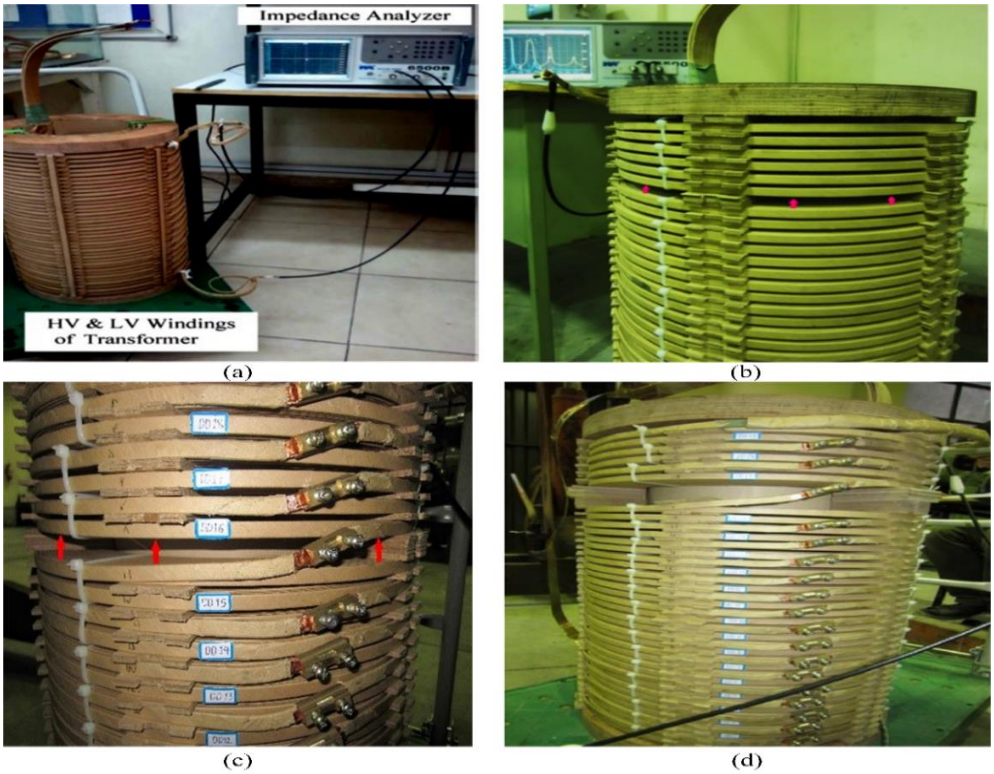


Figure 1. Experimental setup, (**a**) DPT windings and impedance analyzer; (**b**) low severity of DSV deformation; (**c**) average severity of DSV deformation; and (**d**) high severity DSV deformation.

2.2. Data Collection

TF test results are obtained from Equation (1):

$$TF = 20\log\left(\left|\frac{V_{in}(f)}{I_{out}(f)}\right|\right) \quad (1)$$

where TF is the transfer function signature, $I_{out}(f)$ denotes the ground current, and $V_{in}(f)$ represents the input voltage.

The measured TF traces related to the intact and deformed conditions of the windings are illustrated in Figure 2. Both amplitude and phase parts of the TF considering DSV deformations in various locations and severity are shown in the figure.

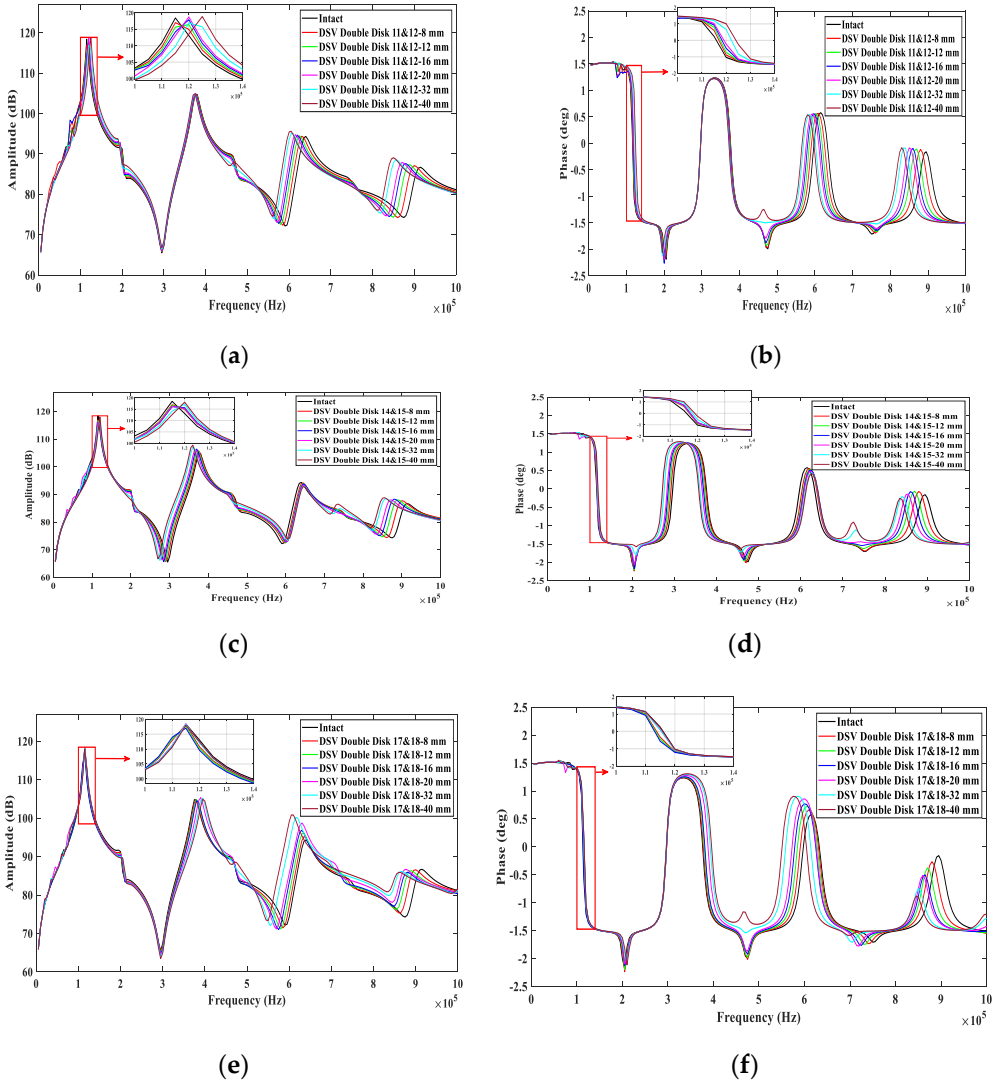


Figure 2. TF signatures illustrating intact and deformed traces, (**a**) the amplitude part of the TF for intact and DSV deformation between DD 11 and 12; (**b**) with the phase part of the TF for intact and DSV deformation between DD 11 and 12; (**c**) the amplitude part of the TF for intact and DSV deformation between DD 14 and 15; (**d**) with the phase part of the TF for intact and DSV deformation between DD 14 and 15; (**e**) the amplitude part of the TF for intact and DSV deformation between DD 17 and 18; and (**f**) with the phase part of the TF for intact and DSV deformation between DD 17 and 18.

Figure 3 illustrates the entire process of the TF results interpretation in order to distinguish faults while monitoring DPT conditions. In order to gather feature sets, ten NIns have been employed in this paper. The NIns are implemented on TF signatures in different frequency ranges. According to the frequencies related to resonance and anti-resonance points, four intervals in the frequency domain are considered, which are (5–225 kHz), (225–500 kHz), (500–775 kHz), and (775–1 MHz).

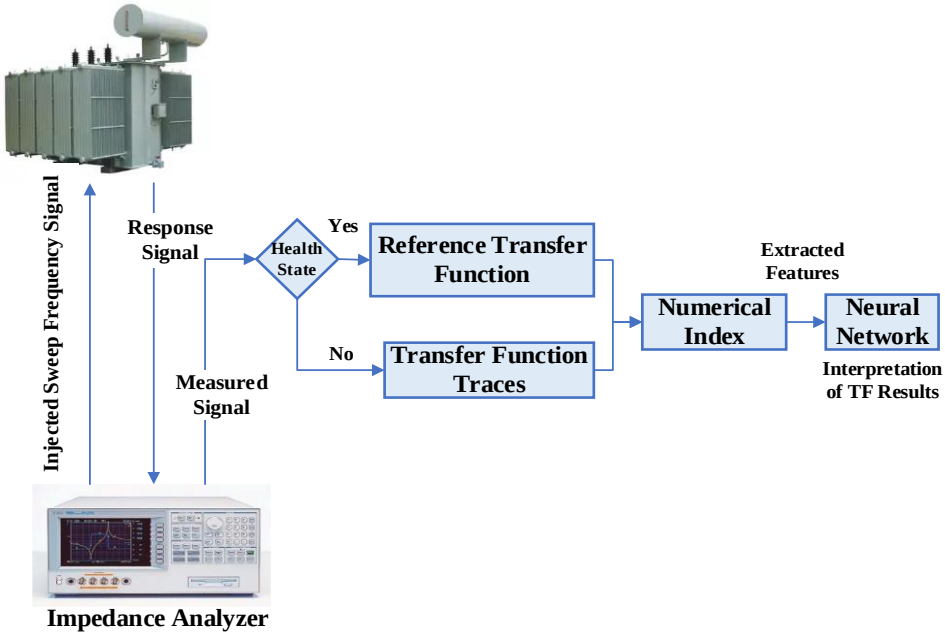


Figure 3. Structure of fault diagnosis and TF test results' interpretation procedure.

It should be noted that DDs 8 up to 11 are labeled as the bottom section, DDs 12 up to 15 are considered as the middle part, and DDs 16 up to 19 are labeled as the top section. Regarding the severity of DSV faults, fault severities between 8 mm and 12 mm are labeled as low, severities between 16 mm up to 20 mm are considered average, and DSV faults with severity greater than 20 mm are labeled as severe. Therefore, there are four feature sets according to the above-mentioned frequency sub-bands as input of the proposed classifier, and three classes are allocated for the classifier's output, indicating the fault's severity or location.

The applied NIns are listed as follows, where, X and Y denote the TF vectors in the intact and deformed cases, respectively, and $\overline{X}$ and $\overline{Y}$ are their average values, and N represents the TF's sample number:

$$SD = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left[\frac{x_i - A_i}{A_i}\right]^2 + \left[\frac{y_i - A_i}{A_i}\right]^2} \tag{2}$$

$$A_i = \frac{x_i + y_i}{2}$$

$$CCF = \frac{\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{X})^2 \sum_{i=1}^{N}(y_i - \overline{Y})^2}} \tag{3}$$

$$CSD = \sqrt{\frac{\sum_{i=1}^{N}[(x_i - \overline{x}) - (y_i - \overline{y})]^2}{N-1}} \tag{4}$$

$$\begin{aligned} LCC &= \frac{2S_{xy}}{(\overline{Y}-\overline{X})^2 + S_x^2 + S_y^2} \\ S_x^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^2, \\ S_y^2 &= \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{Y})^2, \\ S_{xy} &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y}) \end{aligned} \tag{5}$$

$$FP\% = [1 - \sqrt{\frac{\sum_{i=1}^{N} |(x_i - y_i)|^2}{\sum_{i=1}^{N} |(x_i - \overline{X})|^2}}] \times 100 \quad (6)$$

$$NRMSD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{N}} / \overline{Y} \quad (7)$$

$$SE = \frac{\sum_{i=1}^{N} (y_i - x_i)}{N} \quad (8)$$

$$SSE = \frac{\sum_{i=1}^{N} (y_i - x_i)^2}{N} \quad (9)$$

$$SSMMRE = \frac{\sum_{i=1}^{N} \left(\frac{max(y_i, x_i)}{min(y_i, x_i)} - 1\right)^2}{N} \quad (10)$$

$$SSRE = \sqrt{\frac{\sum_{i=1}^{N} \left(\frac{y_i}{x_i} - 1\right)^2}{N - 1}} \quad (11)$$

3. Materials and Methods

In this section, the description of the proposed intelligent interpreter of the TF test results has been prepared.

GMDH Artificial Neural Networks

The GMDH Artificial Neural Networks algorithm was used to find a nonlinear relationship between data inputs and outputs [21]. GMDH is a self-organizing map in which, based on its algorithm, the number of neurons in each layer and the number of layers are defined automatically. An input relates to the output with the Volterra series as formulated in Equation (12.) In this equation, x_i is the ith feature of the input and a_0, a_{ij}, a_{ijk}, ... , are equation variables estimated in the training process:

$$y_n = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M} \sum_{j=1}^{M} a_{ij} x_i x_j + \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{M} a_{ijk} x_i x_j x_k + \ldots \quad (12)$$

GMDH is a feed-forward network. It has an input, middle, and output layer with one neuron that can be used for binary classification. The number of neurons in the first layer is $\binom{n}{2}$ in which n is the number of input features. Two features of data input are fed to each neuron. In Figure 4, the first and second data input features fed the neuron. The Volterra series can be rewritten as a series of second-order equations that pairs of data features input to each of them, and $G(x_1, x_2)$ equation is one of them [22,23].

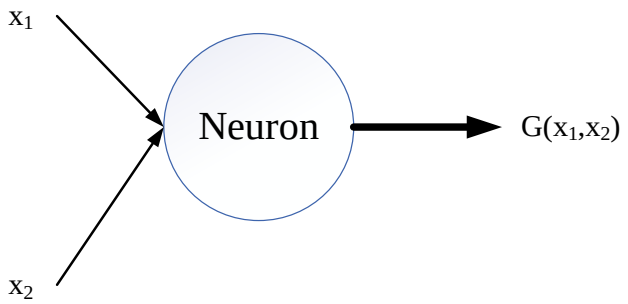


Figure 4. A neuron of the GMDH networks.

