Electrical Power and Energy Systems 156 (2024) 109779

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Machine learning for power transformer SFRA based fault detection

Miloš Bjelić, Bogdan Brković, Mileta Žarković *, Tatjana Miljković

University of Belgrade, School of Electrical Engineering, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia

ARTICLE INFO

Keywords:
Power transformer
Machine learning
SFRA
ANN
ANFIS

ABSTRACT

This paper presents machine learning methods for health assessment of power transformer based on sweep frequency response analysis. The paper presents an overview of monitoring and diagnostics based on statistical Sweep Frequency Response Analysis (SFRA) based indicators that are used to evaluate the state of the power transformer. Experimental data obtained from power transformers with internal short-circuit faults is used as a database for applying machine learning. Machine learning is implemented to achieve more precise asset management and condition-based maintenance. Unsupervised machine learning was applied through the *k*-means cluster method for classifying and dividing the examined power transformer state into groups with similar state and probability of failure. Artificial neural network (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS) as part of supervised machine learning are created in order to detect fault severity in tested power transformers of different lifetime. The presented machine learning methods can be used to improve health assessment of power transformers.

1. Introduction

The power transformer (PT) is one of the most significant and most expensive components in power plants, transmission, and distribution of electrical energy. The reliability of PT operation and its aging are affected by thermal, mechanical, electrical, and chemical processes. Therefore, the PT is a complex element of the power system whose monitoring and diagnostics require special attention to prevent power supply interruptions. Great efforts are being made to clearly define the methods that would be used to switch from Time Based Maintenance (TBM) to Condition Based Maintenance (CBM) that should result in correct and timely decisions regarding PT maintenance and preventing power outages and more serious and costly problems in the power system. For this reason, research in the field of monitoring and diagnostics is perpetual, and one of the monitoring methods that generate many outputs numerical values that can be used by machine learning algorithms is SFRA. SFRA method was introduced by Dick and Erven [1]. SFRA involves measurements performed on transformer windings excited by a sinusoidal voltage source over a wide frequency range. The output of the SFRA is the winding transfer function, which represents the ratio of the output and input signal in the frequency domain. Interpretation of SFRA results is based on comparing measured transfer function with reference measurements. Reference measurements can be made on the same PT before exploitation has commenced, or on another transformer of the same design. However, the interpretation involves a great amount of subjectivity in the comparison, and it is very difficult to make a decision about the prevalence (severity) of the defect. Therefore, this paper presents an overview of all statistical indicators that can be used for clearer diagnostics and open possibilities for machine learning application [2]. The SFRA methodology is well explained in IEEE standards [3] and IEC standard [4]. Fundamentals of SFRA implementation and experience in interpreting the results are presented in the CIGRE reports [5,6]. High-frequency electric circuit model for transformer frequency response analysis is well explained in [7]. Common statistical fault indicators for SFRA interpretation are calculated for different types of short-circuit internal faults in [8]. Papers [9,10] use different methodologies to detect winding deformation, external and internal short-circuit faults. The sensitivity of numerical SFRA–based indicators against different types of faults is analyzed in [11]. Several indicators were shown to exhibit high sensitivity to most types of mechanical faults, however, the issue of defining threshold values remains. Databases generated from experiments can be used in combination with artificial intelligence (AI) based approaches to detect different types of faults [12–21]. Spectral clustering algorithms have been proposed for different winding deformations detection [12]. Paper [13] introduces AI through Decision Tree algorithm and ANN with extraction of differential currents to detect the winding radial deformation and axial displacement. Support Vector Machine (SVM) can be used for identification of

* Corresponding author.
E-mail address: mileta@etf.bg.ac.rs (M. Žarković).

https://doi.org/10.1016/j.ijepes.2023.109779
Received 12 July 2023; Received in revised form 23 December 2023; Accepted 31 December 2023

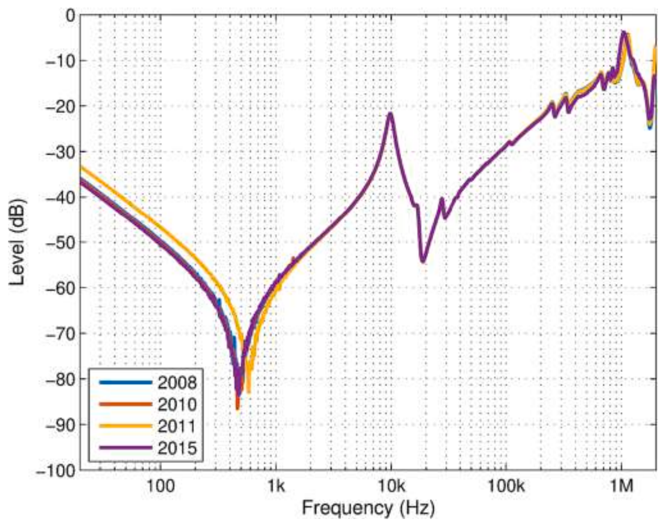


Fig. 1. Frequency responses obtained for power transformers of different age.

transformer winding faults through transfer function analysis [14,15]. Research [18] use ANN to detect the extent of windings axial displacements on the 20 kV winding of a 1600 kVA distribution transformer. The approach presented in this paper uses four inputs for ANN: concordance coefficient (LCC), crosscorrelation factor (CCF), sum of errors (SE), and fitting percentage (FP) as most effective and sensitive indexes. Authors in [19] diagnosing accurate disk space variation locations and fault extents by ANN and numerical index called fitting percentage (FP). Paper [20] uses ANN along with novel statistical indicator FP to detect radial deformation, disk space variation, axial displacement, and short circuit. Authors in [21] use ANN classifier to discriminate between mechanical defects and electrical faults based on EMTP/ATP simulations.

Interpretation of SFRA results is usually performed purely visually, often without applying signal processing to calculate statistical indicators. There is a wide array of statistical indicators that show different behavior depending on the type of fault [8]. Interpretation of these indicators can be ambiguous from the diagnostics standpoint. Paper [22] presents a review of condition monitoring and diagnostic for lifetime estimation of PTs and indicates need to make a correlation between measured parameters of monitoring and aging. Application of machine learning is motivated by the need to achieve more precise interpretation of SFRA results and to correlate fault development with the aging of the PT. The main advantage of applying machine learning is that multiple indicators are observed and processed simultaneously to produce a more reliable and informative conclusion. Results of such a methodology is expected to improve health assessment and help engineers make proper and timely decisions regarding the maintenance and reparation of the PTs. This paper summarizes all indicators obtained from SFRA tests along with novel indicator reverberation time presented in [8]. Until now, such indicators have not been calculated using SFRA results obtained for transformers of different ages and under different operating conditions. Such indicators should more clearly quantify the results of the SFRA method and allow easier and unambiguous detection of PT winding and magnetic circuit deformations and failures. This paper presents the *k*-mean algorithm, ANN and ANFIS as parts of a machine learning procedure intended to improve the CBM of PTs.

