Machine learning algorithms for fault data analytics in power transformers: a literature review

Vitor Esteves[a,*], José Pinto[a] and Ricardo Sousa[a]

[a]LIAAD-INESC TEC, FEUP Campus, Street Dr. Roberto Frias, Porto, 4200 - 465, Portugal

ARTICLE INFO

Keywords:
Literature Review
Fault Diagnosis
Machine Learning
Power Transformer
DGA
FRA

ABSTRACT

One of the most crucial elements of electrical system networks is the power transformer. The failure of this component can be extremely damaging, which is leading to mounting research into fault diagnosis and early warning systems. Given the amount of research, attaining a good grasp of the field can be difficult and time consuming. Thus, to provide a practitioner or researcher with a good starting point, throughout this review we present works pertaining to the usage of machine learning systems for fault data analytics. Furthermore, a discussion of the most relevant methods, either due to proven efficacy and widespread usage or due to innovation and results, as well as, a more in-depth analysis of the most common data sources, namely Dissolved Gas Analysis (DGA), Frequency Response Analysis (FRA) and other uncommon methods and techniques are also presented.

1. Introduction

The power transformer is one of the key elements in the electrical system network. As such, a fault scenario might have a significant impact on the ability of the electrical grid to function properly Rexhepi (2017). Guaranteeing a stable operation of this component is a critical process since replacing a faulty asset is expensive, time-consuming and may degrade the quality of generated power.

Several studies about simple and convenient methods for fault diagnosis were published throughout the years, which is expected due to the importance of preserving power transformers Xie, Zhao and Hong (2020). Meanwhile, with the modernization of industries moving into a new age of development, Industry 4.0, companies are adopting smart systems with intelligent decision-making capabilities in order to improve and optimize current production by making it more customizable and reliable Pasi, Mahajan and Rane (2020). Resort to industrial artificial intelligence, portrayed as the "catalyst to smart manufacturing", has changed how we employ traditional methods in daily activities. The power transformer technical assessment is no exception.

In more recent years, to facilitate the process of identifying potential faults in transformer activity, a task that was usually carried out by technical experts who manually evaluated the condition of the power transformer, machine learning was applied to the domain, as both a mean for prediction of faults and for support to expert decision Khalyasmaa (2019).

In this review paper, we will be analysing how these common techniques can be paired with artificial intelligence by reviewing the current state-of-the-art machine learning algorithms employed in fault detection for power transformers.

The rest of the paper is organized in the following way: Section 2 will delve into the process of selecting relevant articles to the domain explored, including details about the search procedure. Section 3 will give a brief overview of the most often used techniques for fault detection where machine learning can be applied. Section 4 will start by giving an explanation about the machine learning algorithms applied to fault detection and will move into an analysis about which ones thrive in such scenarios. Section 5 will wrap up by discussing the results and summarizing the contribution of this paper.

2. Literature Review

With the recent advances in the field of power transformers and the strong push for the application of AI, driven by Industry 4.0, a necessity on the review of state of the art methodologies is propelled. Thus said, this literature review will provide an overview on the technical aspects of machine learning algorithms applied to this domain, with the

ORCID(s): 0000-0002-2324-5719 (V. Esteves)

objective of expanding the knowledge base of specialists whose area of expertise is generally more related with the energy sector but will benefit from the implementation of these types of tools.

2.1. Methodology

According to Aaron (2008) any methodology section on a literature review should contain a description on how the search was conducted. The first step was obtaining information regarding the topic using Scopus and, to do so, a list of keywords and a combination between them was created. The term “Power Transformer” was grouped with “Machine Learning” as an attempt to have a better grasp about how smart systems were used in the context of power transformers. The input introduced, specifically, TITLE-ABS-KEY ("machine learning" AND "power transformer") which searches within titles, abstracts and keywords was done on 19/05/2021. This search yielded 126 results that were further reviewed to answer the previous contemplation. Step two was pinpointing the more frequent uses, which included, as expected, fault detection. By narrowing said input by adding the keyword “fault detection” the results were filtered to 55 articles. At this point, before neither discarding nor reviewing any articles, the need urged to better understand the domain from the mechanical point of view and, as such, another search was conducted, this time with the following input: TITLE-ABS-KEY ("Power Transformer" AND "Fault Detection"). This search yielded 802 articles, a much broader result reflective of an ever evolving field of study. From these, only surveys and literature reviews were analyzed. This deeper search also allowed for a better understanding of the classification of faults, taxonomies on fault diagnosis and main causes of disruptions. From this analysis, some problems and data types were far more prevalent, thus a second search was performed, aiming at balancing the content. This search specifically consisted of TITLE-ABS-KEY ("machine learning" AND "FRA") and TITLE-ABS-KEY ("machine learning" AND "Vibro Acoustic"). In the last phase, after reviewing and discarding irrelevant information, categorization of papers’ algorithms by technique type and reference analysis brought the total of articles for this review to 31.

3. Fault detection techniques for power transformers

Faults that can disturb the good functioning of a power transformer can be classified into two types, internal or external. Internal faults can then be divided into two subtypes: incipient faults and short circuit faults Yadaiah and Ravi (2011). According to Faiz and Heydarabadi (2014) and the IEEE Guide for Protective Relay Applications to Power Transformers, external faults may occur due to an external short circuit of the power system, or when subjected to inappropriate conditions such as overload or overvoltage while internal transformer faults include winding faults which encompass interturn, layer to layer, phase-to-ground and phase-to-phase faults, tank and oil problems such as oil oxidation, water penetration or dissolution caused by increased temperature, core faults, terminal faults in particular short circuit, bushing faults or turn on the switch faults and lastly other unspecified types of faults, namely cooling system faults. Between 70-80% of power transformer damages arise from internal faults Faiz and Heydarabadi (2014).

The most common fault detection methods for this type of problem which have historically proven to be effective and whose application is linked to machine learning, are Dissolved Gas Analysis (DGA) and Frequency Response Analysis (FRA). More recently, in Bartoletti, Desiderio, Di Carlo, Fazio, Muzi, Sacerdoti and Salvatori (2004) Vibro-Acoustic techniques have been employed successfully in fault and defect location.