Equation (12) can be re-written for two inputs as follows:

$$G(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2 \quad (13)$$

Coefficients including a_0 to a_5 are estimated in the training process. The output of $G(x_1,\ x_2)$ equation is between zero and one, and a_0 to a_5 are initialized. The output of the data is compared with the $G(x_1,\ x_2)$ output, and the squared error is calculated. Based on the principle of having the minimum error, equation variables are estimated.

In order to minimize the differences between pridicted values and actual ones, Equations (14) and (15) can be derived:

$$\hat{y}_n = a_0 + a_1x_{in} + a_2x_{jn} + a_3x_{in}x_{jn} + a_4x_{in}^2 + a_5x_{jn}^2 \quad (14)$$

$$e = \sum_{n=1}^{N} (\hat{y}_n - y_n)^2 \quad (15)$$

where x_{in} and x_{jn} are the ith and jth features of the nth data. Moreover, $\hat{y}_n$ is the output of one random neuron whose inputs are x_{in} and x_{jn} in Equation (14).

Coefficients including a_0 to a_5 are initialized; and $\hat{y}_n$ is calculated for $n = 1$ to N for which N is the number of training data. In addition, y_n is the actual label of the nth data. Then, $\hat{y}_n$ is compared with y_n, and the neuron squared error denoted with e is calculated.

There is an aim to minimize the e value. Therefore, the $\frac{\partial e}{\partial a_i} = 0$ equation should be solved. The neuron variables are denoted with $A = [a_0, a_1, a_2, a_3,\ a_4,\ a_5]$ and variables coefficients are shown as the Y vector, where $Y = \left[1, x_{in},\ x_{jn},\ x_{in}^2,\ x_{jn}^2\right]$. The A vector is estimated by calculating Equation (16). This equation is the result of Ordinary Least Squares fitting. In this equation, $X = Y^TY$ and $B = (y_nY)^T$. The obtained equation is calculated by a computer, and all the neuron variables are estimated. The whole processes are repeated for all neurons of the first layer. Each pair of features is the input of each neuron. Therefore, $\binom{N_f}{2}$ neurons are located in the first layer and N_f is the number of features:

$$\sum_{n=1}^{N} AX = \sum_{n=1}^{N} B \quad (16)$$

From $\binom{N_f}{2}$ neurons in the first layer, N_f neurons remain, and others are removed. Neurons validation loss is calculated with the validation dataset. Selected neurons have the minimum validation losses. The average or top of validation losses of selected neurons is considered the layer validation loss. In addition, layer training loss is calculated with the training dataset. The next layer is created with the same process by regarding the previous layer's neurons' outputs as the inputs. The creation of new layers is then continued.

Suppose the training loss of a new layer reaches a threshold value, or validation loss of it does not decrease. In that case, the training process ends up selecting one neuron of the last layer with the minimum validation loss. This process is shown in Figure 5. GMDH does not need optimization of hyper-parameters and can find a nonlinear relationship between inputs and outputs [24,25]. GMDH Artificial Neural Networks can solve regression or classification problems. In this paper, GMDHs solve several multi-class classification problems, and the results are presented. It should be noted that the One Hot Encoding method has been employed to indicate the output class number, and therefore, the output is not just zero or one; it is a vector of zeros and ones.

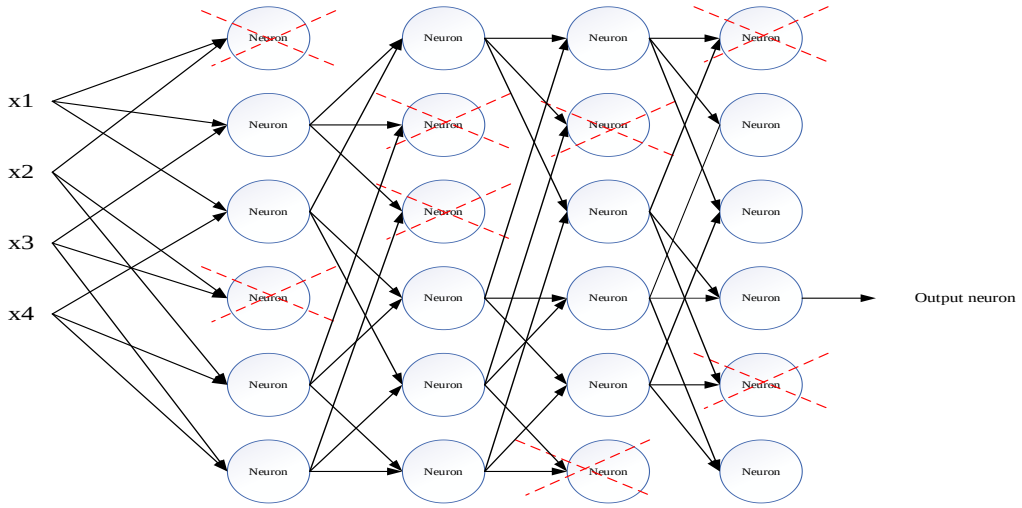


Figure 5. GMDH Artificial Neural Networks.

4. Results

In this section, results of GMDH and MLP Neural Networks for interpreting FRA results have been presented. The MLP Neural Networks has been widely used for medium and large-scale applications. Due to the many applications of this method, its implementation has been carried out completely by the Google Brain Team and provided to users by the Tensor-Flow framework. Selecting the number of layers, determining the number of neurons in each layer, the optimizer, the activation function, the learning rate, the value of Momentum, the number of iterations of the training process, and many other hyper-parameters must be set correctly. By correctly tuning such hyper-parameters, an acceptable and suitable output can be obtained. However, the GMDH model has a different performance, and by automatically calculating the number of neurons and determining the number of layers to the optimal level, there is no need to select the learning rate and the number of training iterations. However, if there are too many input features, it will take a long time to train this network.

The accuracy of the MLP Neural Networks has been high in the DPT fault detection, severity determination, and localization of the apparatus in different research studies. Therefore, MLP Neural Networks can be tuned correctly to determine the severity and location of the DSV fault with high accuracy. In addition, the results must be reproducible to show the impact of selecting different datasets for testing the model. In this paper, 20% of the dataset is selected randomly as the test data, and the rest is used to train the model properly. In order to obtain more reliable accuracy results with appropriate variance among the classification results, this process is repeated with thirty random state values, and, at the end, a box plot of training, validation, and test accuracies is plotted. In addition, average training, validation and test accuracy, corresponding confusion matrix, and ROC curve are provided.

The advantage of the GMDH Artificial Neural Networks is that it does not need hyper-parameter tuning. Therefore, if the GMDH performance is reasonable and similar to the MLP, the time of the training process reduces significantly. In this paper, three GMDH Artificial Neural Networks are used to determine the severity or location of the DSV fault. Therefore, in the GMDH training process, three training and validation loss curves are plotted for each GMDH. In the layer where GMDH has the minimum validation loss, the GMDH variables are saved. MLP variables are saved in the same process based on minimum validation loss. It is crucial to train both models well. The training, validation, and test process are shown in Figure 6. The average training, validation, and test accuracies for the thirty iterations are reported.

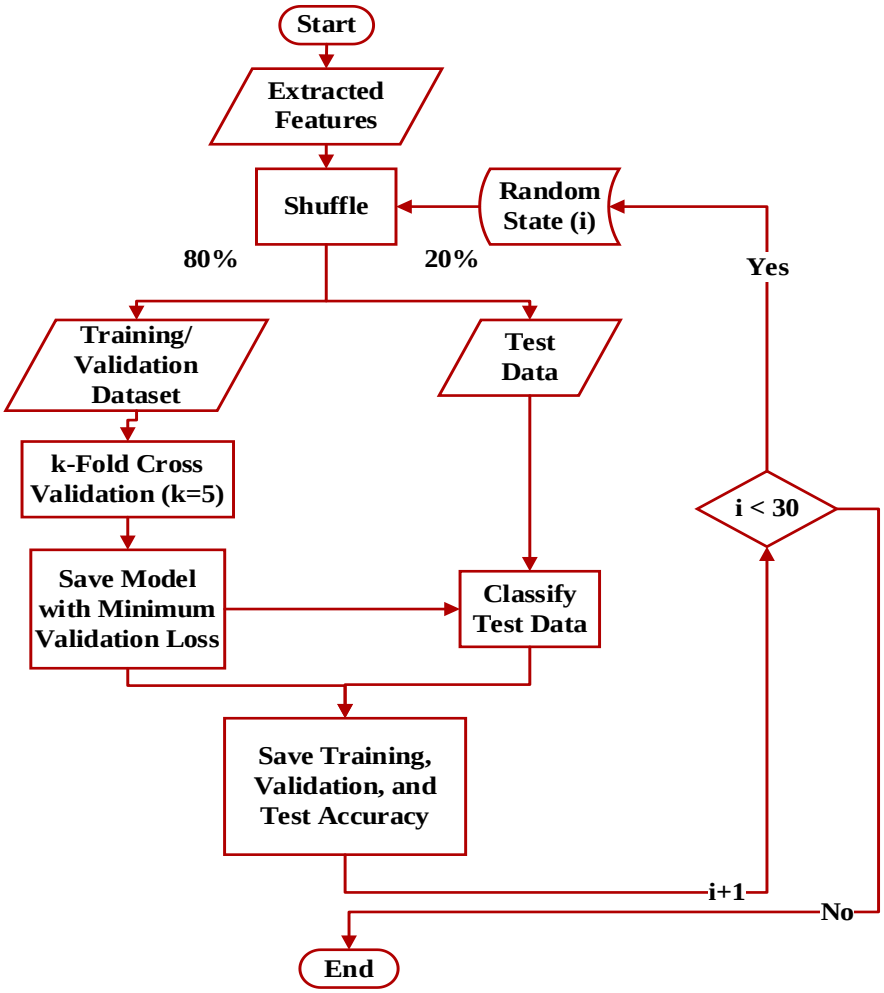


Figure 6. The process of training models and saving training, validation, and test accuracies.

4.1. Localization of DSV Faults

The performance of the MLP and GMDH models with the proposed method in determining the DSV fault location is shown in Figures 7 and 8. The average test accuracy of the MLP model in determining the fault location reaches 94.72% using the extracted features from the real part of the TF with the Normalized Root Mean Square Deviation (NRMSD) index. The GMDH model also localizes the fault with an average test accuracy of 87.22% using the same features. These results are the highest average test accuracies when models are trained with the extracted features. Additionally, the MLP model using the phase part of the TF (direct FRA data without employing NIns) determines the fault location with an accuracy of 99.17%, and the GMDH model determines the fault location with an accuracy of 95.28% using the data of the same part of the TF. When the direct FRA data are used to train the GMDH model, due to a large number of features and the fact that training the GMDH model with these features increases the training time significantly, by applying a dimension reduction method, called the Principal Component Analysis (PCA) algorithm, the number of features is reduced to eight.

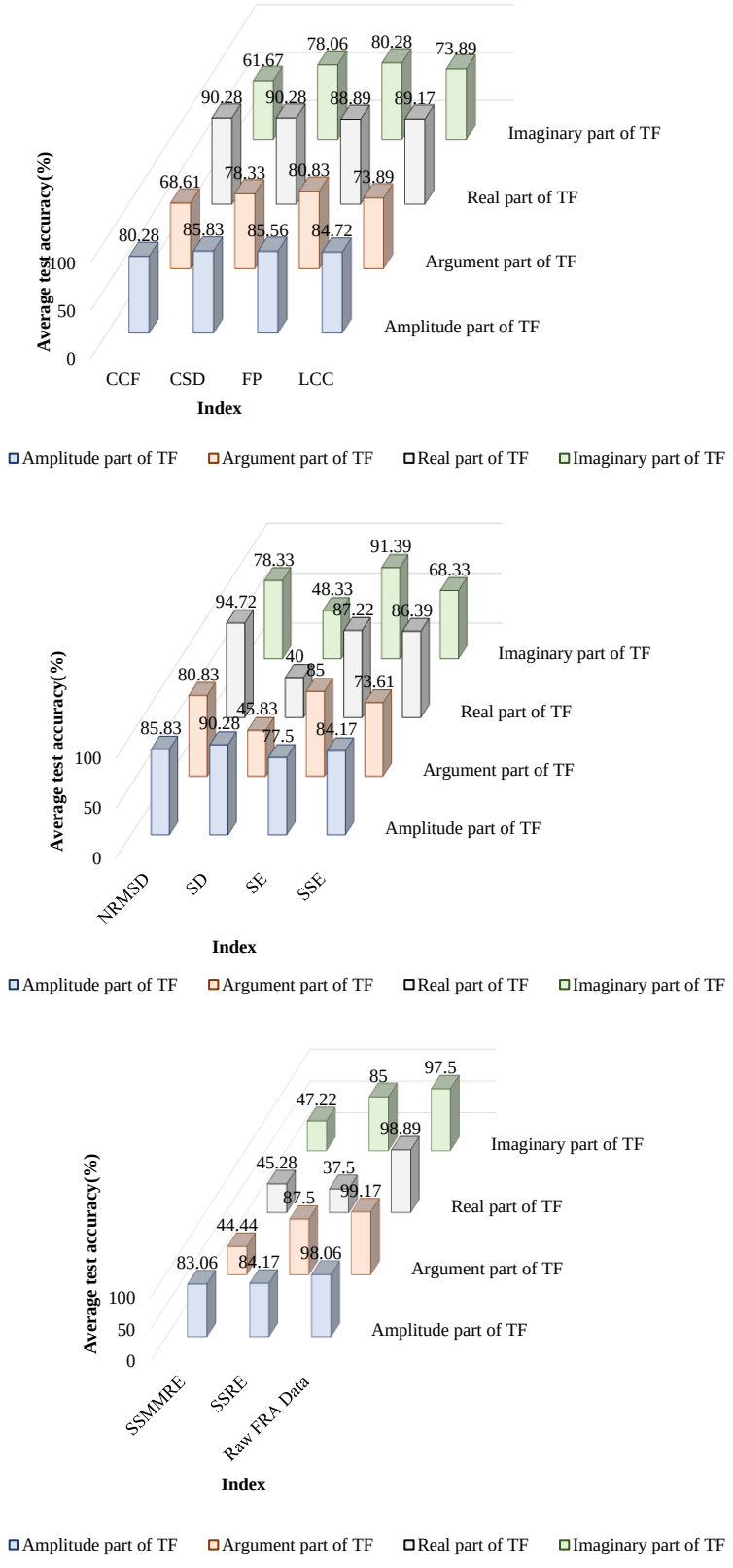


Figure 7. Evaluating the performance of the MLP model in determining the DSV fault location using direct data of FRA and the extracted features by statistical indices.