Contributions of this paper with respect to previous research can be remarked as follows:

- Application of all relevant SFRA statistical indicators with the addition of a new indicator, the reverberation time, to improve the database for applying machine learning algorithms.
- Testing and comparison of different machine learning algorithms. (ANN, ANFIS and k-mean) on real inputs, with power transformers from different periods of exploitation and lifetime from power plants.
- Developed machine learning algorithms can be used to detect fault severity as an output rather than a specific fault condition, thereby properly accounting for the overall insulation degradation due to aging. Fault severity is used to correlate fault development with the aging of the PTs in power plants.

The paper is organized as follows. Section 2 presents background theory for SFRA analysis and provides an overview of the most significant statistical parameters used for PT diagnostics. In section 3, the motivation for the introduction of machine learning is elaborated. The subsection in section 3 described how to use clustering method and AI methods. The results obtained using the previously described methods are compared and discussed in section 4. Conclusions of the research are derived in section 5.

2. SFRA and indicators

The standard SFRA method [23] involves comparing the frequency responses of consecutive periodically conducted tests. If the two compared frequency responses do not differ, then the longitudinal R, L and C parameters did not change significantly, indicating that no deformations occurred in the meantime. Mechanical failures are very progressive and can be detected by analyzing the winding response to an input signal containing a wide range of spectral components. When injecting the input signal into the PT terminal, the PT response is measured. The result of the SFRA method is a frequency response that is represented on a decibel and logarithmic scale. If the two compared frequency responses do not differ, it means that the longitudinal parameters have not changed significantly, i.e., there have been no winding deformations since the previous measurement.

Interpretation of SFRA results is performed empirically based on images generated by specialized devices. The input and output signal are recorded, after which the Fourier transform is performed on the recorded signals. The impulse response in the frequency domain $H(j\omega)$ is obtained as the ratio of the response (output) V_{output} and source (input) V_{input} signal:

$$H(j\omega) = \frac{V_{output}(j\omega)}{V_{input}(j\omega)}. \tag{1}$$

An example of several frequency responses obtained by SFRA analysis is shown in Fig. 1. Fig. 1 displays the SFRA results, transfer function in the frequency domain, of a PT which has been in use since 1968 but was overhauled in 2003. The figure shows the transfer functions obtained from tests conducted in different years. All results on Fig. 1. were recorded with the professional measuring instrument the Doble M5400 SFR Analyzer in power plants of electric power systems of Serbia. The year 2003 is assumed to be the initial time, i.e., the beginning of the transformer lifespan. As seen in Fig. 1, the frequency responses corresponding to different transformer ages do differ, but it is very difficult to make a clear distinction between them purely by means of visual inspection. The same can be said for either of the statistical indicators defined in Table 1. This transition to the frequency domain greatly facilitates the analysis and provides independence from the excitation waveform. By comparing the calculated transfer functions in successive tests, different changes in the winding geometry and changes in the PT magnetic circuit can be detected. Interpretation of the SFRA results relies on personal expertise and can lead to different and biased conclusions. It is not possible to draw clear conclusions about the failure based on the visual comparison of two diagrams.

Moreover, the conventional SFRA fingerprint is very inaccurate in detecting incipient and low mechanical fault levels [9]. The subjectivity of the observer can lead to big errors in the diagnosis of the PT condition.

Table 1
Statistical indicators obtained from SFRA results.

Indicator	Description	Indicator	Description
Correlation coefficient (CC)	$CC = \frac{\sum_{i=1}^{N} x_i y_i}{\sqrt{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}}$	Standard error of mean (SEM)	$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{N}}$
Cross Correlation Factor (CCF)	$CCF = \frac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N} x_i^2}\sqrt{\sum_{i=1}^{N} y_i^2 \sum_{i=1}^{N} x_i y_i}}$	Relation dispersion (RD)	$RD = \frac{\sigma}{\overline{x}}$
Jaccard Distance (JD)	$JD = \frac{\sum_{i=1}^{N} (x_i - y_i)^2}{\sqrt{\sum_{i=1}^{N} (x_i - \overline{x})^2 \sum_{i=1}^{N} (y_i - \overline{y})^2}}$	Stochastic Spectrum Deviation (SSD)	$SSD = \frac{100}{N}\sum_{i=1}^{N}\left\|\frac{y_i - x_i}{x_i}\right\|$
Absolute sum of logarithmic error (ASLE)	$ASLE = \frac{\sum_{i=1}^{N} \|20\log_{10}(y_i) - 20\log_{10}(x_i)\|}{N}$	Euclidean Distance (ED)	$ED = \sqrt{\sum_{i=1}^{N} (y_i - x_i)^2}$
Absolute average difference (DABS)	$DABS = \frac{\sum_{i=1}^{N} \|y_i - x_i\|}{N}$	Standard deviation (SD)	$SD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1}}$
Min–max (MM) ratio	$MM = \frac{\sum_{i=1}^{N} \min(y_i, x_i)}{\sum_{i=1}^{N} \max(y_i, x_i)}$	Sum Squared Ratio Error (SSRE)	$SSRE = \frac{\sum_{i=1}^{N} \left(\frac{Y(i)}{X(i)} - 1\right)^2}{N}$
Maximum of Difference (MD)	$MD = \max(y_i - x_i)$	Sum Squared Max-Min Ratio Error (SSMMRE)	$SSMMRE = \frac{1}{N}\frac{\sum_{i=1}^{N} \min(y_i, x_i)}{\sum_{i=1}^{N} \max(y_i, x_i)}$
Covariance (COVAR)	$COVAR = \frac{1}{N}\sum_{1}^{N} (x_i - \overline{x})(y_i - \overline{y})$	Integral Difference (ID)	$ID = \int (y(f) - x(f))df$
Least Squared Error (LSE)	$LSE = \frac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$	Integral of Absolute difference (AID)	$AID = \int \|(y(f) - x(f))\|df$
Average deviation (AVEDEV)	$AVDEV = \frac{\sum_{i=1}^{N} \|x_i - \overline{x}\|}{N}$	Standardized Difference Area (SDA)	$SDA = \frac{\int \|(y(f) - x(f))\|df}{\int \|x(f)\|df}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (y_i - \overline{x})}{N}$	Sum of Error (SE)	$SE = \frac{\sum_{i=1}^{N} (y_i - x_i)}{N}$

'N' - the overall size of the data set; 'x_i' and 'y_i' - the overall of all variables present in the data set; '$\overline{x}$' and '$\overline{y}$' - mean values of observations; 'L' - lower limit of median class; 'cf' - cumulative frequency of class prior to median class; 'f' - frequency of median class; 'h' - class size.

In order to resolve these problems, various statistical indicators are proposed in CIGRE document [6,9,11]. These indicators are summarized and presented in Table 1. An advantage of numerical values over the engineer's experience may be the elimination of human-introduced subjectivity. Moreover, certain indicators can reveal features of the obtained response that cannot be observed by conventional visual inspection-based approach.