3.1. Dissolved Gas Analysis

According to Faiz and Soleimani (2017); Yadaiah and Ravi (2011); Sun, Huang and Huang (2012) Dissolved gas analysis (DGA) is a technique used to examine oil filled components of a power transformer. While running under normal conditions, the insulation materials of the electrical equipment in the power transformer degenerate slowly and liberate a certain amount of gas over time. The occurrence of faults accelerates this process, therefore resulting in a higher concentration of these compounds, with the rate at which they are produced indicating the severity of the fault. The generated gases from eventual decomposition are Hydrogen (H_2), Carbon dioxide (CO_2), Carbon monoxide (CO), Acetylene (C_2H_2), Ethylene (C_2H_4) Ethane (C_2H_6) and Methane (CH_4). The amount of gas concentration is linked to the number of years the transformer has been operating and the following table represents the maximum amount (μL/L) per activity time.

The type of fault that can be detected is usually not directly inferred through any of the values separately, but by inspecting the ratios of a combination of certain gases. There are several techniques for ratio analysis, with the most common being the Doernenburg Ratio Dornenburg and Strittmatter (1974), the Rogers Ratio Rogers (1978) which are IEEE std C57.104 interpretations IEEE (2019) and the Duval’s Triangle Duval (2003) which is an IEC 60599 interpretation IEC (2015).

Table 1
Table of expected DGA values depending on transformer age.

Gas	<4 years	<10 years	>10 years
CH_4	70	150	300
C_2H_4	150	200	400
C_2H_6	50	150	1000
C_2H_2	30	50	150
H_2	150	300	300
CO	300	500	700
CO_2	3500	5000	12000

Table 2
Table for Dornenberg Ratio interpretation.

Dissolved Gases (Gas Extracted from Oil)
Dornenberg Method

Fault Type \ Ratio	$\frac{CH_4}{H_2}$	$\frac{C_2H_2}{C_2H_4}$	$\frac{C_2H_6}{C_2H_2}$	$\frac{C_2H_2}{CH_4}$
Thermal decomposition (hot spots)	>1.0	<0.75	>0.4	<0.3
Electrical discharges (except corona)	>0.1 <1.0	>0.75	<0.4	>0.3
Corona	<0.1	not significant	>0.4	<0.3

3.1.1. Doernenburg Ratio

By using the Doernenburg Ratio, the proportions of gases considered are CH4 / H2, C2H2 / C2H4, C2H6/C2H2 and C2H2/CH4 (mol/L or ppm v/v) and it is possible to identify thermal decomposition, electrical discharges and corona.

3.1.2. Rogers Ratio

The Rogers Ratio method uses the same methodology as the Doernenburg Ratio but also defines the values for a base case of normal functioning of the power transformer. The fault diagnosis array of possibilities is expanded and consists of low/medium/high thermal faults, partial and continuous discharges and arching.

3.1.3. Duval's Triangle

Duval's Triangle relies on visualization and can be used to support the other methods. By plotting a triangle using only the total concentrations of methane (C2H4), ethylene (C2H4) and acetylene (C2H2) as axes, between 0 to 100%, it's possible to identify faults as seen in the table and example below.

The area of the section represents the probability of the fault occurring and to identify the point as coordinates, refer to the below equations:

$$\%C_2H_2 = C_2H_2 / (C_2H_2 + C_2H_4 + CH_4)$$
$$\%C_2H_4 = C_2H_4 / (C_2H_2 + C_2H_4 + CH_4)$$
$$\% CH_4 = CH_4 / (C_2H_2 + C_2H_4 + CH_4) \quad (1)$$

Table 3
Table for Rogers Ratio interpretation.

Suggested Diagnosis from Gas Ratios—Rogers Ratio Method

$\frac{CH_4}{H_2}$	$\frac{C_2H_6}{CH_4}$	$\frac{C_2H_4}{C_2H_6}$	$\frac{C_2H_2}{C_2H_4}$	Suggested Diagnosis
>0.1 <1.0	<1.0	<1.0	<0.5	Normal
⩽0.1	<1.0	<1.0	<0.5	Partial discharge — Corona
⩽0.1	<1.0	<1.0	⩾0.5 <3.0 or ⩾3.0	Partial discharge — Corona with tracking
>0.1 <1.0	<1.0	⩾3.0	⩾3.0	Continuous discharge
>0.1 <1.0	<1.0	⩾1.0 <3.0 or ⩾3.0	⩾0.5 <3.0 or ⩾3.0	Arc — With power follow through
>0.1 <1.0	<1.0	<1.0	⩾0.5 <3.0	Arc — No power follow through
⩾1.0 <3.0 or ⩾3.0	<1.0	<1.0	<0.5	Slight overheating — to 150°C
⩾1.0 <3.0 or ⩾3.0	⩾1.0	<1.0	<0.5	Overheating 150–200°C
>0.1 <1.0	⩾1.0	<1.0	<0.5	Overheating 200–300°C
>0.1 <1.0	<1.0	⩾1.0 <3.0	<0.5	General conductor overheating
⩾1.0 <3.0	<1.0	⩾1.0 <3.0	<0.5	Circulating currents in windings
⩾1.0 <3.0	<1.0	⩾3.0	<0.5	Circulating currents core and tank; overloaded joints

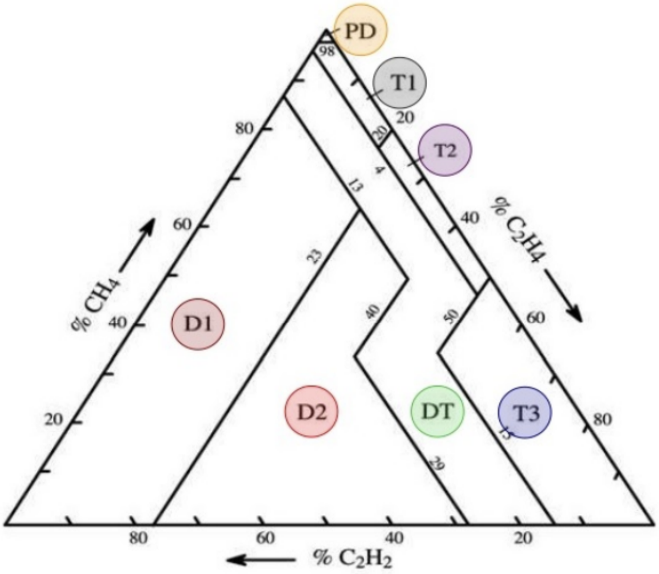
Table 4
Duval's triangle fault interpretation.