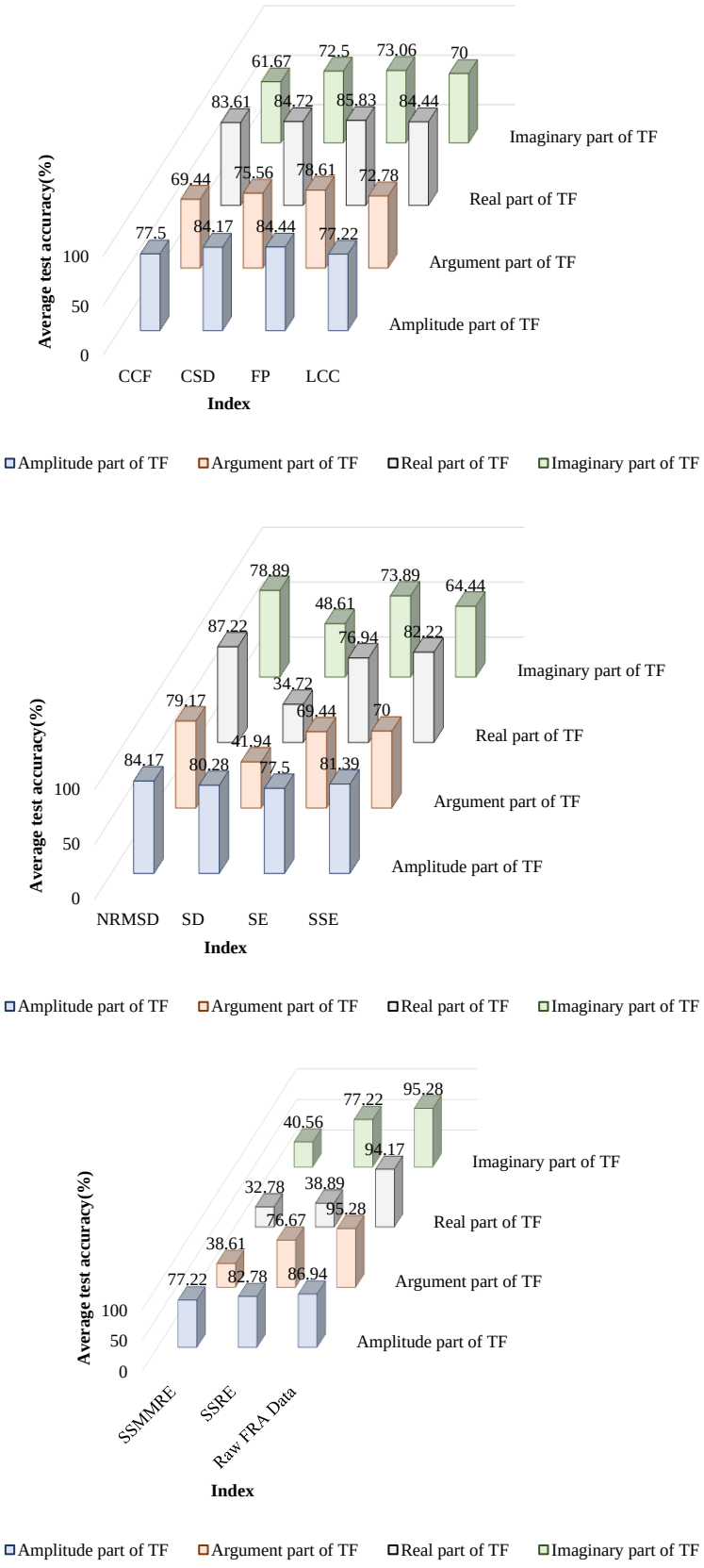


Figure 8. Evaluating the performance of the GMDH model in determining the DSV fault location using direct data of FRA and the extracted features by statistical indices.

4.2. Determining the Severity of DSV Faults

The performance of the MLP and GMDH models with the proposed method in the determination of the DSV fault severity is shown in Figures 9 and 10. The average test accuracy of the MLP model in determining the fault severity reaches 91.39% using the extracted features from the phase part of the TF with the NRMSD index. The GMDH model also determines the severity of the fault with an average test accuracy of 85.56% using the same features. Such results are the highest average test accuracies when models are trained with the extracted features. Furthermore, the MLP model using the data of the real part of the TF (direct FRA data without employing NIns) determines the fault severity with an accuracy of 78.61%, and the GMDH model determines the fault location with an accuracy of 77.78% using the data of the same part of the TF. Although MLP and GMDH models in determining the DSV location had high accuracies when they are trained with raw FRA data without employing NIns, they cannot determine the severity of DSV faults using the raw FRA data. In this regard, the models should be trained with the extracted features with employing NIns to determine the severity of DSV faults with high accuracies.

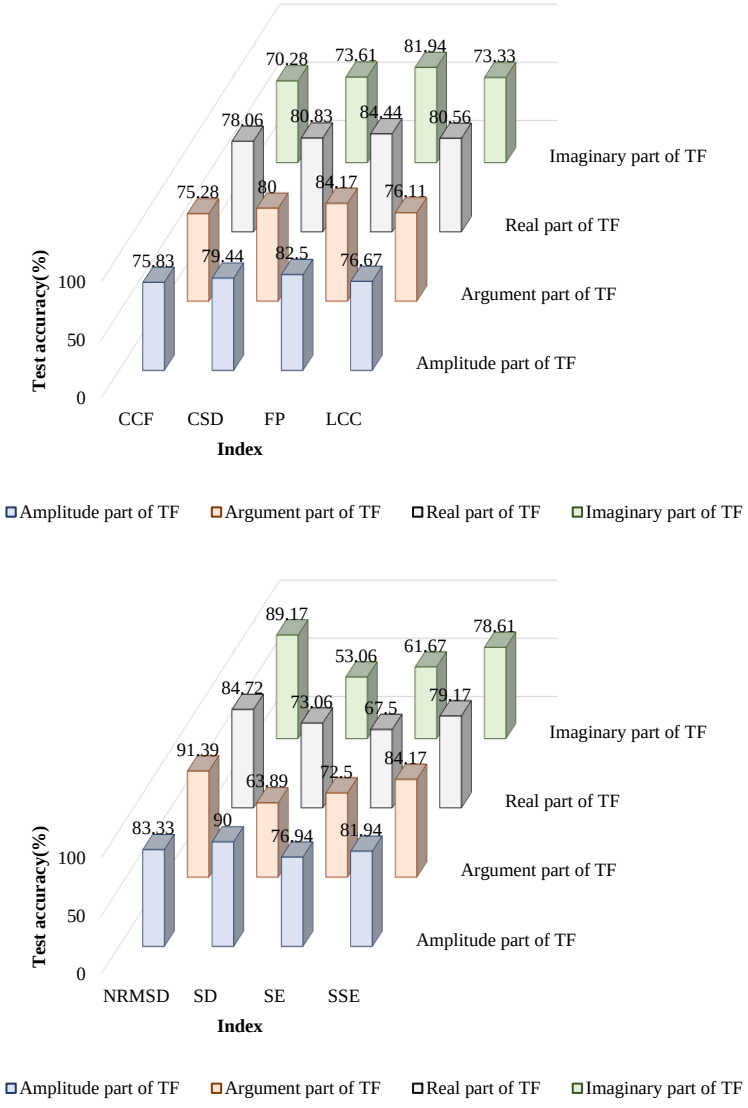


Figure 9. *Cont.*

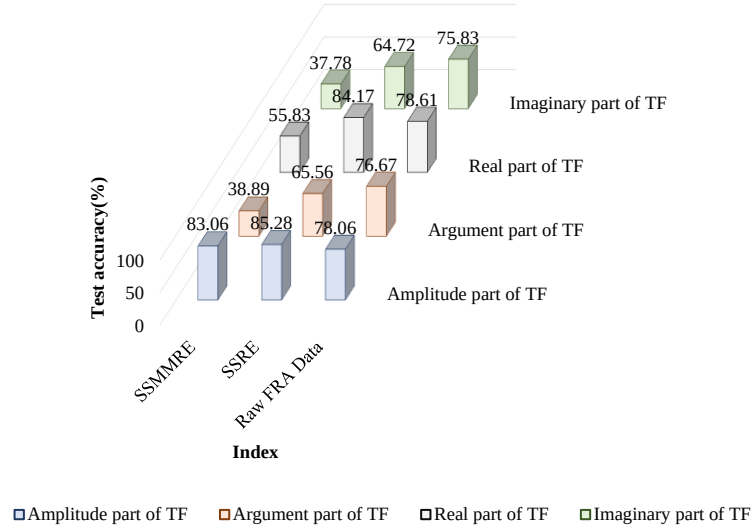


Figure 9. Evaluating the performance of the MLP model in the determination of the DSV fault severity using direct data of FRA and the extracted features by statistical indices.

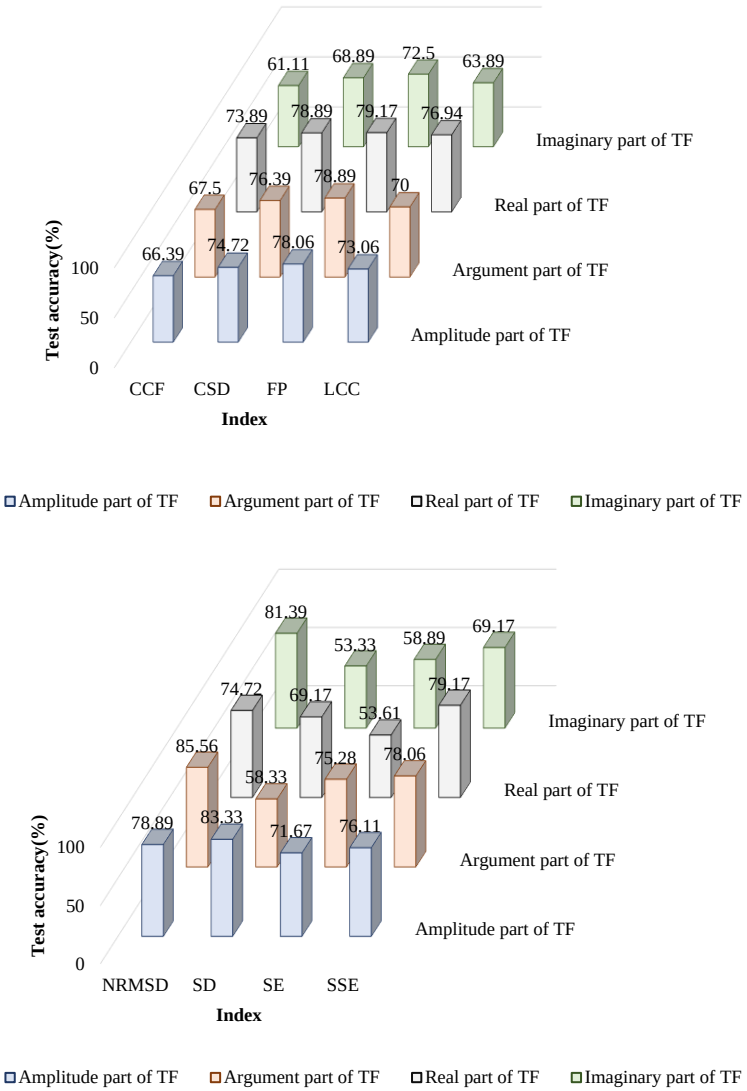


Figure 10. *Cont.*

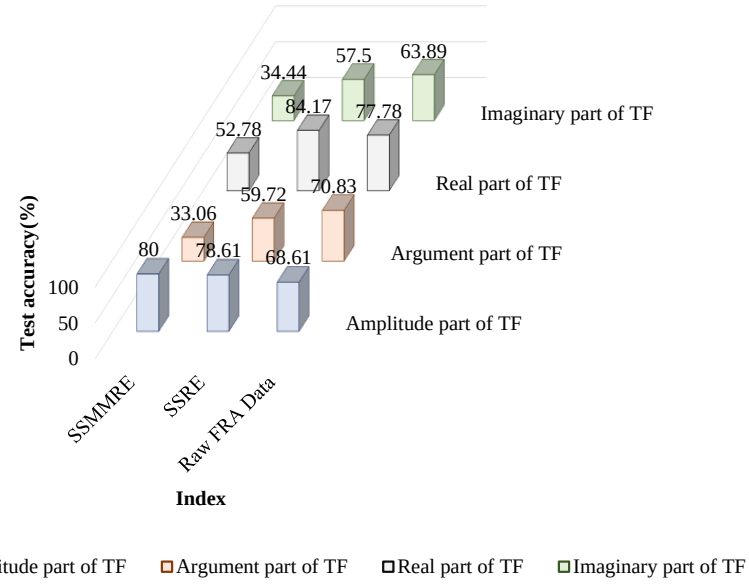


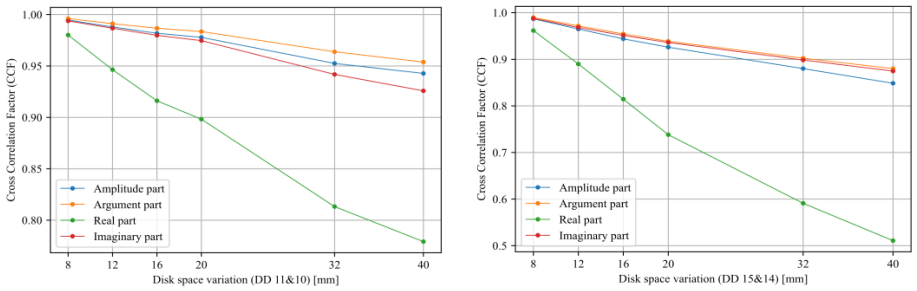
Figure 10. Evaluating the performance of the GMDH model in the determination of the DSV fault severity using direct data of FRA and the extracted features by statistical indices.