In addition to the parameters shown in Table 1, the reverberation time parameter T60 was used in this paper. It has been shown in the literature that the reverberation time T60 can be used as a good identifier of PT failures, but at the same time it can show the degree of PT failure [8]. The reverberation time is calculated based on the impulse response of the system (time domain). A decline curve $L_R(t)$ is calculated based on the impulse response of the system $h(t)$ [24]:

$$L_R(t) = \int_t^{\infty} h^2(t)dt = \int_0^{\infty} h^2(t)dt - \int_0^{t} h^2(t)dt. \tag{2}$$

The reverberation time is defined as the time required for the reverberation curve to reduce by 60 dB from its initial value, which corresponds to a decrease of one million times on a linear scale. The dynamic of 60 dB is difficult to achieve in practice, which is why T60 is determined indirectly using parameters T10, T20, or T30. The interpretation of these parameters is illustrated in [24]. The reverberation time T60 is calculated by multiplying T30 by two, T20 by three, or T10 by six.

None of the existing approaches enables accurate and reliable assessment of the PT condition as it changes due to ageing. The reverberation time was shown to be very reliable in terms of fault detection, its ability to differentiate between faults of different severities is somewhat limited. Namely, the reverberation time does exhibit a monotonous variation with respect to the fault severity, but this variation is not pronounced enough to perceive subtle changes in winding condition that occur due to ageing. It is therefore the authors' goal to employ this parameter along with the previously reported indicators (listed in Table 1) as inputs of machine learning algorithms. This approach is expected to yield a more reliable and accurate estimate of the transformer condition.

3. Machine learning

Machine learning is a discipline that allows computers to learn without explicit programming. This learning technique represents a subfield of artificial intelligence, and it arose from the study of pattern recognition and the theory of computational learning, as well as the use of statistics for the purpose of learning based on previously available data, designing and developing appropriate algorithms. In addition to being a combination of statistics, optimization and linear algebra, it also relies on graph theory, functional analysis, and other mathematical fields. Machine learning algorithms fall into one of two main categories: Supervised Learning and Unsupervised Learning. One of the most common forms of machine learning is unsupervised learning, which is gaining importance because it does not require the algorithm to be fed data with predefined outputs. Algorithms from this group allow approaching the problem without prior knowledge of what the results should be. This is exactly the case with PT diagnostics. There are no output values in the monitoring database of SFRA. Therefore, for the set of monitoring results, there is no output value related to the probability of failure and the urgency of intervention. One of the most common unsupervised learning approaches is cluster analysis, which is used to classify PTs into groups with a certain probability of failure. In existing literature, clustering methods were used to indicate mechanical deformities. However, no attempt to define the probability or the severity of fault using the clustering approach has been made so far.

Database is formed based on results of experiments described in [8]. One measurement is performed under healthy condition, indicated as OK state with fault severity 0, and 7 measurements are performed with PT failures of varying severity: 0–1, 0–2, 0–3, 0–4, 0–5, 0–6, 0–7. To expand the database and thereby improve the ML algorithm training procedure, additional measurements are performed under the following internal short circuit conditions:

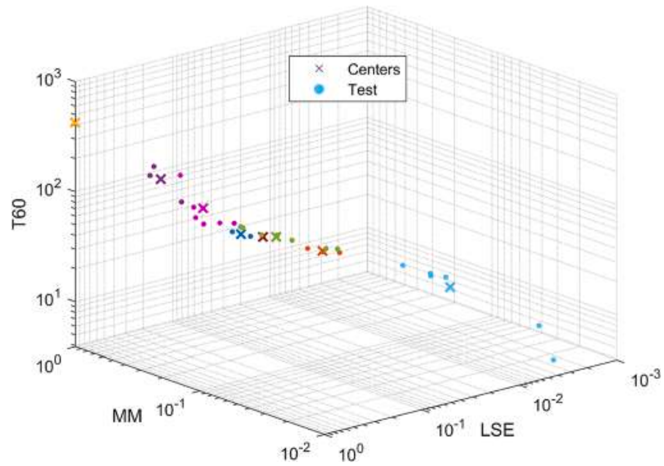


Fig. 2. Centers of cluster and measurements for different fault severities obtained from laboratory measurements.

- 1–2, 2–3, 3–4, 4–5, 5–6, 6–7, indicating the lowest degree of fault severity 14.28 %,
- 1–3, 2–4, 3–5, 4–6, 5–7, indicating fault severity of 28.56 %,
- 1–4, 2–5, 3–6, 4–7, indicating fault severity of 48.84 %,
- 1–5, 2–6, 3–7, indicating fault severity of 57.12 %,
- 1–6, 2–7, indicating fault severity of 71.40 %,
- 1–7, indicating fault severity of 85.68 %.

Indicators were calculated and memorized for each of the 29 experiments and representative values are shown in Table 4 in the Appendix.

3.1. Unsupervised Machine learning

3.1.1. K-mean

Clustering analysis is a multi-criteria optimization technique used to classify objects into groups. So, each of the objects is classified into one of the clusters with a certain accuracy. This grouping should provide the following cluster characteristics:

- homogeneity within the cluster, i.e., data belonging to the same cluster should be as similar as possible, and
- heterogeneity between clusters, i.e., data belonging to different clusters should be as different as possible.

Cluster analysis can be applied in data research as a stand-alone tool for gaining insight into the distribution of data, observing the characteristics of each cluster, and focusing on a specific set of clusters for further analysis. Clustering represents an iterative process of merging objects into groups so that in the next stage the new objects and previously formed groups are merged. This actually means that previously formed groups are only expanded with new objects according to the specifics of the selected criterion, and at the same time, there is no possibility of moving objects from one group to another during the iteration process. In order to carry out clustering analysis in general, similarity between two objects is quantified based on their characteristics using the Euclidean distance, given by the following relationship:

$$d_E(x,y) = \|x-y\| = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} = d_E(y,x) \tag{1}$$

where $x = (x1, \ldots, xn)$ and $y = (y1, \ldots, yn)$ are attribute values of two objects of length n. In this implementation x and y are values of different statistical indicators. For optimal performance, cluster algorithms require data normalization, to prevent attributes with a larger range from having a greater impact on the results. The normalized values are in the range [0,1], unless the new data represent extreme values that lie outside the standardized range. According to [25] one of the proposed solutions is the min–max normalization defined in the following manner:

$$x^* = \frac{x-\min(x)}{range(x)} = \frac{x-\min(x)}{\max(x)-\min(x)} \tag{2}$$

All the indicators shown in Table 1 and reverberation time T60 are used as diagnostic input parameters. Along with these SFRA-based indicators, the transformer reverberation time introduced in [8] is also used as an input parameter. The values of all indicators must be normalized. Additionally, it is necessary to consider whether a higher value of an indicator indicates a better or worse state of the energy transformer. Such expert knowledge is presented in [8]. The goal of the methodology is to obtain the centers of eight clusters, where the center closest to the origin represents the healthy state of the PT. The most distant cluster represents the most severe form of failure and requires immediate intervention. Fig. 2 shows the cluster centers in a subspace defined by three indicators. The data is displayed on a logarithmic scale to achieve better dispersion. The given three-dimensional representation serves to provide a visual interpretation, however, it should be noted that the cluster centers are defined in a multidimensional space, where each statistical indicator represents one dimension. The cluster centers were formed based on laboratory measurements performed under different transformer conditions, i.e., with inter–turn short–circuits spanning different portions of the phase winding. Transfer functions were recorded for seven different short–circuit faults of different severities, where 0–1 is the least severe (between adjacent coils), and 0–7 is the most severe (between the top and bottom coil). The complete dataset containing the values of all indicators under different winding conditions is given in Table 4 in the Appendix.