Region	Fault
PD	Partial discharge
D1	Discharges of low energy
D2	Discharges of high energy
T1	Thermal fault, t < 300°C
T2	Thermal fault, 300°C< t < 700°C
T3	Thermal fault, t > 700°C
DT	Mixtures of thermal and electrical faults

3.2. Frequency Response Analysis

Today, Frequency Response Analysis (FRA) is one of the most well-established and accurate power transformer mechanical fault detection and diagnostics tools, with very solid foundations both in the research space, as well as in industry practical applications and boasting a great degree of standardization Nurmanova, Akhmetov, Bagheri, Zollanvari, Gharehpetian and Phung (2020); Mao, Wang, Crossley, Jarman, Fieldsend-Roxborough and Wilson (2020). Originally, FRA was developed with the goal of identifying winding mechanical movements and deformation, a task for which it still is commonly employed; with this first variant of FRA being known as sweep FRA Zhao, Tang, Zhou, Xu, Gui and Yao (2017); Mao et al. (2020); Mao, Wang, Jarman and Fieldsend-Roxborough (2019). With this, FRA quickly gained attention as one of the most effective methods for winding deformation and movement detection, having multiple

Figure 1: Duval's triangle.



advantages, such as, detecting faults at early stages, being simple to perform, inexpensive and non-invasive Liu, Zhao, Tang, Yao, Li and Islam (2019); Nurmanova et al. (2020); Zhao et al. (2017).

Transformer windings are susceptible to many forms of mechanical faults, such as unintended movement, deformation (such as tilting, forced bulking, free buckling, hoop tension, telescoping, etc.); which can lead to total transformer failure Zhao et al. (2017); Mao et al. (2020). Furthermore, short-circuit faults can induce much greater forces than expected on the winding, leading to deformation, which in turn makes short-circuits more likely to occur, thus leading to a positive feed-back loop. The detection of these faults is very hard to at early stages, due to the fact that they have very little impact on transformer operations, which makes FRA (that can do so) all the more important Zhao et al. (2017).

In general, FRA works by injecting a set of low-voltage sinusoidal signals at one end of the winding and reading the response at another, with the gathered data consisting of the magnitude ratio and phase shift between the injected and obtained signals Liu et al. (2019); Mao et al. (2020). These measurements are typically done in the frequency range between 20Hz and 2MHz, with multiple possible configurations, such as end-to-end open-circuit test, end-to-end short-circuit test, capacitive inter-winding test and inductive inter-winding test Nurmanova et al. (2020). A more recent version of FRA, impulse FRA works by the same mechanism but employing a wider range of frequencies, from a few Hz to several MHz, with it being employed to build the frequency response signature Zhao et al. (2017); Mao et al. (2020). The advantage of impulse FRA over sweep FRA is that it has a higher signal to noise ratio and smaller energy injection, making it more feasible to perform in an online setting Zhao et al. (2017). In general, going from low to high frequencies, FRA traces are, in order dominated by, the core, winding interactions, winding-under-test, and the measurement leads and set-up Mao et al. (2019).

For the analysis of these signals, there are several alternatives. The most common is known as time basis, where the measured signal is compared to an healthy, original signal obtained at the time of manufacture, at the factory, known as fingerprint Nurmanova et al. (2020); Contreras, Sanz-Bobi, Banaszak and Koch (2011); Liu et al. (2019); Mao et al. (2020, 2019). Two other types include type basis, where the signal is compared to that of identical units; and phase basis, where different frequency responses of different phases are compared Nurmanova et al. (2020). As time based is by far the most common, it is the one that will be elaborated upon.

This comparison to the fingerprint is still by no means simple, with there also being multiple competing alternatives. One common approach is based on the fact that windings have equivalent electrical circuits Liu et al. (2019). These circuits are built according to resistance, capacitance, and inductance in a high frequency range Liu et al. (2019); Zhao et al. (2017); Mao et al. (2020, 2019). The most common approach is the analysis of a magnitude frequency spectrum, for which multiple methods exist, such as visual inspection by an operator, statistic indicators, and transfer function expressions Zhao et al. (2017); Nurmanova et al. (2020); Mao et al. (2020).

Any discrepancy to the fingerprint can indicate a mechanical fault, with different frequency response signatures representing the winding status, that can be used not only to identify the existence of a failure, but also its type and even severity Liu et al. (2019); Zhao et al. (2017); Mao et al. (2020). This is possible due to the fact that faults of the same type tend to present similar characteristics, with identifiable conditions including healthy, short circuit between turns and mechanical deformation Liu et al. (2019); Contreras et al. (2011).

Despite the large variety of FRA analysis methods, none are perfect or the best in all situations, which leads to some, still in use, mechanical fault FRA alternatives. These are ones such as short-circuit impedance, which, being based on the principle of short-circuit impedance measurement, compares the short-circuit impedance phase of the winding to the nameplate or factory test results Liu et al. (2019). Another option is low-voltage impulse, which being based on the principal of signal analysis, compares time domain signals of the winding before and after a detected fault Liu et al. (2019).

The aforementioned limitations of FRA mostly lie in the ability to interpret the signals. As visual analysis of response plots by an operator is common, there is a great degree of subjectivity, which is increased by the complexity and difficulty of the task Nurmanova et al. (2020); Contreras et al. (2011); Zhao et al. (2017). Another difficulty is that statistical mathematical calculations are not very accurate when identifying fault type and severity, compounded with different winding types presenting wildly varying FRA characteristics Liu et al. (2019); Mao et al. (2020). Combining this with the fact that most operating personnel do not have access to the winding type information turns FRA analysis into an extremely complex challenge Mao et al. (2020, 2019). Because of this, it is no surprise that machine learning methods have been garnering increased attention, in particular with the problem of identifying winding type from FRA results, which is in itself a hard task, as gathering large complete FRA datasets with winding information, deformation types, severity and location is not easy to perform Mao et al. (2019); Zhao et al. (2017).

3.3. Vibro Acoustic Methods

Vibro acoustic methods can be divided into vibration-based methods and Acoustic Emission (AE) based methods. Due to their interactions, with each interfering or generating another (sound is in effect just a vibration), as well as the similarities in target problem and approach, they are in most cases handled simultaneously and interchangeably.