5. Discussion

The main aim of this paper is to localize the DSV fault and determine its severity in a DPT. For that reason, such a fault has been investigated with six severity in ten locations. Based on the amplitude and phase parts of the frequency responses, it is found that there are resonance and anti-resonance in 5–250, 250–500, 500–750, and 750–1000 kHz frequency sub-intervals. Six severity and ten locations of DSV faults are categorized into three severity and three locations. Using ten statistical indices, which are used in many research studies, four features are extracted from frequency sub-intervals. The extracted features have different sensitivities to the DSV faults using different parts of the TF data. The part of the TF for which each index has the most sensitivity is defined in Table 1. In this regard, the features are extracted from whole frequency intervals. Usually, the extracted features from the real part of the TF have the most sensitivity to the occurrence of the DSV fault. Feature variations based on the severity of the DSV in different locations for the considered indices are shown in Figures 11–20.

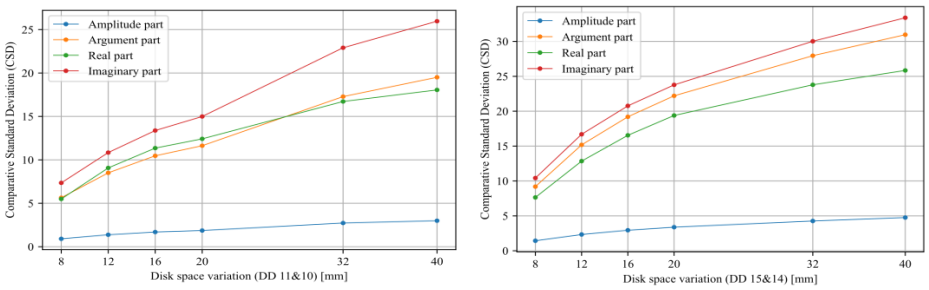
Table 1. The part of the TF whose features extracted from its data are most sensitive to the occurrence of the DSV fault.

Index	Part of the TF
CCF	real
CSD	imaginary
FP	real
LCC	real
NRMSD	real
SD	cannot be defined
SE	phase
SSE	imaginary
SSMMRE	real
SSRE	real



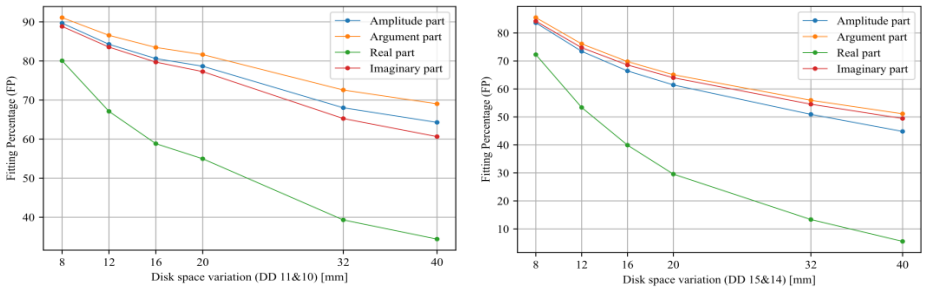
(**a**) (**b**)

Figure 11. Feature variations based on the severity of the DSV fault in different locations for extracted by the CCF index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



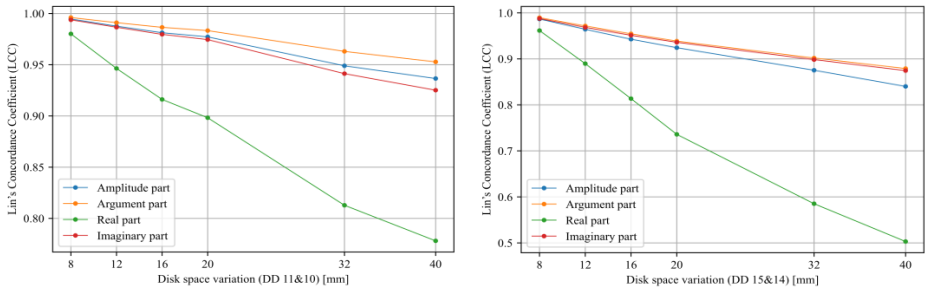
(**a**) (**b**)

Figure 12. Feature variations based on the severity of the DSV fault in different locations for extracted by the CSD index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



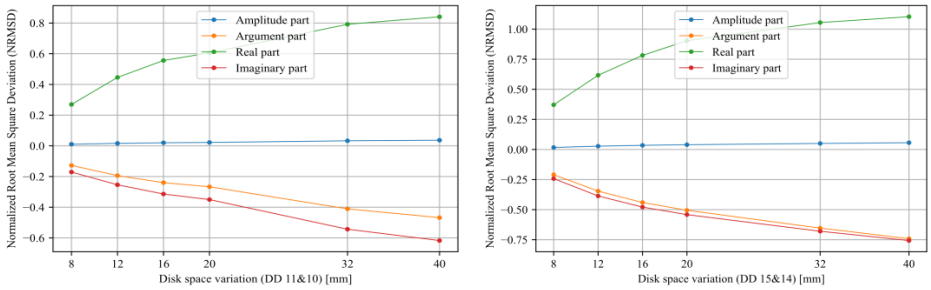
(**a**) (**b**)

Figure 13. Feature variations based on the severity of the DSV fault in different locations for extracted by the FP index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



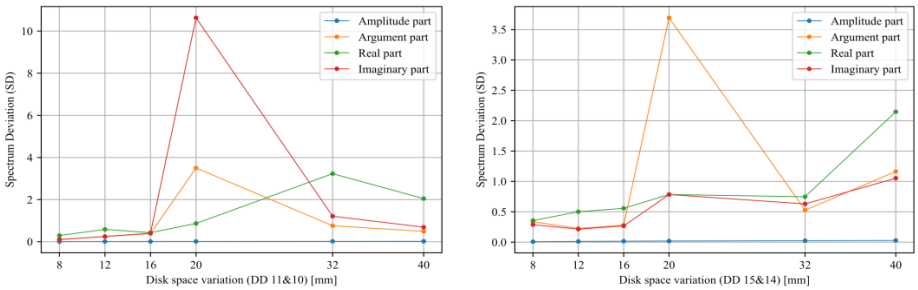
(**a**) (**b**)

Figure 14. Feature variations based on the severity of the DSV fault in different locations for extracted by the LCC index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



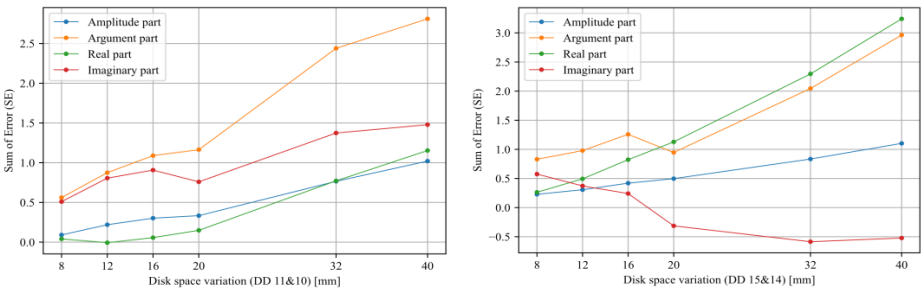
(**a**) (**b**)

Figure 15. Feature variations based on the severity of the DSV fault in different locations for extracted by the NRMSD index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



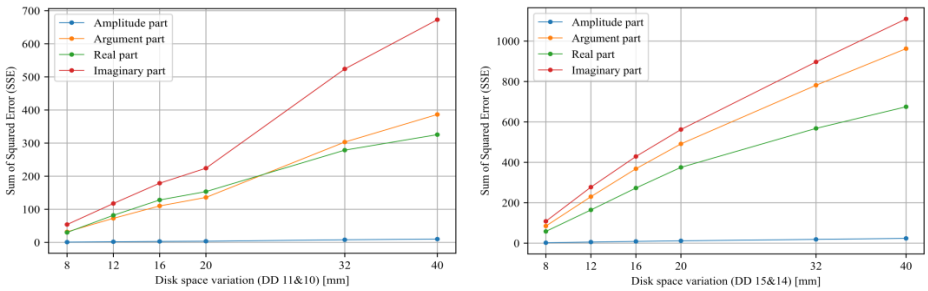
(**a**) (**b**)

Figure 16. Feature variations based on the severity of the DSV fault in different locations for extracted by the SD index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



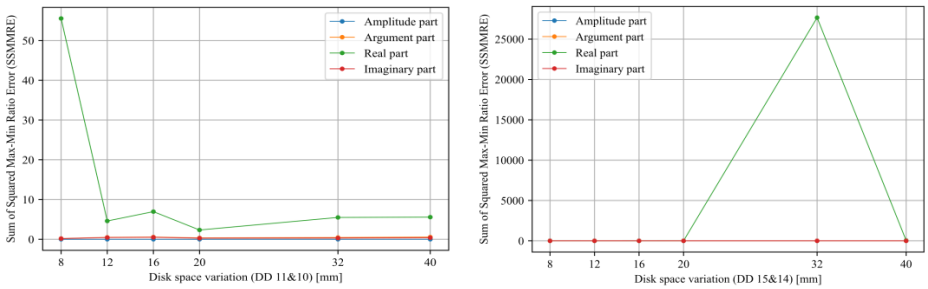
(**a**) (**b**)

Figure 17. Feature variations based on the severity of the DSV fault in different locations for extracted by the SE index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



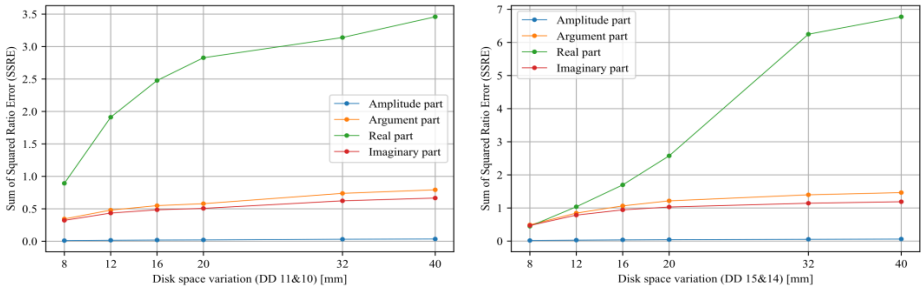
(**a**) (**b**)

Figure 18. Feature variations based on the severity of the DSV fault in different locations for extracted by the SSE index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



(**a**) (**b**)

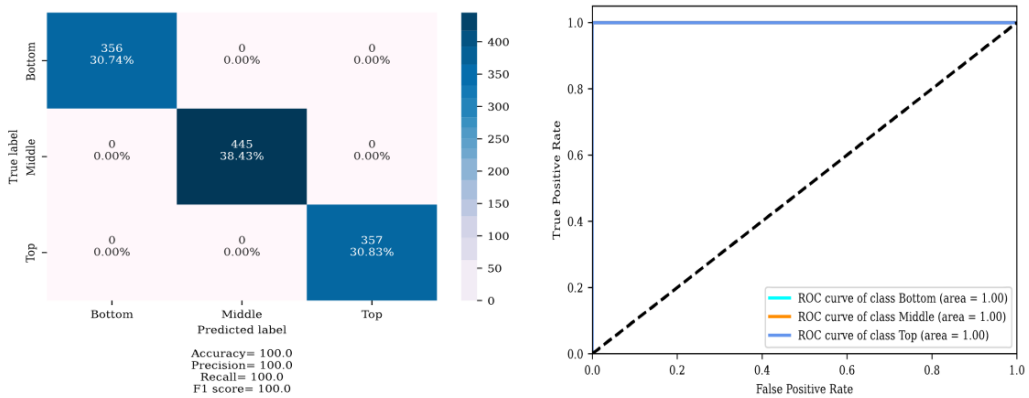
Figure 19. Feature variations based on the severity of the DSV fault in different locations for extracted by the SSMMRE index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.



(**a**) (**b**)

Figure 20. Feature variations based on the severity of the DSV fault in different locations for extracted by the SSRE index: (**a**) DSV in DD 11 and 10 (**b**) DSV in DD 15 and 14.

MLP and GMDH models have the maximum average accuracies in localization of the DSV fault when trained with the extracted features from the real part of the TF by the NRMSD index. Figures 21 and 22 show the training, validation, and test confusion matrices and ROC curves. The confusion matrix provides accuracy, precision, recall, and F1-score values. A box plot of the training, validation, and test accuracies in the thirty iterations is also provided.