The cluster centers correspond to specific fault conditions, i.e., interturn short circuits spanning different portions of the winding. The condition of a transformer operating in realistic conditions would never completely coincide with these laboratory conditions. The diagnostic procedure would be based on the proximity of the coordinates (inputs) obtained for a given transformer to the cluster centers. Namely, when new measurements are added to the database, they are assigned to one of the clusters depending on the distance from the cluster centers, thereby indicating the severity of the fault and the urgency of the intervention. New measurements are used to obtain new transfer functions from the SFRA test. Based on these transfer functions, relevant indicator values are calculated.

3.2. Supervised machine learning

3.2.1. Artificial neural network

Artificial Neural Network (ANN) represents a machine learning-based approach which aims to determine the output of a system based on its inputs. It is the most useful in the case of stochastic systems, where it is very difficult to define the relationship between the input and output values. Therefore, it is necessary to apply a method that can estimate the output of the system without knowing the exact definition of the system transfer function. ANN represents a mathematical model of a human neural network which can, based on the relevant input data and transferring the information between artificial neurons, determine output information which represent the solution of the analyzed problem. Fig. 3 shows the mathematical representation of a neuron in ANN. Each neuron has inputs that are connected to it through synapses. Those inputs can originate from the outer world or from the neurons in the previous layer. In Fig. 3, x_j represents the jth input, $j = 1,..n$. The strength of synapses is defined through synaptic weight ω_{ij}, where the positive value of ω_{ij} corresponds to excitatory synapses, while negative values model inhibitory synapses.

All inputs multiplied by the appropriate weight coefficients are in-

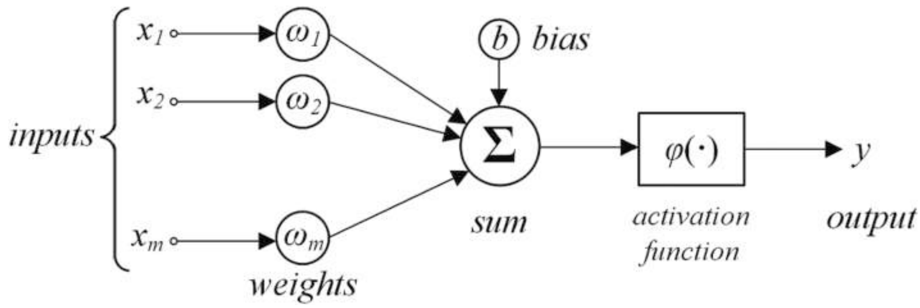


Fig. 3. ANN mathematical model.

Table 2
Statistical indicators obtained from SFRA results for 7 different types of faults.

Inputs				Outputs	
CC	MM	LSE	T60	Fault Severity [%]	Fault type
1.00	1.00	1.00	543.71	0	OK
0.97	0.94	1.01	512.70	14.28	0–1
0.96	0.95	0.97	484.91	28.56	0–2
0.95	0.93	0.95	386.49	42.84	0–3
0.93	0.90	0.92	291.02	57.12	0–4
0.91	0.90	0.89	151.75	71.4	0–5
0.86	0.90	0.78	76.48	85.68	0–6
0.79	0.89	0.78	29.67	100	0–7

tegrated into the transfer function and this value is compared to the value of the threshold. In the case that the transfer function is greater than the threshold, activation value equals 1, otherwise it equals 0. This statment can formally be written in the following form:

$$\phi = \phi\left(\sum_{j=1}^{n} x_j \cdot w_{i,j} - \theta_j\right) \tag{3}$$

The most used activation function of a neuron is sigmoid function, defined by:

$$o_j = \frac{1}{1 + e^{-\alpha \cdot \phi}} \tag{4}$$

The structure of ANN has a different number of neurons in different layers. There is no clear way to determine the optimal number of neurons and layers, but it depends on the nature of the problem and the input–output relationship. Mean Average Percentage Error (MAPE) is used as a performance index to evaluate the prediction capability of each model. Typically, the model and structure of ANN is chosen to minimize the MAPE, which is formally given by:

$$MAPE(x, y) = \frac{\sum_{i=1}^{N} \frac{|y_i - x_i|}{x_i}}{N} \cdot 100\% \tag{5}$$

The training and testing process of ANN is carried out through three phases. AI is used to predict the fault severity which takes values from the range 0–100 %. As in the previous case (k-mean algorithm), the inputs of the algorithm are the statistical indicators listed in Table 1 and the reverberation time. Part of the used database is shown in Table 2. The entire data set is displayed in Table 4 (Appendix). Different types of faults are characterized by different probabilities and fault severities. Based on [8], the indicators CC, MM, LSE and T60 provide the most reliable fault detection and identification of fault severity.

The training data are selected from the entire set of available data. The training set uses 70 % of the database, the testing set uses 15 % of the database, and 15 % of the dataset is used for validation. These datasets are randomly generated from the database shown in Table 4 (Appendix). The data is then normalized to avoid saturation. Training of the ANN is performed using the back-propagation algorithm. In this case, the smallest MAPE error is 17.329 % and is achieved using a structure with three layers and 64, 24 and 8 neurons per layer, respectively. The results of training ANN prediction of fault severity are presented in Fig. 4.

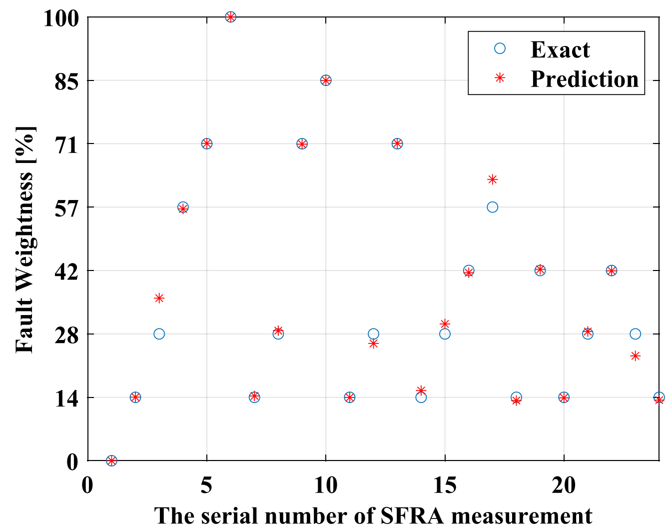


Fig. 4. Training results for ANN – prediction of fault severity based on statistical indicators.