These methods are typically used to tackle two types of problems. The first, and arguably more prevalent, is partial discharge (PD) identification and diagnosis, while the second is on-load tap-changer (OLTC) failure detection. The power transformer is a source of many low frequency signals (Important ones tend to lie in the 10 Hz to 100 Hz range Jancarczyk, Bernaś and Boczar (2019)), from which those resulting from PD play a small role. PDs are nevertheless one of the biggest threats to insulation systems, with their detection being crucial but difficult Kunicki and Wotzka (2019); Jancarczyk et al. (2019). This difficulty stems from the fact that most of the signals are blocked or reflected by the internal and external transformer components, with as little as 1% actually reaching the sensor Kunicki and Wotzka (2019). To compound this difficulty, signals can be masked by other transformer operations, such as, transformer vibration, fans motor noise, pumps, noise of tap selector and diverter switchViereck and Saveliev (2019); Woon, El-Hag and Harbaji (2016). However, this also means that other possible defects and problems can be identified Kunicki and Wotzka (2019).

The primary advantage of these methods is that they are noninvasive and measuring devices can be installed and operated even when a power transformer is in an energised state Wotzka and Cichoń (2020); Kunicki and Wotzka (2019). AE measuring devices are typically magnetically placed in the transformer tank wall Woon et al. (2016); Woon, Aung and El-Hag (2018). Thus, AE sensors have the advantages of being very inexpensive with simple installation and permitting remote operation, as such not requiring an on-site engineer Woon et al. (2016, 2018). Vibration sensors share many of the same characteristics, similarly being inexpensive and simple to operate.

Of course, although these sensors have their advantages, there are other alternatives with their strengths and weaknesses. A common alternative is ultra-high-frequency (UHF), which is far more sensitive to electromagnetic pulses, while being less prone to external noise, at the cost of being far more expensive to install and operate Kunicki and Wotzka (2019); Woon et al. (2016). A disadvantage of vibration sensors is that the gathered data can be far more complex, requiring a very in-depth understanding of the domain and operating principles Viereck and Saveliev (2019).

This raw data gathered from vibro acoustic sensors is typically complex and difficult to analyse or employ in most methods, with multiple processing approaches are employed. These include short time Fourier transform for spectrogram analysis Jancarczyk et al. (2019), discrete and continuous wavelet transformations, as well as Hilbert, and Gabor transformationsWotzka and Cichoń (2020); Woon et al. (2016).

Overall, vibro acoustic methods are inexpensive, can be continuously monitored without disruption to the transformer operations, are simple to install and provide large amounts of data. However, due to the complexity of said data, the interference of multiple systems, noise and large measuring error margins, the usage of intelligent machine learning data-driven methods is a must.

4. Machine Learning Algorithms and its application on fault detection

Machine learning is one of the most important branches of artificial intelligence and its concept is based on systems being given access to data that they can learn from in order to make decisions without human intervention. It can be divided into three main classes of techniques: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is a technique where the main objective is learning to approximate a function (f) from pairs of inputs and outputs, so that the input (x) can match the output (y) when given new information, that is, finding y=f(x). Supervised learning then can be divided into two main groups, classification and regression, with the main difference here being the output variable type, which is categorical for the former and numerical for the latter. In power transformer fault detection, classification is the most common approach followed since we want to predict if a fault has occurred. Unsupervised learning is a technique used to attempt to create labels in data, such as clustering, where the objective is grouping unlabeled data that share similarities in their features. Although some authors only distinguish between the above two classes, reinforcement learning is widely included in the discussion due to algorithms' ability to identify patterns that will later on change taking into account the conditions of the environment it was built on. Thus said, a literature review about machine learning techniques was compiled to understand how any of the aforementioned techniques relates to power transformer fault diagnosis and detection. This review was divided by each of the traditional techniques which were described before:

4.1. Dissolved Gas Analysis

DGA is by far the most common data source used for fault prediction and diagnosis in power transformers. Having already proven its worth in identifying and discriminating a broad selection of possible faults with classical approaches, it is now subject to a great amount of machine learning experiments and research.

Many works utilize DGA gas values directly. Such is the case for the work in Zheng and Shioya (2020), which uses 7 gases for fault diagnosis in Oil-immersed power transformers, where a Multi Layer Perceptron (MLP) model is created, achieving good results.

The work in Su, Cui, Yu, Chen, Xu, Hou and Sheng (2019) uses 5 gases for fault prediction and diagnosis, not only identifying 8 possible faults but also possibly outputting an healthy state. For this is used an hierarchical classification model, splitting the problem region into multiple spaces, where each region gets assigned either a MLP model or an EasyEnsemble model. The new method is shown to achieve better accuracy than MLPs, Random Forests (RF) and Support Vector Machines (SVM).

A similar problem is tackled in Cheemala, Asokan and Preetha (2019) where 7 gases are used for 7 fault type or healthy state prediction. Here, 3 different common methods are compared, SVMs, K-Nearest Neighbours (kNN) and a bagged tree ensemble. It was concluded that preprocessing the DGA gases with a logarithmic transformation, rather than standardization, as well as providing balanced classes can greatly improve results. Furthermore, the ensemble method was shown to achieve the best accuracy.

The task of fault diagnosis was undertaken in Kaur, Brar and Leena (2019), where 5 gases were used to predict 6 fault types (but not a healthy state). The Random Neural Network (RNN) model was selected and two optimization methods were contrasted, Levenberg-Marquardt and Broyden-Fletcher-Goldfarb-Shanno. From these the Broyden-Fletcher-Goldfarb-Shanno optimization method yielded the better results.

In Basuki et al. (2018) fault diagnosis (no healthy state) is performed, predicting from 9 possible faults and employing 7 gases for the task. The work presents the contrast between a simple Machine learning method, C4.5 Decision Trees (DT) and 5 classical methods; Rogers ratio, Duvals triangle, IEC table, doernenburg and key gas. The results show that even a simple and interpretable DT is capable of better classification accuracy than the standard methods.

The task of diagnosing between 5 faulty states is undertaken in Dong, Yang, Li, Xie and Zuo (2019), where a new Artificial Neural Network (ANN) optimization method, Modified Bat Algorithm (MBA), is tested on data containing 5 gases. This new method is contrasted with another 3 exotic optimization methods, Bat Algorithm (BA), Particle Swarm Optimization (PSO) and Genetic Algorithm (GA), with the new one achieving the best results.