(**a**) (**b**)

Figure 21. *Cont.*

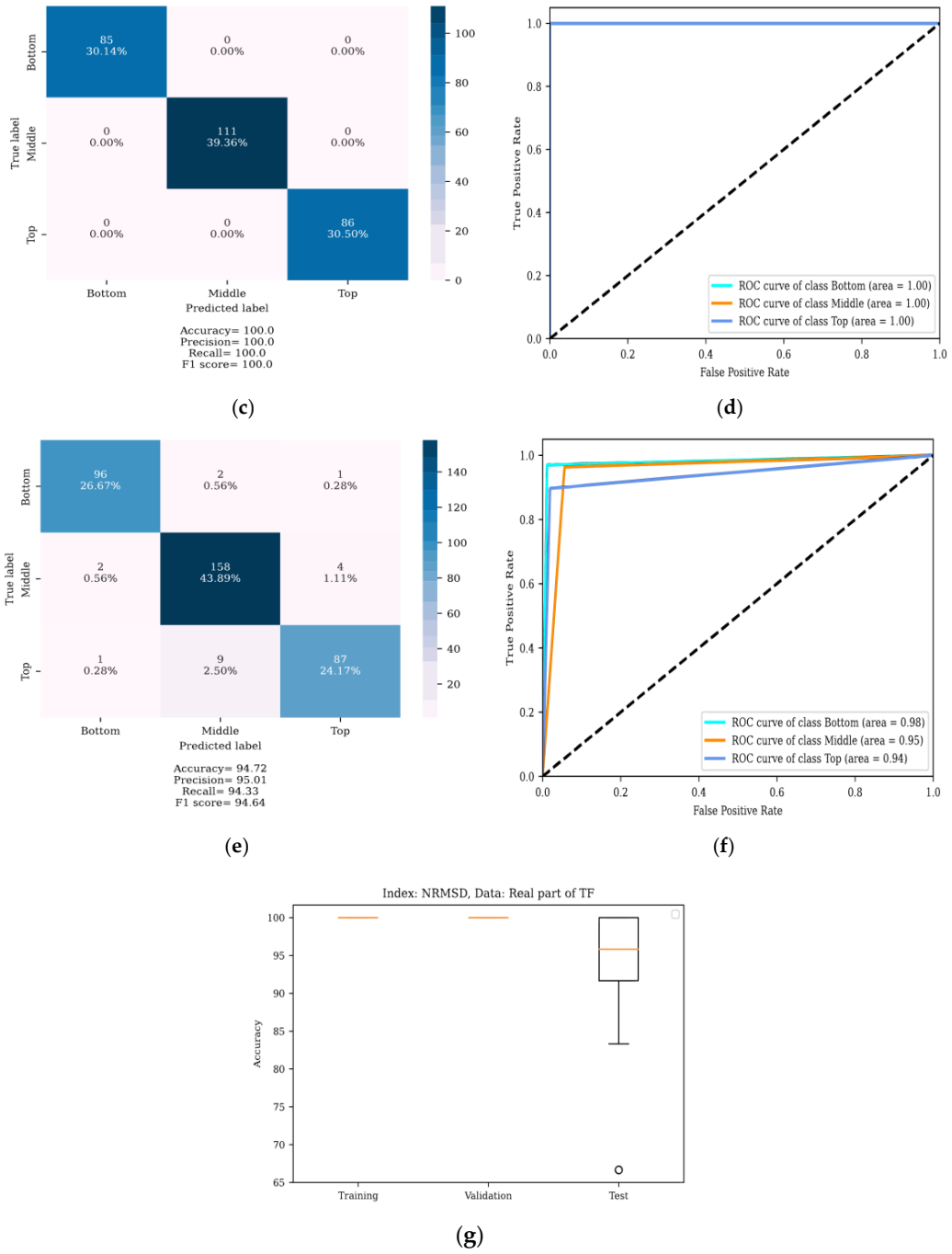


Figure 21. MLP performance in localization of the DSV fault using the extracted features from the real part of the TF using NRMSD index: (**a**) training confusion matrix; (**b**) training ROC curve; (**c**) validation confusion matrix; (**d**) validation ROC curve; (**e**) test confusion matrix; (**f**) test ROC curve; (**g**) box plot of the training, validation, and test accuracies in thirty iterations.

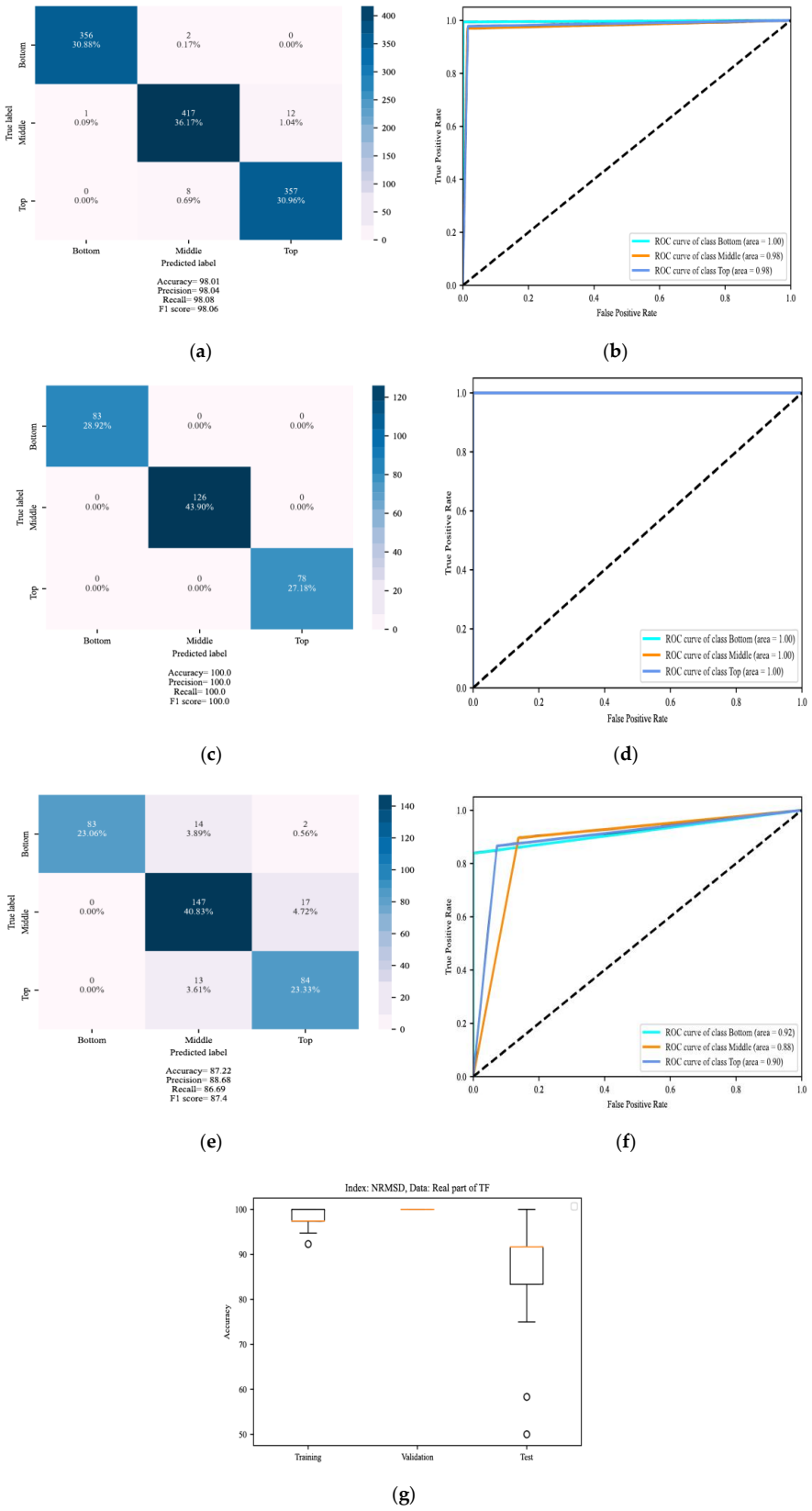


Figure 22. GMDH performance in localization of the DSV fault using the extracted features from the real part of the TF using NRMSD index: (**a**) training confusion matrix; (**b**) training ROC curve; (**c**) validation confusion matrix; (**d**) validation ROC curve; (**e**) test confusion matrix; (**f**) test ROC curve; (**g**) box plot of the training, validation, and test accuracies in thirty iterations.

The MLP model is trained with an average accuracy of 100%. The average validation and test accuracies are 100% and 94.72%, respectively. In the box plot of the test accuracies, the maximum value is 100%, and the minimum value is close to 80%. The lower and upper quartiles are close to 90% and 100%, respectively. The GMDH model is trained with an average accuracy of 98.01%. The average validation and test accuracies are 100% and 87.22%, respectively. In the box plot of the test accuracies, the maximum value is 100%, and the minimum value is close to 70%. The lower and upper quartiles are close to 80% and 90%, respectively. Therefore, both GMDH and MLP models can localize the DSV fault with a maximum test accuracy of 100%. In the box plot of test accuracies, minimum, lower, and upper quartile values of the MLP model are 10% above the GMDH results. It is shown that, if the MLP and GMDH models are trained with the phase part of the TF data, they can localize the DSV fault with high average test accuracies. Therefore, training, validation, and test confusion matrices and box plots when models are trained with the phase part of the TF are shown in Figures 23 and 24, respectively. It must be noted that the training process time is doubled, leading to higher average test accuracies.

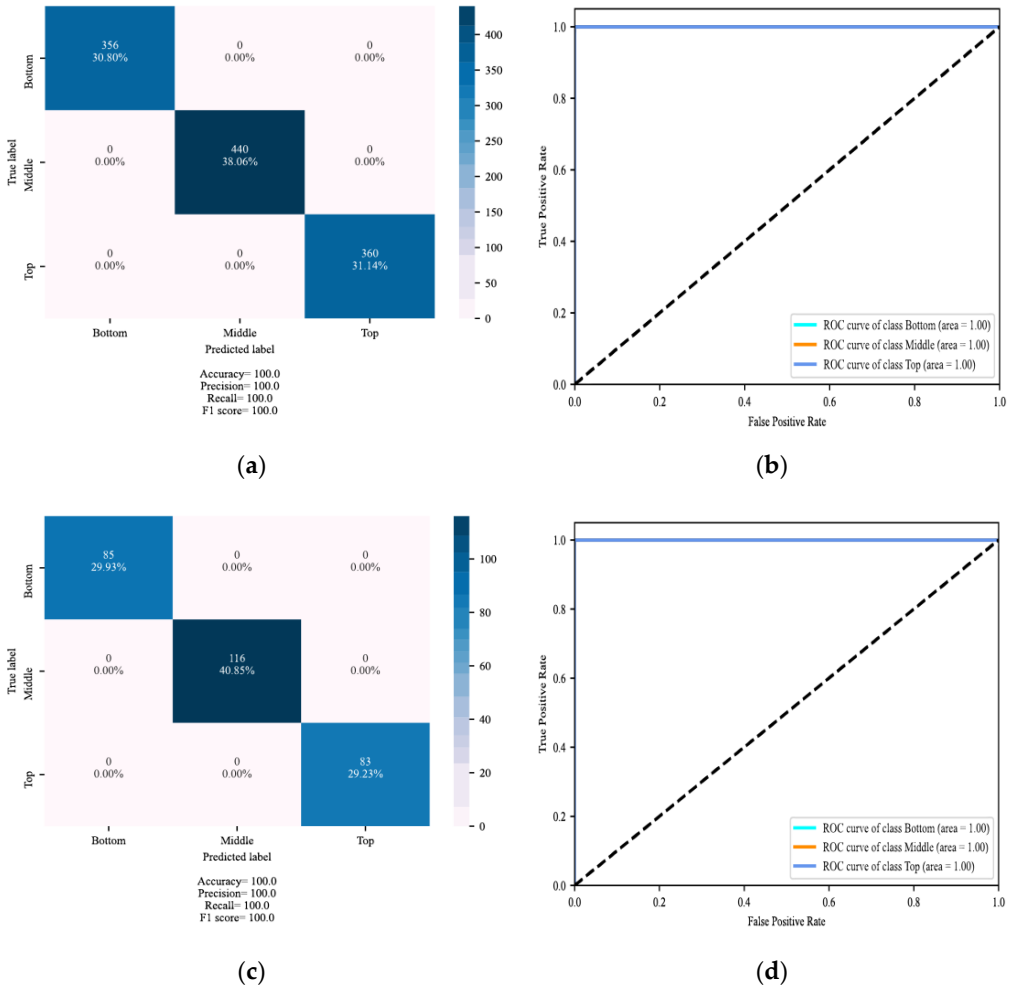


Figure 23. *Cont.*

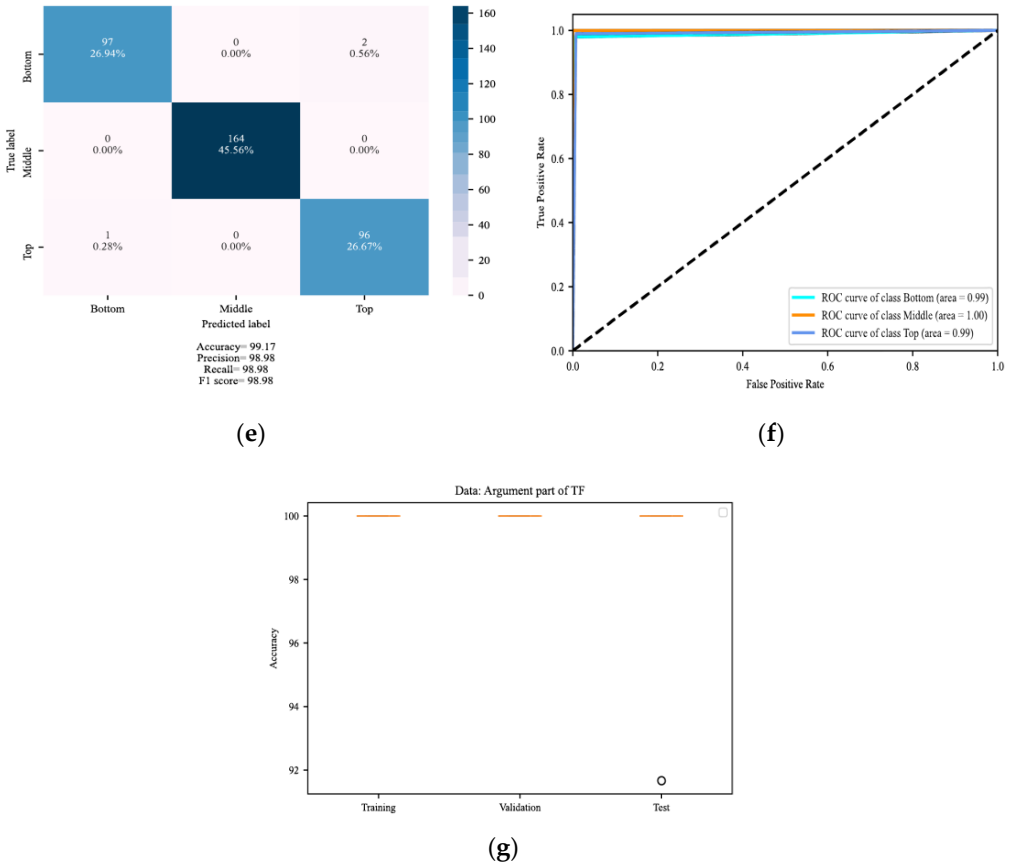


Figure 23. MLP performance in localization of the DSV fault using data of phase part of the TF: (**a**) training confusion matrix; (**b**) training ROC curve; (**c**) validation confusion matrix; (**d**) validation ROC curve; (**e**) test confusion matrix; (**f**) test ROC curve; (**g**) box plot of the training, validation, and test accuracies in thirty iterations.