3.2.2. *Adaptive network based Fuzzy inference system*

Neural networks and Fuzzy systems both have the ability to manage non-linear, dynamic systems for which there is no corresponding mathematical model. The disadvantage of neural networks is that it is not clear how the management problem is solved, that is, the lack of interpretability. Neural networks do not have the ability to generate or establish any kind of structural knowledge, for example in the form of rules, nor the ability to use prior knowledge to reduce training time. In contrast, Fuzzy systems perform inference completely transparently through a set of concrete linguistic rules, but lack the appropriate learning algorithms to perform adjustments based on the database. Neural networks and Fuzzy systems are dynamic systems that perform parallel processing in order to evaluate the input–output relationship without using a mathematical model, but only by experience-based learning and the obtained sample. In Fuzzy systems, input–output relationships are given explicitly in the form of if-then rules. On the other hand, no explicit relationships are defined by neural networks, but they are implicitly contained within the network's parameters and structure. One of the results of the most successful integration of fuzzy systems and neural networks is ANFIS (Adaptive Network based Fuzzy Inference System), which is one of the most applied neuro-fuzzy models. Thus, ANFIS is a class of adaptive networks that are functionally equivalent to Fuzzy inference, using the advantages of both neural networks and Fuzzy logic. One of the main features of this system is its adaptability, i.e., its membership functions are derived based on the data describing the

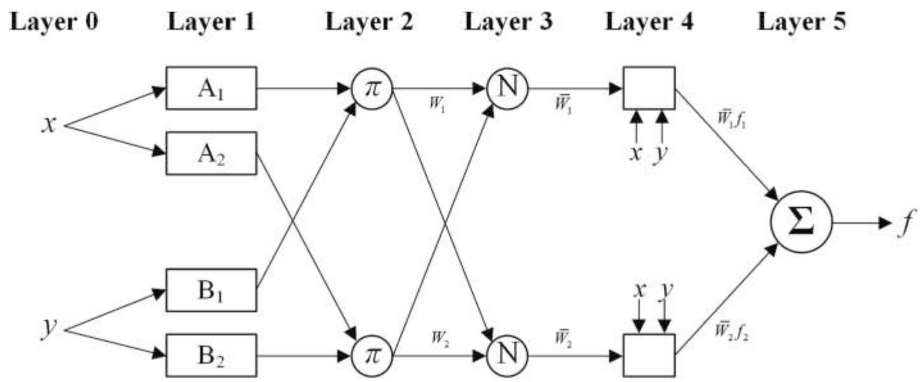


Fig. 5. ANFIS structure.

system behavior given through the input–output system variables. With this approach, it is not necessary to manually perform phasing and set the membership functions, which are now formed based on the monitoring database of the key power equipment. For this reason, it is desirable to have as large a database of monitoring results as possible, because in this way the system will be trained better and more realistically, with the aim of drawing accurate and useful conclusions.

In order to explain the ANFIS technique, a five-layer ANFIS architecture is displayed in Fig. 5. Each of the layers has its clearly defined role. Square nodes represent adaptive nodes which have parameters assigned to them, whereas circular, fixed nodes have no parameters. The set of parameters of a certain adaptive network is the union of the set of parameters of all adaptive nodes. In order to achieve the desired input–output mapping, parameters are adjusted according to the training set, based on the steepest descent method (which is described in section 3.3). This method is based on the fastest passage of the extremum via the derivative. So the nodes are classified into the following layers:

- **Layer 1**: each node of this layer is a square node and performs the following function:

$$O_i^1 = \mu_{Ai}(x), \quad i = 1, 2, 3... \tag{6}$$

where x is the input to node i, and A is a linguistic term (such as "small", "large", …) that is related to the given function of the node. In other words, O is the value of the membership function A that indicates to what extent a given x belongs to the Phase set Ai. Membership functions can be represented in different forms: triangular, trapezoidal, sigmoid… The most used is a bell function with a maximum value of 1 and a minimum value of 0:

$$\mu_{Ai}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i}\right)^2\right]^{b_i}} \tag{7}$$

where (a_i, b_i, c_i) is a set of parameters related to the premise of the rule that needs to be optimized through the learning process.

- **Layer 2:** each node of this layer is circular and is labeled π and its task is to multiply the input signals and then forward them to the next layer.

$$w_i = \mu_{Ai}(x) \times \mu_{Bi}(x), \quad i = 1, 2, 3... \tag{8}$$

The output of each node represents the total strength of the rule. Instead of multiplication, other operators can be applied here. Multiplication represents the conjunction of linguistic terms A and V and forms a rule.

- **Layer 3:** The nodes of this layer are circular and marked *N*. Their task is to perform calculations that obtain the normalized strength of the *i*-th rule:

$$\overline{w_i} = \frac{w_i}{w_1 + w_2}, i = 1, 2, 3... \tag{9}$$

where w_i are the strengths of all the rules. For each node, the ratio of the strength of the *i*-th rule to the sum of the strengths of all other rules is calculated.

- **Layer 4:** Each node *i* of this layer is a square node, which performs a function:

$$O_i^4 = \overline{w_i} f_i = \overline{w_i}(p_i x_1 + q_i x_2 + r_i), \quad i = 1, 2, 3... \tag{10}$$

where (p_i, q_i, r_i) are the set of parameters related to the conclusion.

- **Layer 5:** is represented by a circle and calculates the total output as the sum of all input signals:

$$O_i^5 = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}, \quad i = 1, 2, 3... \tag{11}$$

The purpose of using the presented structure and algorithm is that the parameters of the membership functions are not chosen arbitrarily but are determined according to the input–output data. The parameters associated with the membership functions change through the learning process. Calculating these parameters (or adjusting them) is facilitated by the gradient vector. This gradient vector determines how well the Fuzzy Inference system models the input–output data for a given set of parameters. Once the gradient vector is obtained, any of several optimization methods can be applied to adjust the parameters and reduce the error. The error is usually defined as the sum of the squared differences between the actual and desired outputs. ANFIS uses a hybrid

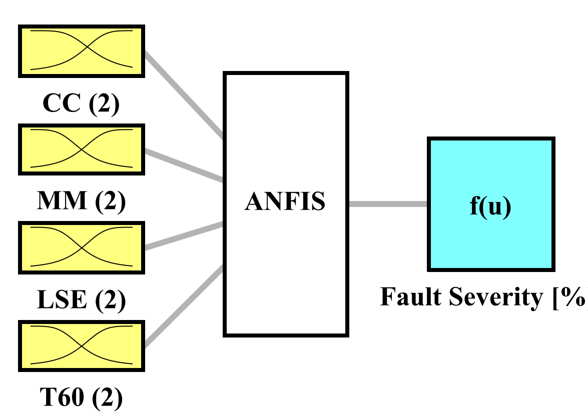
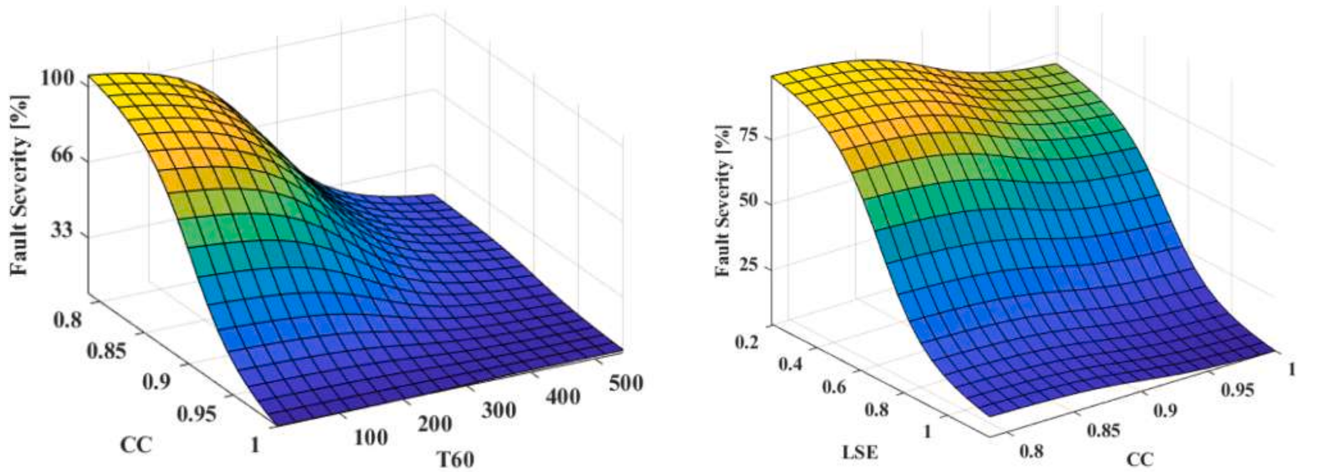


Fig. 6. Fuzzy expert system based on ANN.