Another novel method for fault diagnosis and detection is presented in Huang, Wang and Tian (2018), where 5 gases are used to differentiate between 5 fault types or a healthy state. This novel method entails a Kernel Extreme Learning Machine (KELM) optimized via the grey wolf algorithm, whose parameters are initialized with a logistic chaotic map. This method was proven to achieve better results than a simple Extreme Learning Machine (ELM) or an ANN to which it was compared.

Like in most classical methods, rather than gas concentrations, as in the previous presented work, it is quite common for gas ratios to be used for model creation. Soto, Lima and Saavedra (2019) used 4 gas ratios to predict amongst 5 fault types. This work employs data generated by the mean shift algorithm to compare 3 variants of MLP backproagation. All variants achieved perfect results (100% in validation), which raises some concerns regarding the use and validity of the mean shift data generation.

In Shang, Xu, Zheng, Qi and Zhang (2019) 5 gas ratios are used to predict from 5 fault types or healthy state. A new method is developed where PSO is used to optimize the parameters of a Hypersphere Multiclass SVM (HMSVM), thus creating the PSO-HMSVM method. The method was first contrasted with an hybrid immune algorithm and a KELM, achieving improved results. Finally, improved D–S evidence theory (IDET) was used to combine all 3 methods, obtaining even better results.

Using a novel method, Multiscale Information Fusion-based Extreme Learning Machine (MIF-ELM), Huang, Huang, Long and Long (2019) perform fault identification and diagnosis of 6 fault types and healthy state. A total of 3 gas ratios are used, showing that the new method outperforms both ELM and ANN models.

Li, Yang, Dong, Xie and Yang (2018) is last work presented that utilizes gas ratios. Fault diagnosis is performed with 5 fault types being predicted from 3 gas ratios. Modified Cuckoo Search (MCS) was used to optimize an ANN creating the MCS Back Propagation (MCS-BP) optimization method. This optimization method was compared with other optimization methods, BP, Cuckoo Search (CS-BP), PSO-BP and GA-BP, as well as, Multiverse Optimized-Multi-Layer Perceptrons, SVMs and Probabilistic Neural Networks (PNNs), achieving the best results.

The only work found using exclusively gas percentage values is Islam, Lee and Hettiwatte (2017), where 5 percentages are used to predict 7 fault types. A Duval triangle-kNN hybrid is developed, where the kNN makes the predictions for values close to Duval's prediction boundaries. This new method achieves better results than the classical ones it was compared to, Duvals triangle, Rogers ratio and IEC table.

In some instances, other data sources are used to complement the DGA data, providing further context into the operating conditions. This can be seen in the work Yang, Chen, Li and Yang (2020), where besides 5 gas values; transformer capacity, temperature and humidity conditions are also present. This work pertains to oil-immersed power transformer fault diagnosis, where the Multi-Verse Optimizer (MVO) algorithm is used to optimize a PNN, creating an hybrid method, MVO-PNN. This method was compared with MVO-MLP, GA-PNN, PSO-PNN, BA-BP, CS-BP, GA-BP, IEC and PNN, achieving the best accuracy of 94%.

In Katooli and Koochaki (2020), on top of 5 gas values; 5 gas ratios, 3 gas proportions and a total gas concentration sum are used for fault detection and diagnosis, predicting 3 faulty conditions or healthy state. Here is employed an hybrid method where an association rule mining technique is used for feature selection and an Adaptive Neuro-Fuzzy Inference System (ANFIS) is optimized via Black Widow Optimization (BWO) to create the Opt-ANFIS method. This method was compared in 2 databases with Roger ratio, Doernenburg's ratio, SVM, Radial Basis Function Neural Network (RBFNN), MLP and PNN, achieving the best results. It was also compared with two state of the art methods, Binary Multi-Objective Particle Swarm Optimization (BMOPSO) and a MLP trained by a Time-Varying Acceleration Coefficient Version of Particle Swarm Optimization (TVACPSO), while achieving better results and requiring less computational time.

A 5 gas dataset is expanded with n-octane (C_8H_{18}) and eicosane ($C_{20}H_{42}$) values in Taha, Dessouky and Ghoneim (2021). Other than fault diagnosis for which 6 fault types are predicted, the problem of fault severity prediction is also separately tackled. In this work, multiple Neural Pattern Recognition (NPR) models are arranged in a tree-like manner, subdividing the problem. A 2-stage scenario using 3 models and a 3-stage one with 5 models were tested. The new models were shown to outperform Duvals triangle, Rogers ratio, IEC table, California State University Sacramento method (CSUS), clustering, conditional probability and SVM in the tasks of fault diagnosis and severity prediction.

Unlike works presented previously, were multiple DGA gas formats are used in tandem, in Benmahamed, Teguar and Boubakeur (2018), gases in ppm, in percentage, Dornenberg ratios, Rogers ratios and Duval values are alternated and tested individaully. Here, PSO was used for SVM optimization, creating the PSO-SVM method, which was compared with kNN. It was concluded that the PSO-SVM method achieved the best results while using Duvals DGA values.

In the work Guo, Dong and Wu (2019), 7 gases are complemented with the total hydrocarbon amount to differentiate between 5 fault types. ANNs, PNNs and DTs are compared, with the DT model achieving the best results.

The final work that we will present that utilizes DGA gases as its core data is Li, Wu, Gao, Hao, Xin and Yin (2016), where complementing 5 gases; 3 gas ratios are present. In this work, the task of fault diagnosis and identification is undertaken, with the prediction of 6 fault types or healthy state. A new method is presented involving the preprocessing of the data with an arctangent transformation, later being fed into a self-adaptive evolutionary extreme learning machine model. It is concluded that the new model achieves better results than ANN, SVM and ELM models.

4.2. Frequency Response Analysis

Unlike DGA, which is used to predict from a variety of fault types, FRA is far more effective for the prediction of winding related problems and faults. In Contreras et al. (2011), a DT method is employed on FRA data to predict from 2 healthy and 2 unhealthy winding states, with good and interpretable results being achieved.

The task of differentiating between 3 winding fault types is presented in Zhao et al. (2017). A SVM model is trained on a variant of FRA, Impulse FRA, where although good results are obtained, no comparisons to other methods are presented.