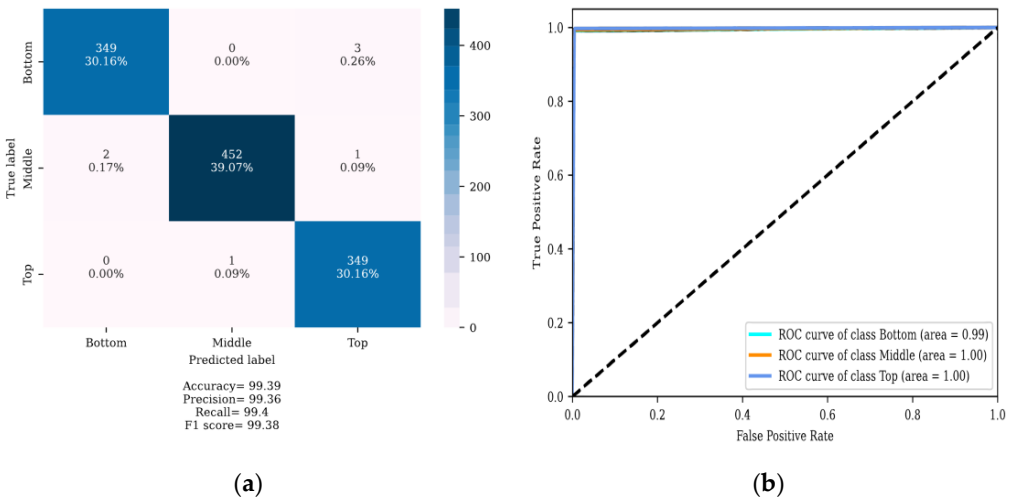


Figure 24. *Cont.*

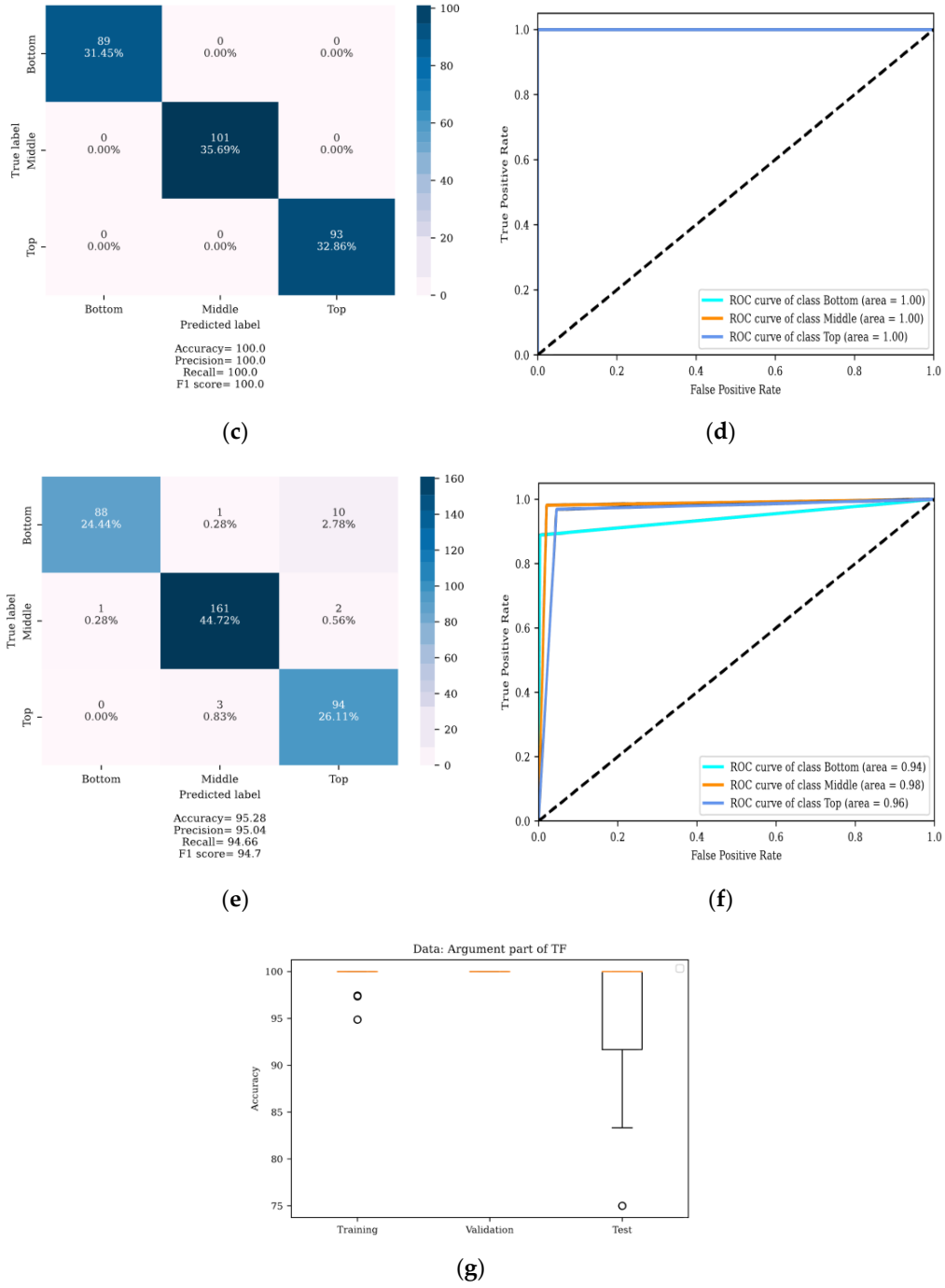


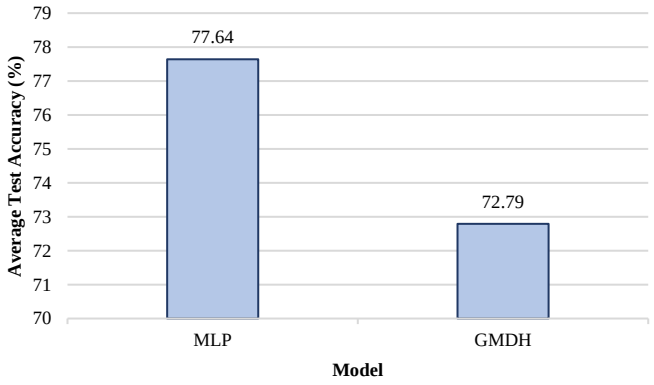
Figure 24. GMDH performance in localization of the DSV fault using data of phase part of the TF: (**a**) training confusion matrix; (**b**) training ROC curve; (**c**) validation confusion matrix; (**d**) validation ROC curve; (**e**) test confusion matrix; (**f**) test ROC curve; (**g**) box plot of the training, validation, and test accuracies in thirty iterations.

The results of training models with the data of the phase part of the TF reveal that the MLP model is trained with an average accuracy of 100%. The average validation and test accuracies are also 100% and 99.17%, respectively. The minimum test accuracy in the box plot of the test accuracies is close to 92%. The lower quartile, upper quartile, and maximum values are 100%. The GMDH model is trained with an average accuracy of 99.39%. The average validation and test accuracies are 100% and 95.28%, respectively. In the box plot of the test accuracies, the minimum test accuracy is close to 85%. The lower quartile is above 90%, and the upper quartile and maximum values are 100%. Therefore, both GMDH and

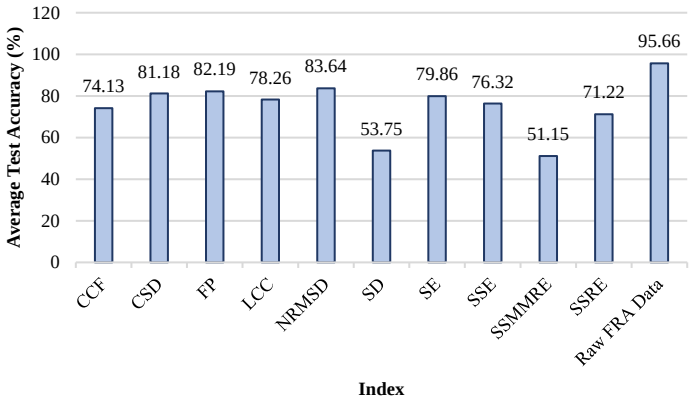
MLP models can localize the DSV fault with a maximum test accuracy of 100%. In the box plot of test accuracies, it is evident that the GMDH model can track MLP results very well in localization of the DSV fault.

It is also found that that, if the NRMSD index is used to extract features from data of the real part of the TF and models are trained with the features, the location of the DSV fault is determined with a higher accuracy. Additionally, in the box plots, it is shown that, if the data of the phase part of the TF is used to train models, the minimum lower quartile of test accuracies increases, but the time of the training process is doubled. The GMDH localize the fault with high accuracies near the MLP results with no need for parameter tuning. Therefore, GMDH can reach the MLP results faster.

Figures 7 and 8 show the extracted features from data of different parts of the TF using different NIns. The MLP and GMDH models are trained with such features and localize the DSV fault with varying test accuracies. The best results are chosen and their confusion matrices, ROC curves, and box plots are shown. The test accuracies for each model, data of part of the TF, and index are also averaged. The results are shown in Figure 25. It is evident that the average accuracy of the MLP and GMDH models in localization of DSV fault is close to 75%. The average performances of models when they are trained with the extracted features by the NRMSD index and raw FRA data are 83.64% and 95.66%, which is better than using other NIns. In addition, considering the extracted features from the data of the amplitude part of the TF, the average performance of models is 83.32%, which is higher than using the data of other parts of the TF.

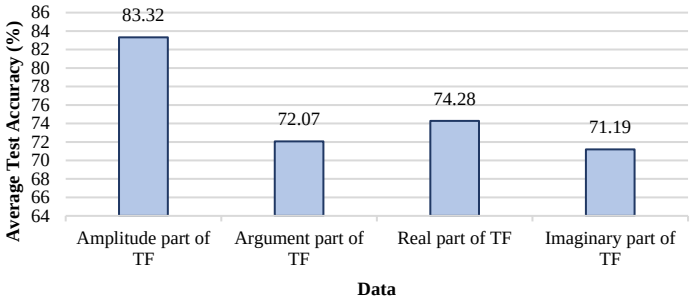


(**a**)



(**b**)

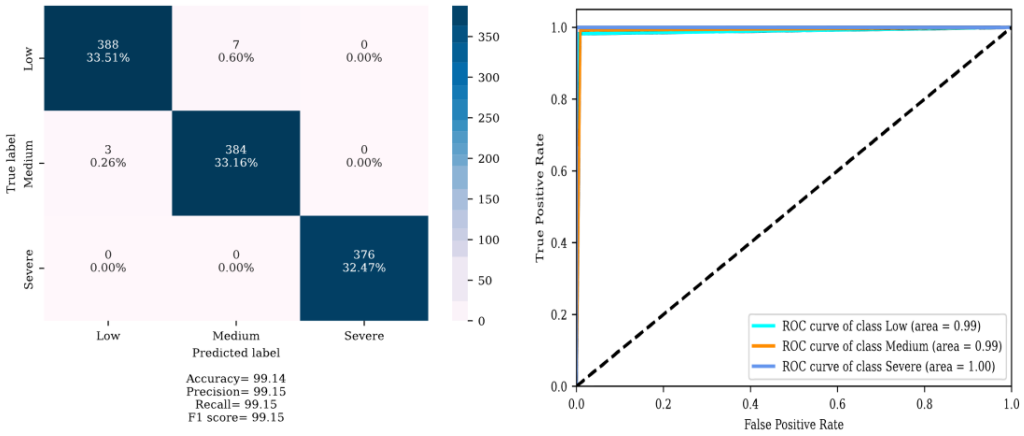
Figure 25. *Cont.*



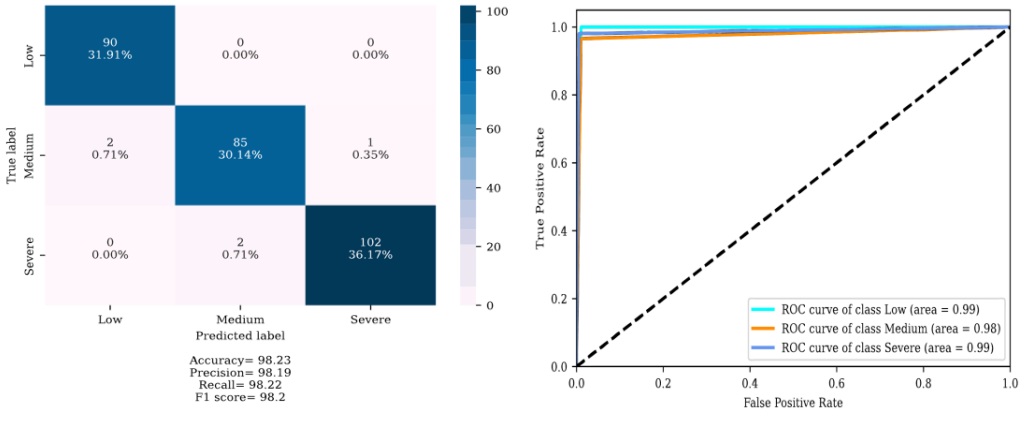
(**c**)

Figure 25. Average performance of models, using indices and parts of the TF in localization of the DSV fault: (**a**) average performance of the GMDH and MLP models; (**b**) average performance when models are trained with the extracted features by different indices; (**c**) average performance of models when the extracted features are from different parts of the TF.