(a) *(b)*

Fig. 7. Fault severity as the function of the combinations of: (a) CC and T60, (b) LSE and CC.

Table 3
Results obtained using the k-mean algorithm. ANFIS, ANN and CBM.

Transformer	CC	MM	LSE	T60	Age	*k*-mean	ANFIS	ANN	**CBM**
PT1	1.00	1.00	1.00	322.35	5	1	26.33	7.70	**9**
PT1	0.99	0.92	0.95	301.56	7	2	27.42	14.35	**8**
PT1	1.00	0.88	0.90	213.93	8	2	36.66	20.90	**7**
PT1	0.70	0.71	0.82	211.86	12	5	32.06	43.58	**6**
PT1	0.83	0.74	0.78	198.15	13	5	29.22	58.78	**5**
PT1	1.00	0.96	1.10	288.75	20	4	40.55	59.83	**4**
PT2	1.00	0.91	1.11	184.32	23	4	86.34	64.82	**3**
PT2	0.99	0.88	1.14	186.79	27	4	82.88	67.87	**2**
PT2	1.00	0.88	1.16	183.38	28	4	84.81	75.17	**1**

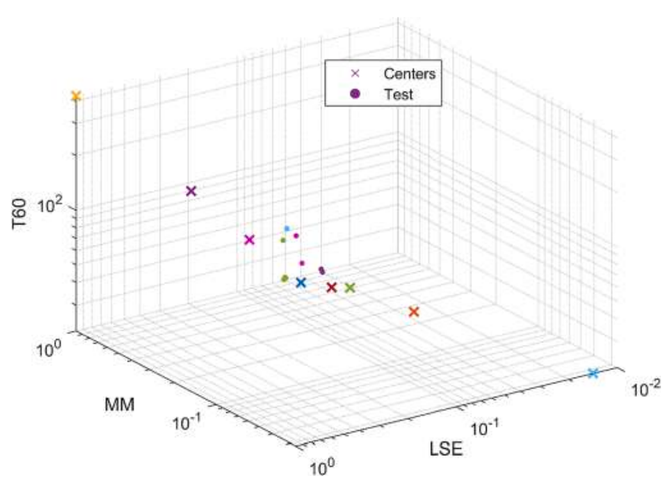


Fig. 8. Results for k-mean from Table 3.

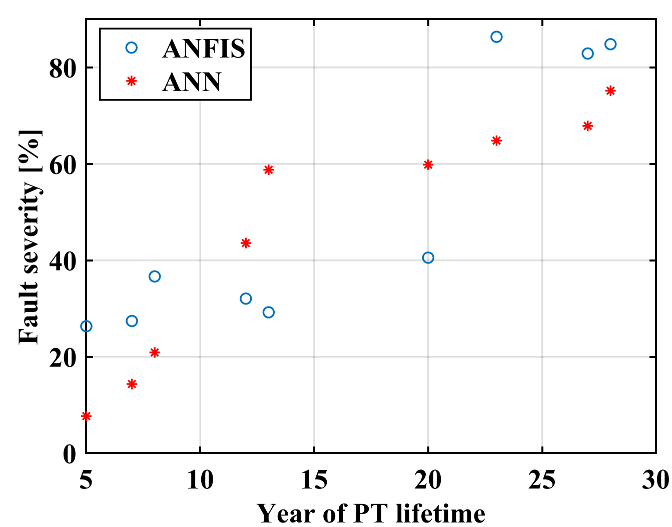


Fig. 9. Comparison of ANN and ANFIS results for CBM.

algorithm that combines backpropagation gradient descent and least squares to create a learnable fuzzy inference system (membership functions are iteratively adjusted based on the input–output training set). The backpropagation algorithm adjusts the parameters of the membership functions of the premises, while the least squares method adjusts the coefficients of the linear combination.

Fig. 6 presents the ANFIS system created in Matlab. Fig. 7 shows the dependency of the failure probability with respect to the input data and actually represents the transfer function between the input and the probability of failure. The presented results point to the logical conclusions: smaller values of CC, LSE and T60 indicate a higher probability of failure.

4. Results

Based on data from laboratory experiments given in the Appendix, three AI methodologies are trained, validated, and tested. To correlate the probabilities and severity of the failure with the years of PT exploitation and the methodology that best indicates it, measurements from power plants were applied as new test inputs. To compare the results of the presented methods of supervised and unsupervised machine learning, 9 SFRA monitoring data sets are used. The first four measurements are for one PT and other five are for PTs of same brands and

Table 4
Values of indicators obtained from tests on a laboratory setup for different fault severities.