In Nurmanova et al. (2020) the identification, diagnosis and severity prediction of short circuit faults is performed. Here a new FRA measurement configuration was tested, while sequential forward search (SFS) was employed for feature selection and several interpolation models were compared for prediction. It was concluded that piecewise cubic hermite interpolation attains better results than linear and cubic spline interpolation.

Despite a lot of search for scientific work related to the use of Machine Learning (ML) in FRA data for fault diagnosis, only a total of 4 examples were found, which when contrasted with the DGA, for which the presented research is still far from exhaustive, indicates a large imbalance and a search space almost totally uncharted. In the last identified work, Liu et al. (2019), a PSO optimized SVM for fault diagnosis is presented, where 3 fault types and various fault severity values are predicted, obtaining good results.

4.3. Vibro Acoustic Methods

Similarly to FRA, Vibro acoustic methods have been proven to be more effective in very specific scenarios. While data from Acoustic Emissions (AE) can better detect Partial Discharge (PD) fault types, vibrational data can better identify mechanical failure.

A variety of ML methods are tested in Kunicki and Wotzka (2019) for the diagnosis of PD faults and identification of other fault types. The methods used consist of fine trees, SVMs, kNNs and bagged DTs, which were trained on AE data. From these, SVMs achieved the best results.

The work Woon et al. (2016) performs fault diagnosis by distinguishing from 3 types of PD. AE data is used by DTs, gradient boosted DTs, RFs, SVMs and linear discriminant analysis. To verify the generality of the models, the training and testing of models was attempted in different datasets which present different operating conditions. For same dataset training and testing SVMs attained the better results, while for the transfer to another dataset RFs surpassed them. Regardless, SVMs, RFs and Boosted DTs achieved very similar results in all tests.

A continuation of the previous work by some of the same authors is presented in Woon et al. (2018). Here the same set of models and datasets were used to test a new method, one-dimensional convolutional neural networks. Although the new model did not achieve any significant level of performance improvement over the others, due to its nature, the difficult and domain specific feature extraction step could be skipped, saving complexity and simplifying the work.

Two problems, using AE data, were tackled in the work Wotzka and Cichoń (2020). The first entails the detection and diagnosis of On Load Tap Changer (OLTC) defects. The second involves determining the best type of AE sensor to detect these fault types. Multiple variants of SVMs, DTs, kNNs and model ensembles were tested and compared. It was concluded that an ensemble subspace discriminant is capable of obtaining the best results, while the best sensor type is a WD sensor.

The final work regarding vibro acoustic methods, and the only one to use vibrational sensors is presented in Tamilselvan and Wang (2013). This work presents a deep belief network which is composed of multiple stacked restricted Boltzmann machines, each trained at at time. The model was trained on data from 9 simulated vibration sensors to diagnose also simulated joint failures. The method was compared with SVMs, ANNs, self-organizing maps and Mahalanobis distance, obtaining the best accuracy.

4.4. Other data sources

There are, of course, instances of other more uncommon data sources being used for fault diagnosis. In this subsection we will identify these data sources, the works that employ them and the approaches presented.

PD waveforms are employed in Jia and Zhu (2018) for PD failure diagnosis, where 4 types of PD are discerned. Two methods for PD feature extraction are compared, pulse shape characterization and phase resolve PD. Variable Predictive Model-based Class Discrimination Kernel Partial Least Squares (KPLS-VPMCD) is also contrasted with ANNs, SVMs, and VPMCD. For all models, pulse shape characterization achieved better results, while the new model outperformed the others in all cases.

A work with many similarities to the previous is presented in Kartojo, Wang, Zhang et al. (2019), where PD waveforms are used to predict from amongst 3 types of PD. RFs, SVMS, kNNs and linear discriminant analysis were compared, with RFs obtaining the best accuracy.

The work in Li and Ma (2019) is a strange case, for the actual data content is not clearly stated. Nevertheless, for completeness, we still decided to present this work, where 9 fault types or healthy state are predicted. An Online Sequence Extreme Learning Machine (OS-ELM) optimized via improved PSO, thus creating the IPSO-OS-ELM method, is presented. The new method was contrasted with OS-ELM and ELM, achieving the best results.

The final work that we will detail in this chapter is Raichura, Chothani and Patel (2020), where current data is utilized to diagnose different fault types (internal, external, etc.). Current data feature extraction was performed with a discrete wavelet transform. A Hierarchical Ensemble Extreme Learning Machine (HE-ELM) was contrasted with SVMs, ELMs and PNNs, with it attaining the best accuracy.

5. Discussion

In the previous section we presented many works using ML for fault detection and diagnosis. Given the high amount of information, it might be difficult for a reader to extract the most important details and to either extend the work presented with their own contribution or put the methodologies detailed into practice. Thus, now we will present the main takeaways from the work aggregated and presented.

First, it is important to analyse the most common methods employed, which can be seen on table 5, to present those which are well established and which have a high likelihood of success in a large array of problems. Overall, there are 5 methods which are quite common for DGA interpretation, ANNs/MLPs, SVMs, DTs, kNNs and ELMs. From these DTs and SVMs tend to obtain the best results, while as a bonus, DTs create very interpretable results. When the methods are compared, in none of the works gathered, do simple ANNs/MLPs or ELMS obtain the best results; however, with these coming very close, testing them could still be a good idea. On the other hand, kNNs presented by far the worst results in all works when contrasted to other methods, so we do not believe it to be useful to utilize them in their standard form.

A second type of models, which is still quite prevalent in the works gathered, but not as much as the previous ones, is DT ensembles, whether bagging or boosting, which tend to show similar or even superior results. RFs in particular have shown a tendency for obtaining amongst the best results, while tending to not overfit data. In fact, by some, DT ensembles in their various forms, are considered the state of the art for tabular data.

The third type of methods that we will discuss are those that although obtain great results, are only present in at maximum a couple of works, and which are not as widespread in the ML community at large.

Both MVO-PNN and Opt-ANFIS where contrasted with a large number of other methods, while still obtaining the best results, thus being the most promising. The unnamed hierarchical classification method presented in Su et al. (2019), using a MLP model or an EasyEnsemble obtained better results than the staple models previously presented. Self-adaptive evolutionary extreme learning machines, MIF-ELM, KELM optimized via the grey wolf algorithm and the NPR tree model, presented in Taha et al. (2021), surpass some of the standard methods, being worth looking at. Finally, the PSO-HMSVM method and the combination of it with hybrid immune algorithm and a KELM, presented in Shang et al. (2019), also seem promising.