MLP and GMDH models have the maximum average accuracies in determining the severity of the DSV fault when trained with the extracted features from the phase part of the TF by the NRMSD index. Figures 26 and 27 demonstrate the training, validation, and test confusion matrices and ROC curves. A box plot of the training, validation, and accuracies in the thirty iterations is also provided.



(**a**) (**b**)



(**c**) (**d**)

Figure 26. *Cont.*

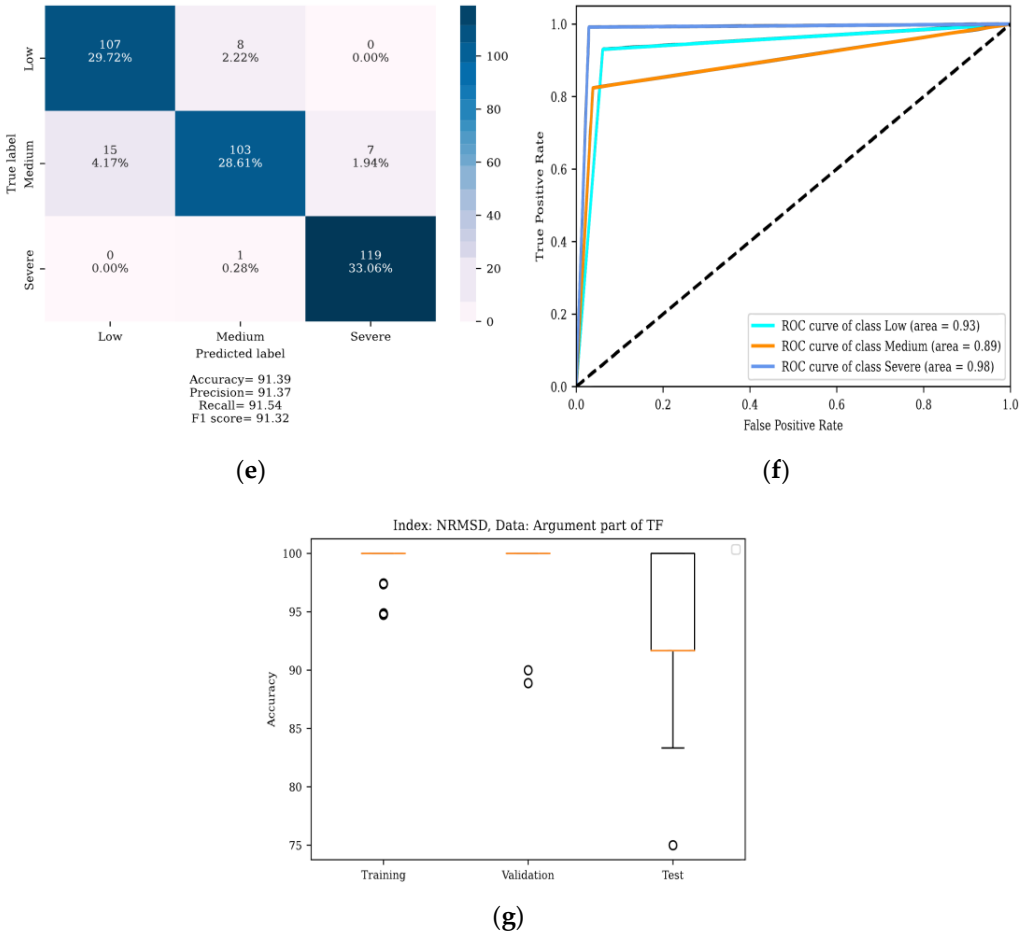


Figure 26. MLP performance in determining the severity of the DSV fault using the extracted features from the phase part of the TF using the NRMSD index: (**a**) training confusion matrix; (**b**) training ROC curve; (**c**) validation confusion matrix; (**d**) validation ROC curve; (**e**) test confusion matrix; (**f**) test ROC curve; (**g**) box plot of the training, validation, and test accuracies in thirty iterations.

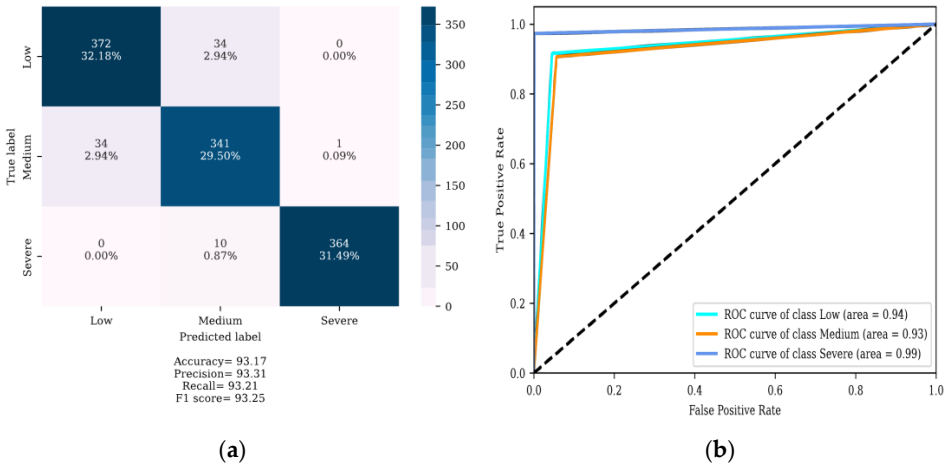
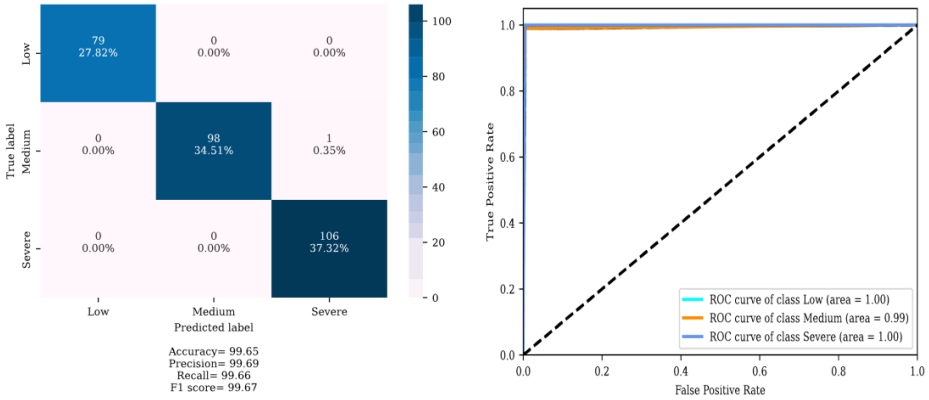
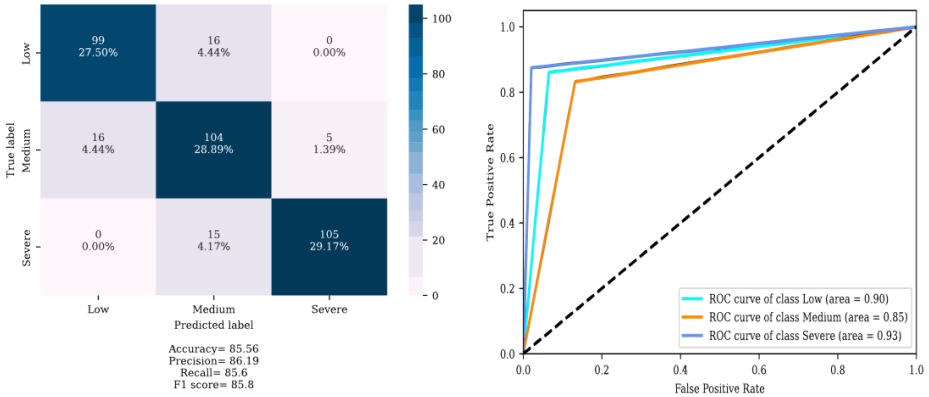


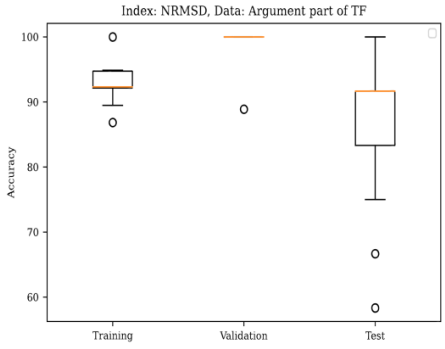
Figure 27. *Cont.*



(**c**) (**d**)



(**e**) (**f**)

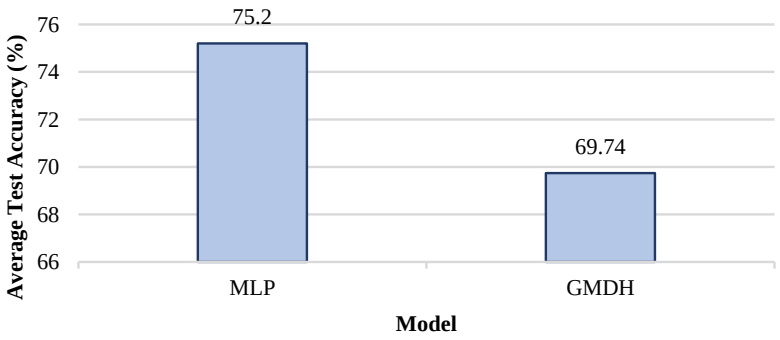


(**g**)

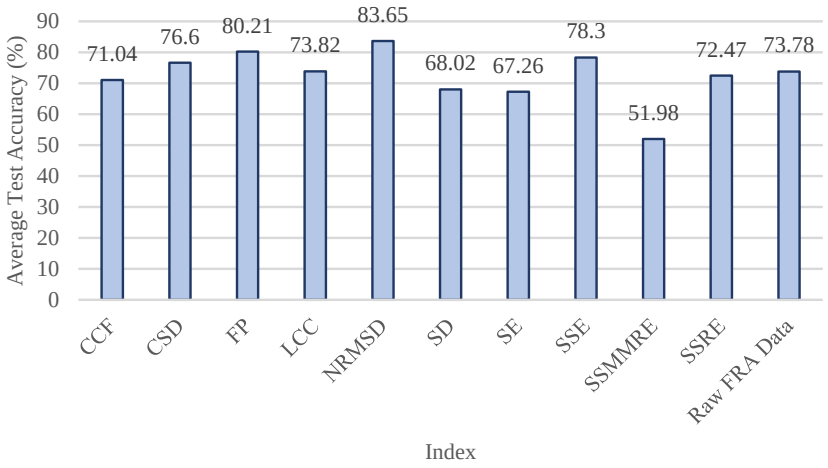
Figure 27. GMDH performance in determining the severity of the DSV fault using the extracted features from the phase part of the TF using NRMSD index: (**a**) training confusion matrix; (**b**) training ROC curve; (**c**) validation confusion matrix; (**d**) validation ROC curve; (**e**) test confusion matrix; (**f**) test ROC curve; (**g**) box plot of the training, validation, and test accuracies in thirty iterations With regard to determining DSV severity, the MLP model is trained with an average accuracy of 99.14%. The average validation and test accuracies are 98.23% and 91.39%, respectively. In the box plot of the test accuracies, the maximum value is 100%, and the minimum value is close to 85%. The lower and upper quartiles are close to 90% and 100%, respectively. The GMDH model is trained with an average accuracy of 93.17%. The average validation and test accuracies are 99.65% and 85.56%, respectively. In the box plot of the test accuracies, the maximum value is 100%, and the minimum value is close to 75%. The lower and upper quartiles are close to 80% and 90%, respectively. Therefore, both GMDH and MLP models can determine the severity of the DSV fault with a maximum test accuracy of 100%. In the box plot of test accuracies, minimum, lower, and upper quartile values of the MLP model are 10% above the GMDH results.

It is also noted that, if the NRMSD index is used to extract the features from data of the phase part of the TF and models are trained with the obtained features, the severity of the DSV fault is determined with high accuracy. The GMDH determines the severity of the fault with high accuracies near the MLP result with no need for parameter tuning. Therefore, GMDH can reach the MLP results faster. The accuracy of models in localization of the DSV fault is higher than determining the severity.