FAULT	CC	CCF	JD (xE2)	ASLE (xE1)	DABS (xE-2)	MM	MD (xE2)	CVAR	LSE	ADEV (xE-1)	VAR (xE-10)	SEM (xE-1)	RD (xE2)	SSD (xE2)	SD (xE-10)	SSRE (xE1)	SSMMRE	-SE	1-SSMMRE	T60 (s)
OK	1.00	0.92	0.00	0.00	0.00	1.00	1.65	1.46	1.00	4.50	4.50	6.10	9.07	3.33	7.65	0.00	0.00	0.91	1.00	0.43
0–1	0.99	0.99	0.02	0.82	1.50	0.43	1.72	0.67	0.46	2.68	3.57	6.28	8.08	2.03	0.55	0.31	0.32	0.65	0.68	0.20
0–2	0.99	0.74	0.02	0.84	1.58	0.44	1.05	0.72	0.49	3.17	3.79	6.00	8.33	2.40	2.46	1.48	0.31	0.97	0.69	0.17
0–3	0.99	0.50	0.02	1.38	1.92	0.32	0.48	0.51	0.35	2.54	3.46	5.57	7.95	1.95	1.37	3.17	0.51	1.45	0.49	0.10
0–4	0.98	0.30	0.05	1.48	2.18	0.19	0.21	0.29	0.20	1.62	2.98	5.19	7.38	1.27	3.42	4.62	0.54	1.83	0.46	0.06
0–5	0.97	0.14	0.12	1.27	2.36	0.09	0.05	0.12	0.08	0.90	2.59	4.88	6.88	0.73	1.23	6.34	0.47	2.14	0.53	0.04
0–6	0.90	0.05	0.51	1.29	2.38	0.03	0.01	0.03	0.02	0.32	2.29	4.87	6.46	0.28	0.39	3.62	0.48	2.26	0.52	0.03
0–7	1.00	0.72	3.80	1.13	2.34	0.02	0.91	0.00	0.00	0.10	2.27	5.85	6.45	0.11	0.05	0.37	0.43	1.19	0.57	0.00
1–2	0.99	0.65	0.03	0.81	1.76	0.33	0.80	0.50	0.34	2.14	3.28	5.83	7.75	1.64	1.50	0.67	0.30	1.17	0.70	0.17
1–3	0.99	0.40	0.03	1.43	1.98	0.28	0.33	0.43	0.30	2.15	3.26	5.36	7.72	1.66	4.78	2.74	0.51	1.68	0.49	0.07
1–4	0.98	0.25	0.07	1.43	2.23	0.15	0.13	0.20	0.14	1.17	2.75	5.06	7.09	0.94	0.14	3.04	0.51	1.97	0.49	0.05
1–5	0.96	0.12	0.17	1.70	2.35	0.07	0.03	0.08	0.06	0.63	2.46	4.86	6.71	0.53	0.38	3.73	0.57	2.17	0.43	0.04
1–6	0.92	0.06	0.64	1.96	2.37	0.03	0.01	0.02	0.02	0.27	2.27	4.86	6.44	0.24	0.34	2.75	0.59	2.26	0.41	0.03
1–7	0.99	0.48	2.83	2.02	2.34	0.02	0.45	0.00	0.00	0.12	2.27	5.44	6.44	0.12	0.15	0.57	0.60	1.63	0.40	0.01
2–3	0.99	0.36	0.05	1.26	2.05	0.20	0.26	0.27	0.19	1.29	2.85	5.26	7.21	1.01	2.46	0.72	0.47	1.81	0.53	0.07
2–4	0.99	0.25	0.08	1.70	2.21	0.13	0.13	0.17	0.12	0.93	2.66	5.07	6.97	0.75	1.57	1.40	0.55	1.99	0.45	0.06
2–5	0.97	0.16	0.17	2.04	2.32	0.07	0.06	0.09	0.06	0.59	2.46	4.92	6.71	0.49	1.57	2.25	0.58	2.11	0.42	0.05
2–6	0.96	0.14	0.37	2.31	2.37	0.04	0.04	0.04	0.03	0.37	2.33	4.87	6.52	0.32	0.41	3.01	0.62	2.13	0.38	0.03
2–7	0.99	0.46	0.52	2.36	2.39	0.03	0.40	0.03	0.02	0.33	2.28	5.41	6.46	0.29	0.51	4.22	0.67	1.66	0.33	0.03
3–4	0.99	0.33	0.06	1.45	2.10	0.18	0.23	0.25	0.17	1.22	2.81	5.21	7.17	0.96	1.85	1.02	0.50	1.85	0.50	0.06
3–5	0.99	0.27	0.10	1.89	2.26	0.11	0.16	0.14	0.10	0.85	2.61	5.11	6.90	0.69	1.57	1.90	0.58	1.94	0.42	0.05
3–6	0.98	0.28	0.14	2.31	2.33	0.08	0.17	0.10	0.07	0.69	2.51	5.13	6.77	0.56	0.10	3.15	0.65	1.89	0.35	0.05
3–7	0.99	0.46	0.14	1.85	2.37	0.08	0.43	0.10	0.07	0.79	2.53	5.44	6.80	0.64	0.00	6.11	0.63	1.63	0.37	0.04
4–5	0.99	0.40	0.05	1.79	2.10	0.18	0.34	0.26	0.18	1.28	2.84	5.36	7.20	1.00	0.48	1.22	0.58	1.70	0.42	0.06
4–6	0.99	0.44	0.07	1.81	2.21	0.15	0.40	0.20	0.14	1.14	2.75	5.45	7.09	0.90	0.48	2.24	0.58	1.58	0.42	0.06
4–7	0.99	0.54	0.06	1.75	2.23	0.16	0.56	0.24	0.16	1.37	2.85	5.57	7.22	1.08	3.76	4.39	0.61	1.50	0.39	0.05
5–6	0.99	0.61	0.04	1.25	2.02	0.22	0.72	0.33	0.22	1.54	2.97	5.74	7.37	1.19	3.01	1.10	0.48	1.29	0.52	0.07
5–7	1.00	0.65	0.03	1.22	2.01	0.26	0.77	0.40	0.27	1.95	3.17	5.73	7.61	1.50	1.78	2.48	0.46	1.34	0.54	0.06

constructions, but of different ages. The signal processing of data was done with the Doble M5400 SFR Analyzer. These data sets represent historical SFRA measurements on the block transformers in power plants in the production units of the Serbian power system. SFRA measurements are obtained from PTs of different ages. Accordingly, the condition of the analyzed PTs is expected to differ. The proposed methodology should provide information regarding the condition of the transformer and the probability of its failure depending on its age, i.e., the year of its exploitation. Table 3 presents the results obtained using different methodologies. Table 3 displays only the values of four indicators selected for demonstrative purposes. However, it should be noted that all statistical indicators listed in Table 4 in the Appendix are obtained and used as inputs of machine learning algorithms. The displayed results indicate that the *k*-mean algorithm follows a logically consequential dependency (the older the transformer, the higher the probability of failure). The *k*-mean algorithm is clustering the older transformers into the most vulnerable eighth cluster, while “younger” transformers are sorted into clusters with lower numbers. Fig. 8 shows the test data (denoted by small circular markers), which are assigned to individual clusters whose centers can be seen in the picture (denoted by large cross markers). However, the algorithm does not delimit clusters well enough and that can be seen in Fig. 8. The results obtained using ANFIS indicate that PTs of ages twelve and eight (rows 3 and 4 in Table 3) are more likely to suffer a failure than a 13-year-old transformer (row 5). This apparent inconsistency is due to the fact that, while all results are obtained for PTs of the same type, the actual units under observation (PT1 and PT2, see Table 3) are different and operate under different conditions. This explains the notable “leap” in fault severity between years 20 and 23, as these results correspond to two different PT units. Of course, such a deterioration of transformer condition could occur for a single transformer over a three-year span. This would indicate that the transformer had suffered a serious fault over that time period and that an immediate intervention is in order.

In Fig. 9, the fault severity depending on the age of the PT is displayed. Results from Table 3 obtained using the ANFIS and ANN techniques are compared. These results can be used for asset management and CBM to propose a sequence of maintenance. The last column of the Table 3 shows the value of CBM based on the ANN results. The ANN results are selected as they exhibit the best correlation with the transformer age (based on Fig. 9). Results such as this could be used for the purpose of implementing CBM in order to properly schedule corrective, preventive maintenance and to prevent the spread of defects and the occurrence of dangerous accidents in power systems.