There are also a few works which focus not on the exact model used, but on the optimization algorithm employed. The Broyden-Fletcher-Goldfarb-Shanno optimization method for RNNs shows promise, while MBA for ANN optimization provides significant improvements.

Finally, we will also present the methods which we believe, that without further research show little promise. For these, either the amount or quality of comparisons made was found lacking, or the results seem dubious. Given this, both the Duval triangle-kNN hybrid and PSO-SVM were found to be of little interest.

Table 5
Table of DGA articles.

Article	Data	Best model	Other Models
Taha et al. (2021)	5 DGA gases, C8H18, C20H42	Tree arranged NPR	Duvals triangle, Rogers ratio, IEC, CSUS, clustering, conditional probability, SVM
Zheng and Shioya (2020)	7 DGA gases	MLP	None
Yang et al. (2020)	5 DGA gases, temperature, capacity, humidity	MVO-PNN	MVO-MLP, GA-PNN, PSO-PNN, BA-BP, CS-BP, GA-BP, IEC, PNN
Katooli and Koochaki (2020)	5 DGA gases, 5 DGA ratios, 3 DGA proportions, total DGA	Opt-ANFIS	Doernenburg's ratio, SVM, RBFNN, MLP, PNN, BMOPSO, TVACPSO
Su et al. (2019)	5 DGA gases	hierarchical classification with MLP and EasyEnsemble	MLP, SVM and RF
Cheemala et al. (2019)	7 DGA gases	Bagged tree ensemble	SVM, kNN
Kaur et al. (2019)	5 DGA gases	Broyden-Fletcher-Goldfarb-Shanno optimization for RNN	RNN with Levenberg-Marquardt optimization
Guo et al. (2019)	7 DGA gases, total hydrocarbon	DT	ANN, PNN
Soto et al. (2019)	4 DGA gas ratios	IDET fused PSO-HMSVM, hybrid immune algorithm and KELM	PSO-HMSVM, hybrid immune algorithm, KELM
Huang et al. (2019)	7 DGA gas ratios	MIF-ELM	ANN, ELM
Dong et al. (2019)	5 DGA gases	MBA-MLP	BA-MLP, PSO-MLP, GA-MLP
Huang et al. (2018)	5 DGA gases	Grey wolf algorithm optimized KELM	ELM, ANN
Basuki et al. (2018)	7 DGA gases	DT	None
Li et al. (2018)	3 DGA gas ratios	MCS-BP	BP, CS-BP, PSO-BP, GA-BP, Multi-verse Optimized MLP, SVM, PNN
Benmahamed et al. (2018)	DGA ppm, %, Dornenberg, Rogers, Duval	PSO-SVM	kNN
Islam et al. (2017)	5 DGA %	Duval triangle-kNN hybrid	Duvals triangle, Rogers ratio, IEC table
Li et al. (2016)	5 DGA gases, 3 DGA gas ratios	self-adaptive evolutionary ELM	ANN, SVM, ELM

For FRA, given the reduced amount of existing research, which is present on table 6, only the most common staple methods were found to be particularly interesting. Despite PSO-SVM also being present here, the lack of any real comparisons still leads us to not recommend it.

For Vibro Acoustic methods the staple methods are very similar, as can be seen on table 7, with ELM models being less common, while DT ensembles, in particular RFs, gaining far more traction. Here, two standout methods are present. The first is deep belief networks, which outperform most of the staple methods. The second one is one that we believe to be particularly useful, one-dimensional convolutional neural networks, which on top of achieving competitive results, allows the practitioner to skip the laborious and domain specific feature extraction task for time series data, being particularly useful at establishing a quick baseline.

For the fault diagnosis tasks using other types of miscellaneous data, as shown on table 8, the staple methods are once again similar to those found for DGA, with ELM models being far more prevalent. There are 3 standout methods which seem to have potential, KPLS-VPMCD, OS-ELM and HE-ELM.

Table 6
Table of FRA articles.

Article	Data	Best model	Other Models
Nurmanova et al. (2020)	FRA	Hermite interpolation	Linear and cubic spline interpolation
Liu et al. (2019)	FRA	PSO-SVM	None
Zhao et al. (2017)	Impulse FRA	SVM	None
Contreras et al. (2011)	FRA	DT	None

Table 7
Table of Vibro acoustic articles.

Article	Data	Best model	Other Models
Kunicki and Wotzka (2019)	AE	SVM	DT, kNN
Liu et al. (2019)	AE	Ensemble subspace discriminant	SVM, DT, kNN
Woon et al. (2018)	AE	1D convolutional neural network	SVM, DT, gradient boosted DT, RF, linear discriminant analysis
Woon et al. (2016)	AE	SVM, RF	DT, gradient boosted DT, linear discriminant analysis
Tamilselvan and Wang (2013)	Simulated vibration from 9 sensors	Deep belief network	SVM, ANN, self-organizing map, Mahalanobis distance

Table 8
Table of various data source articles.

Article	Data	Best model	Other Models
Nurmanova et al. (2020)	PD waveforms	RF	SVM, kNN, linear discriminant analysis
Raichura et al. (2020)	Current	HE-ELM	SVM, ELM, PNN
Kartojo et al. (2019)	PD waveforms	SVM	None
Li and Ma (2019)	Undisclosed	IPSO-OS-ELM	OS-ELM, ELM
Jia and Zhu (2018)	PD waveforms	KPLS-VPMCD	ANN, SVM, VPMCD

Lastly, it is important to state that the focus of this analysis has been placed on the underlying models and optimization methods, rather that data generation and preprocessing techniques, which can greatly affect results. Given this and the fact that most methods are not tested on the same datasets, we for the most part refrain from presenting and comparing performance metrics, as these cannot be equated.