Figures 9 and 10 show the extracted features from the data of different TF parts using different NIns. The MLP and GMDH models are trained with these features and determine the severity of the DSV fault with varying test accuracies. The best results are selected and their confusion matrices, ROC curves, and box plots are shown. The test accuracies for each model, the data of part of the TF, and index are all averaged. The results are shown in Figure 28. The results reveal that the average accuracy of the MLP and GMDH models in determining the severity of DSV fault is close to 75% and 70%, respectively. The average performance of models when they are trained with the extracted features by the NRMSD index is 83.65%, which is better than using other NIns. Furthermore, when the extracted features from the data of amplitude part of the TF, the average performance of models is 78.3%, which is higher than using the data of other parts of the TF.

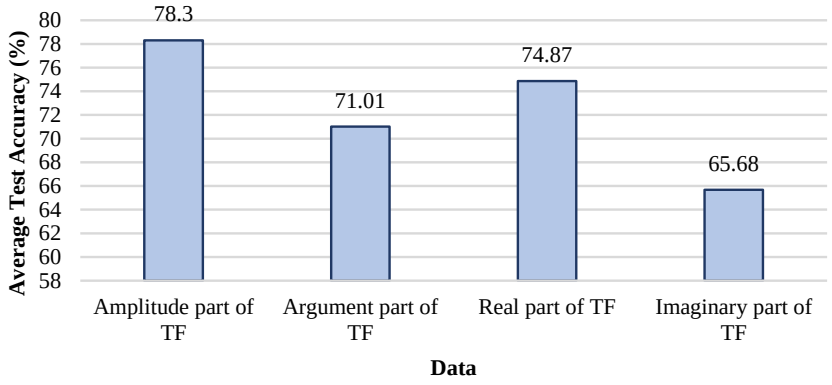


(**a**)



(**b**)

Figure 28. *Cont.*



(**c**)

Figure 28. Average performance of models, using indices and parts of the TF in determining the severity of the DSV fault: (**a**) average performance of the GMDH and MLP models; (**b**) average performance when models are trained when the extracted features are from different indices; (**c**) average performance of models when the extracted features are from different parts of the TF.

Some additional information for further consideration of the model is provided in Table 2. The average time of training features and raw FRA data to determine the severity of the DSV fault is less than localization. In addition, tuning the hyper-parameters of the MLP is more time-consuming than the tuning of the hyper-parameters of the GMDH. The stop factor of training condition (epsilon value) for the GMDH is set to 0.001. The MLP model has two hidden layers with six and four neurons, respectively. The activation function of neurons is the tangent hyperbolic. The learning rate, momentum, and decay values of the SGD optimizer are set to 0.015, 0.9, and 0.005, respectively.

Table 2. Average time of training models with extracted features and raw FRA data.

Models	Determination of Location OR Severity	Average Time of Training Process Using NIns	Average Time of Training Process Using Raw FRA Data
GMDH	Location	19.56 s	50.62 s
	Severity	7.9 s	29.62 s
MLP	Location	11.22 s	28.89 s
	Severity	9.85 s	11.02 s

It should be noted that, while comparing other machine learning methods, such as SVM utilized to identify DSV faults in [26], with the results of ANNs in this paper, the MLP outperforms. It can be concluded that utilizing the NRMSD indicator to feed MLP results in the highest accuracy values in localizing and detecting the intensity of DSV faults.

6. Conclusions

Data analysis is a vital step in monitoring the condition of electric power components in modern power systems. Distribution Power Transformers (DPTs) installed in modern networks play a prominent role in delivering electricity to the customers; therefore, their health assessment should be considered. The Frequency Response Analysis (FRA) method, which is also known as the Transfer Function (TF) technique, is a widely accepted technique in the maintenance and monitoring of DPTs. Nevertheless, the operators cannot access reliable and precise codes to interpret the TF measurements and test results. In this paper, the application of Group Method of Data Handling (GMDH)-based intelligent interpreter of the TF test results is investigated. In addition, sensitivity analysis of all

TF parts (real, phase, imaginary, and amplitude components) considering ten suitable Numerical Indicators (NIns), i.e., Spectrum Deviation (SD), Cross Correlation Factor (CCF), Comparative Standard Deviation (CSD), Lin's Concordance Coefficient (LCC), Fitting Percentage (FP), Normalized Root Mean Square Deviation (NRMSD), Sum of Error (SE), Sum of Squared Error (SSE), Sum of Squared Max-Min Ratio Error (SSMMRE), and Sum of Squared Ratio Error (SSRE) indices, is conducted. In the preprocessing stage of the TF datasets, a feature extraction algorithm that applies NIns to four frequency sub-bands has been proposed, leading to saving time in determining the severity and location of Disk-Space Variation (DSV) deformations. On average, using the NRMSD index, in comparison to the other NIns, leads to having high accuracy of 84.64% in determining the DSV location. Based on the average test accuracy of the intelligent interpreters utilizing datasets of various TF components, the amplitude part of the TF with 83.32% accuracy on average has resulted in obtaining the highest accuracy. In determining the severity of DSV deformations, utilizing the NRMSD index also made the intelligent identifiers interpret the TF results with the highest average accuracy of 83.65% considering all ten NIns. It should be noted that models are trained with extracted features, and all locations are considered, and all extracted features are simultaneously taken into account. It is noticed that there is no linear relationship between extracted features and DSV fault intensity or location. As a result, the role of Neural Networks becomes essential. The MLP and GMDH models can find such a nonlinear relationship accurately with the use of the NRMSD index. Moreover, MLP and GMDH outperform the other machine learning methods, such as SVM, in localizing and identifying the severity of DSV faults. Using the amplitude part of the TF results as input to identify DSV severity leads to outperforming the other parts with an average test accuracy of 78.3%. Finally, in order to clarify the future scope of this research study, the following items can be considered:

- ➢ Using AI-based interpreters to identify simultaneous winding faults, such as detecting the simultaneous occurrence of RD and DSV faults with various intensities and locations;
- ➢ Considering the effects of adjacent substation devices, such as Current Transformers (CTs), Potential Transformers (PTs), and other phases of a three-phase transformer on the phase in which the FRA test is performed for online monitoring.

Author Contributions: O.E.: Software, Methodology, Writing—original draft preparation, Investigation, Visualization; R.B.: Conceptualization, Software, Methodology, Writing—original draft preparation, Investigation, Resources; G.B.G.: Validation, Supervision, Project administration, Methodology, Data curation; F.M.: Formal analysis, Investigation, Validation, Supervision, Project administration, Methodology, Data curation, Writing—original draft preparation, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

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

Nomenclature

SD	Spectrum Deviation
CCF	Cross Correlation Factor
CSD	Comparative Standard Deviation
LCC	Lin's Concordance Coefficient
FP	Fitting Percentage
NRMSD	Normalized Root Mean Square Deviation
SE	Sum of Error

SSE	Sum of Squared Error
SSMMRE	Sum of Squared Max-Min Ratio Error
SSRE	Sum of Squared Ratio Error
DSV	Disk-Space Variation
DD	Double-Disk
GMDH	Group Method of Data Handling
NIns	Numerical Indicators
DPT	Distribution Power Transformer
FRA	Frequency Response Analysis
TF	Transfer Function
ANN	Artificial Neural Network
SVM	Support Vector Machine

References

1. Wei, X.; Gao, S.; Huang, T. Analysis of electrical network vulnerability using segmented cascading faults graph. *Comput. Electr. Eng.* **2019**, *81*, 106519. [CrossRef]
2. Beura, C.P.; Beltle, M.; Wenger, P.; Tenbohlen, S. Experimental Analysis of Ultra-High-Frequency Signal Propagation Paths in Power Transformers. *Energies* **2022**, *15*, 2766. [CrossRef]
3. Miyazaki, S. Detection of Winding Axial Displacement of a Real Transformer by Frequency Response Analysis without Fingerprint Data. *Energies* **2021**, *15*, 200. [CrossRef]
4. Mohammadi, F.; Nazri, G.A.; Saif, M. A Fast Fault Detection and Identification Approach in Power Distribution Systems. In Proceedings of the IEEE 5th International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), Istanbul, Turkey, 26–27 August 2019.
5. Yang, Q.; Su, P.; Chen, Y. Comparison of Impulse Wave and Sweep Frequency Response Analysis Methods for Diagnosis of Transformer Winding Faults. *Energies* **2017**, *10*, 431. [CrossRef]
6. Yoon, Y.; Son, Y.; Cho, J.; Jang, S.; Kim, Y.-G.; Choi, S. High-Frequency Modeling of a Three-Winding Power Transformer Using Sweep Frequency Response Analysis. *Energies* **2021**, *14*, 4009. [CrossRef]
7. Tahir, M.; Tenbohlen, S. A Comprehensive Analysis of Windings Electrical and Mechanical Faults Using a High-Frequency Model. *Energies* **2019**, *13*, 105. [CrossRef]
8. Bigdeli, M.; Abu-Siada, A. Clustering of transformer condition using frequency response analysis based on k-means and GOA. *Electr. Power Syst. Res.* **2021**, *202*, 107619. [CrossRef]
9. Thango, B.A.; Nnachi, A.F.; Dlamini, G.A.; Bokoro, P.N. A Novel Approach to Assess Power Transformer Winding Conditions Using Regression Analysis and Frequency Response Measurements. *Energies* **2022**, *15*, 2335. [CrossRef]
10. Banaszak, S.; Kornatowski, E.; Szoka, W. The Influence of the Window Width on FRA Assessment with Numerical Indices. *Energies* **2021**, *14*, 362. [CrossRef]
11. Tahir, M.; Tenbholen, S.; Miyazaki, S. Analysis of Statistical Methods for Assessment of Power Transformer Frequency Response Measurements. *IEEE Trans. Power Deliv.* **2020**, *36*, 618–626. [CrossRef]
12. Lu, S.; Gao, W.; Hong, C.; Sun, Y. A newly-designed fault diagnostic method for transformers via improved empirical wavelet transform and kernel extreme learning machine. *Adv. Eng. Inform.* **2021**, *49*, 101320. [CrossRef]
13. Behkam, R.; Moradzadeh, A.; Karami, H.; Nadery, M.S.; Mohammadi-Ivatloo, B.; Gharehpetian, G.B.; Tenbohlen, S. Mechanical Fault Types Detection in Transformer Windings Using Interpretation of Frequency Responses via Multilayer Perceptron. *J. Oper. Autom. Power Eng.* **2022**. [CrossRef]
14. Zhao, Z.; Tang, C.; Zhou, Q.; Xu, L.; Gui, Y.; Yao, C. Identification of Power Transformer Winding Mechanical Fault Types Based on Online IFRA by Support Vector Machine. *Energies* **2017**, *10*, 2022. [CrossRef]
15. Li, Z.; Zhang, Y.; Abu-Siada, A.; Chen, X.; Li, Z.; Xu, Y.; Zhang, L.; Tong, Y. Fault Diagnosis of Transformer Windings Based on Decision Tree and Fully Connected Neural Network. *Energies* **2021**, *14*, 1531. [CrossRef]
16. Behkam, R.; Karami, H.; Naderi, M.S.; Gharehpetian, G.B. Condition Monitoring of Distribution Transformers Using Frequency Response Traces and Artificial Neural Network to Detect the Extent of Windings Axial Displacements. In Proceedings of the 26th International Electrical Power Distribution Conference (EPDC), Tehran, Iran, 11–12 May 2022; pp. 18–23. [CrossRef]
17. Mohammadi, F.; Zheng, C. A precise SVM classification model for predictions with missing data. In Proceedings of the 4th National Conference on Applied Research in Electrical, Mechanical Computer and IT Engineering, Tehran, Iran, 4 October 2018.
18. Mohammadi, F.; Zheng, C.; Su, R. Fault Diagnosis in Smart Grid Based on Data-Driven Computational Methods. In Proceedings of the 5th International Conference on Applied Research in Electrical, Mechanical, and Mechatronics Engineering, Tehran, Iran, 24 January 2019.
19. Tahir, M.; Tenbohlen, S. Transformer Winding Condition Assessment Using Feedforward Artificial Neural Network and Frequency Response Measurements. *Energies* **2021**, *14*, 3227. [CrossRef]
20. Moradzadeh, A.; Pourhossein, K.; Mohammadi-Ivatloo, B.; Mohammadi, F. Locating Inter-Turn Faults in Transformer Windings Using Isometric Feature Mapping of Frequency Response Traces. *IEEE Trans. Ind. Inform.* **2020**, *17*, 6962–6970. [CrossRef]

21. MolaAbasi, H.; Khajeh, A.; Chenari, R.J. Use of GMDH-type neural network to model the mechanical behavior of a cement-treated sand. *Neural Comput. Appl.* **2021**, *33*, 15305–15318. [CrossRef]
22. Cremer, J.L.; Strbac, G. A machine-learning based probabilistic perspective on dynamic security assessment. *Int. J. Electr. Power Energy Syst.* **2021**, *128*, 106571. [CrossRef]
23. Mikheev, M.Y.; Gusynina, Y.S.; Shornikova, T.A. Building Neural Network for Pattern Recognition. In Proceedings of the 2020 International Russian Automation Conference (RusAutoCon 2020), Sochi, Russia, 6–12 September 2020; pp. 357–361. [CrossRef]
24. Sesmero, M.P.; Alonso-Weber, J.M.; Sanchis, A. CCE: An ensemble architecture based on coupled ANN for solving multiclass problems. *Inf. Fusion* **2019**, *58*, 132–152. [CrossRef]
25. Onwubolu, G. *GMDH-Methodology and Implementation in MATLAB*; Perlego Ltd.: London, UK, 2016.
26. Tarimoradi, H.; Karami, H.; Gharehpetian, G.B.; Tenbohlen, S. Sensitivity analysis of different components of transfer function for detection and classification of type, location and extent of transformer faults. *Measurement* **2021**, *187*, 110292. [CrossRef]