5. Conclusion

For the purposes of condition monitoring and diagnostics of PTs, SFRA measurements and statistical analysis of as many relevant indicators as possible need to be performed on a yearly basis. Standard diagnostics and maintenance procedures do not require the calculation and comparison of statistical indicators or analysis of their trends. This paper points out the importance of such analyses for proper transformer state monitoring and maintenance. A total of 23 indicators were calculated from SFRA measurements for PTs with faults of different severities. Based on the new measurement data and the trend of past measurements, fault severity and probability of failure of PTs during exploitation is determined. The unsupervised machine learning approach based on statistical indicators is applied to cluster PTs according to their state (health). The obtained results indicate that the clustering approach, while applicable, is less accurate than the results of supervised machine learning. ANFIS and ANN approaches provide very good results that can be used to improve asset management. This concept of CBM enables automatic proper and timely decision-making regarding maintenance planning. Decisions based on health assessment ensure timely fault detection and removal, shorter interruptions, and savings in electric power system maintenance. Monitoring multiple statistical indicators and their interpretation is an extremely difficult, if not impossible engineering task. On the other hand, Machine Learning methodologies presented in this paper are intended precisely to derive conclusions from a great number of apparently unrelated parameters, i.e., the SFRA indicators. Future work will be focused on testing the new algorithms on a greater number of PTs of different ages and conditions and achieving predictive maintenance.

CRediT authorship contribution statement

Miloš Bjelić: Software, Visualization, Writing – review & editing. **Bogdan Brković:** Investigation, Supervision, Writing – original draft, Writing – review & editing. **Mileta Žarković:** Investigation, Supervision, Writing – original draft, Writing – review & editing. **Tatjana Miljković:** Project administration, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

The authors would like to thank the Ministry of Science and Technological Development of the Republic of Serbia which, within the framework of Project III 42009 ‘Smart grid’ and Project TR36026, made this work possible.

Appendix A

The values of all indicators used in the presented analysis obtained during the laboratory measurements for different short–circuit conditions are presented in Table 4. These values were used to form training, validation, and testing datasets for the applied machine learning algorithms.

References

[1] Dick EP, Erven CC. Transformer diagnostic testing by frequency response analysis. IEEE Trans PAS 1978;97:2144–53. https://doi.org/10.1109/TPAS.1978.354718.
[2] Secue JR, Mombello EE. New SFRA measurement interpretation methodology for the diagnosis of power transformers. Electr Eng 2014;96:183–98. https://doi.org/10.1007/s00202-013-0286-6.
[3] IEEE PC57.149/D8 (2012) Guide for the Application and Interpretation of Frequency Response Analysis for Oil Immersed Transformers.
[4] IEC 60076-18 (2012) Power transformers - Part 18: Measurement of frequency response.
[5] CIGRE Working Group A2.26 (2008) Mechanical-condition assessment of transformer windings using frequency response analysis.
[6] CIGRE Working Group A2,53 (2020) Advances in the interpretation of transformer Frequency Response Analysis (FRA) Reference 812.
[7] Zhao X, Yao C, Abu-Siada A, Liao R. High frequency electric circuit modeling for transformer frequency response analysis studies. Int J Electr Power Energy Syst 2019;111:351–68. https://doi.org/10.1016/j.ijepes.2019.04.010.
[8] Bjelić M, Brković B, Žarković M, Miljković M. Fault detection in a power transformer based on reverberation time. Int J Electr Power Energy Syst 2021;137:1–12. https://doi.org/10.1016/j.ijepes.2021.107825.
[9] Abbasi AR, Mahmoudi MR. Application of statistical control charts to discriminate transformer winding defects. Electr Pow Syst Res 2021;191:1–11. https://doi.org/10.1016/j.epsr.2020.106890.
[10] Shamlou A, Feyzi MR, Behjat V. Winding deformation classification in a power transformer based on the time-frequency image of frequency response analysis using Hilbert-Huang transform and evidence theory. Int J Electr Power Energy Syst 2021;129:1–15. https://doi.org/10.1016/j.ijepes.2021.106854.

[11] Tahir M, Tenbholen S, Miyazaki S. Analysis of statistical methods for assessment of power transformer frequency response measurements. IEEE Trans Power Delivery 2021;36(2):618–26. https://doi.org/10.1109/TPWRD.2020.2987205.
[12] Zhao Z, Tang C, Chen Y, et al. Interpretation of transformer winding deformation fault by the spectral clustering of FRA signature. Int J Electr Power Energy Syst 2021;130:1–12. https://doi.org/10.1016/j.ijepes.2021.106933.
[13] Bagheri S, Moravej Z, Gharehpetian GB Classification and Discrimination among Winding Mechanical Defects, Internal and External Electrical Faults and Inrush Current of Transformer. IEEE Transactions on Industrial Informatics 14: 484-493. 10.1109/TII.2017.2720691.
[14] Bigdeli M, Vakilian M, Rahimpour E. Transformer winding faults classification based on transfer function analysis by support vector machine. IET Electr Power Appl 2012;6:268–76. https://doi.org/10.1049/iet-epa.2011.0232.
[15] Liu JN, Zhao ZY, Tang C, Yao CG, Li CX, Islam S. Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine. IEEE Access 2019;7:112494–504. https://doi.org/10.1109/ACCESS.2019.2932497.
[16] Ahmadi H, Vahidi B, Nematollahi AF. A simple method to detect internal and external short-circuit faults, classify and locate different internal faults in transformers. Electr Eng 2021;103:825–36. https://doi.org/10.1007/s00202-020-01122-3.
[17] Ekojono P, R.A., Apriyani, M.E, et al. Investigation on machine learning algorithms to support transformer dissolved gas analysis fault identification. Electr Eng 2022; 268. https://doi.org/10.1007/s00202-022-01532-5.
[18] 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, 26th Electrical Power Distribution Conference, May 11-12, 2022, Tehran, Iran.
[19] Behkam R., Karami H., Naderi M.S., Gharehpetian G.B. Application of Artificial Neural Network on Diagnosing Location and Extent of Disk Space Variations in Transformer Windings Using Frequency Response Analysis, 2022 30th International Conference on Electrical Engineering (ICEE).
[20] Behkam R., Naderi M.S., Gharehpetian G.B. Detection of transformer defects in smart environment using frequency response analysis and artificial neural network based on data-driven systems. 11th Smart Grid Conference (SGC); 2021.
[21] Ghanizadeh AJ, Gharehpetian GB. ANN and cross-correlation based features for discrimination between electrical and mechanical defects and their localization in transformer winding. IEEE Trans Dielectr Electr Insul 2014;21(5):2374–82. https://doi.org/10.1109/TDEI.2014.004364.
[22] Islam MM, Lee G, Hettiwatte SN. A review of condition monitoring techniques and diagnostic tests for lifetime estimation of power transformers. Electr Eng 2018;100: 581–605. https://doi.org/10.1007/s00202-017-0532-4.
[23] Ghani SA, Thayoob YHM, Ghazali YZY, Khiar MSA, Chairul IS. Condition monitoring of distribution transformer's mechanical parts using sweep frequency response analysis (SFRA). Procedia Eng 2013;68:469–76. https://doi.org/10.1016/j.proeng.2013.12.208.
[24] Schroeder MR. New method of measuring reverberation time. J Acoust Soc Am 1965;37(6):409–12. https://doi.org/10.1121/1.1909343.
[25] Virmani D, Taneja S, Malhotra G (2015) Normalization based K means Clustering Algorithm. arXiv preprint arXiv:1503.00900.