6. Conclusion

In recent years, the development of manufacturing companies through Industry 4.0 has been growing steadily, especially since the integration of artificial intelligence in common processes. Various machine learning algorithms have been put to service in the automatization of regular tasks such as, fault detection on power transformers. With the purpose of enlightening the future development of the field, this literature review presents a comprehensive approach to techniques that deal with failure detection where artificial intelligence can be applied. Moreover, by concatenating two distinct fields of knowledge such as electrical engineering and artificial intelligence, insights can be provided for a wider range of scientific experts in any of the fields. In Fault Detection for Power Transformers, the study delves into the fundamentals of traditional techniques such as DGA, which have been used throughout the years, and afterwards, in Machine Learning Algorithms and its application on fault detection after a brief introduction to algorithm families used in this domain, research is presented where these methods are combined with machine learning to achieve better and faster results, with the best matches being presented in Discussion. The authors hope that this review serves as a guideline to expand the usability of machine learning for power transformers and that propels further research about artificial intelligence scenarios related to smart manufacturing.

References

Aaron, L., 2008. Writing a literature review article. Radiologic technology 80, 185–186.

Bartoletti, C., Desiderio, M., Di Carlo, D., Fazio, G., Muzi, F., Sacerdoti, G., Salvatori, F., 2004. Vibro-acoustic techniques to diagnose power transformers. IEEE Transactions on power Delivery 19, 221–229.

Basuki, A., et al., 2018. Online dissolved gas analysis of power transformers based on decision tree model, in: 2018 Conference on Power Engineering and Renewable Energy (ICPERE), IEEE. pp. 1–6.

Benmahamed, Y., Teguar, M., Boubakeur, A., 2018. Diagnosis of power transformer oil using pso-svm and knn classifiers, in: 2018 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM), IEEE. pp. 1–4.

Cheemala, V., Asokan, A.N., Preetha, P., 2019. Transformer incipient fault diagnosis using machine learning classifiers, in: 2019 IEEE 4th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), IEEE. pp. 1–6.

Contreras, J.V., Sanz-Bobi, M., Banaszak, S., Koch, M., 2011. Application of machine learning techniques for automatic assessment of fra measurements, in: Proc. 17th Int. Symp. High Voltage Eng., pp. 1109–1119.

Dong, H., Yang, X., Li, A., Xie, Z., Zuo, Y., 2019. Bio-inspired phm model for diagnostics of faults in power transformers using dissolved gas-in-oil data. Sensors 19, 845.

Dornenburg, E., Strittmatter, W., 1974. Monitoring oil-cooled transformers by gas-analysis. Brown Boveri Review 61, 238–247.

Duval, M., 2003. New techniques for dissolved gas-in-oil analysis. IEEE Electrical Insulation Magazine 19, 6–15.

Faiz, J., Heydarabadi, R., 2014. Diagnosing power transformers faults. Russian Electrical Engineering 85, 785–793.

Faiz, J., Soleimani, M., 2017. Dissolved gas analysis evaluation in electric power transformers using conventional methods a review. IEEE Transactions on Dielectrics and Electrical Insulation 24, 1239–1248.

Guo, C., Dong, M., Wu, Z., 2019. Fault diagnosis of power transformers based on comprehensive machine learning of dissolved gas analysis, in: 2019 IEEE 20th International Conference on Dielectric Liquids (ICDL), IEEE. pp. 1–4.

Huang, W., Huang, Z., Long, S., Long, Y., 2019. Multiscale information fusion-based extreme learning machine for power transformer fault diagnosis, in: 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2), IEEE. pp. 2182–2186.

Huang, X., Wang, X., Tian, Y., 2018. Research on transformer fault diagnosis method based on gwo optimized hybrid kernel extreme learning machine, in: 2018 Condition Monitoring and Diagnosis (CMD), IEEE. pp. 1–5.

IEC, 2015. Mineral oil-impregnated electrical equipment in service—guide to the interpretation of dissolved and free gases analysis. International Standard 60599.

IEEE, 2019. Ieee guide for the interpretation of gases generated in mineral oil-immersed transformers. IEEE Std C57.104-2019 (Revision of IEEE Std C57.104-2008) , 1–98doi:10.1109/IEEESTD.2019.8890040.

Islam, M.M., Lee, G., Hettiwatte, S.N., 2017. A nearest neighbour clustering approach for incipient fault diagnosis of power transformers. Electrical Engineering 99, 1109–1119.

Jancarczyk, D., Bernaś, M., Boczar, T., 2019. Classification of low frequency signals emitted by power transformers using sensors and machine learning methods. Sensors 19, 4909.

Jia, Y., Zhu, Y., 2018. Partial discharge pattern recognition using variable predictive model-based class discrimination with kernel partial least squares regression. IET Science, Measurement & Technology 12, 360–367.

Kartojo, I.H., Wang, Y.B., Zhang, G.J., et al., 2019. Partial discharge defect recognition in power transformer using random forest, in: 2019 IEEE 20th International Conference on Dielectric Liquids (ICDL), IEEE. pp. 1–4.

Katooli, M.S., Koochaki, A., 2020. Detection and classification of incipient faults in three-phase power transformer using dga information and rule-based machine learning method. Journal of Control, Automation and Electrical Systems 31, 1251–1266.

Kaur, A., Brar, Y.S., Leena, G., 2019. Fault detection in power transformers using random neural networks. International Journal of Electrical and Computer Engineering 9, 78.

Khalyasmaa, 2019. Machine learning algorithms for power transformers technical state assessment, in: 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), pp. 0601–0606. doi:10.1109/SIBIRCON48586.2019.8958395.

Kunicki, M., Wotzka, D., 2019. A classification method for select defects in power transformers based on the acoustic signals. Sensors 19, 5212.

Li, A., Yang, X., Dong, H., Xie, Z., Yang, C., 2018. Machine learning-based sensor data modeling methods for power transformer phm. Sensors 18, 4430.

Li, S., Wu, G., Gao, B., Hao, C., Xin, D., Yin, X., 2016. Interpretation of dga for transformer fault diagnosis with complementary sae-elm and arctangent transform. IEEE Transactions on Dielectrics and Electrical Insulation 23, 586–595.

Li, Y., Ma, L., 2019. Fault diagnosis of power transformer based on improved particle swarm optimization os-elm. Archives of Electrical Engineering 68.

Liu, J., Zhao, Z., Tang, C., Yao, C., Li, C., Islam, S., 2019. Classifying transformer winding deformation fault types and degrees using fra based on support vector machine. IEEE Access 7, 112494–112504.

Mao, X., Wang, Z., Crossley, P., Jarman, P., Fieldsend-Roxborough, A., Wilson, G., 2020. Transformer winding type recognition based on fra data and a support vector machine model. High Voltage 5, 704–715.

Mao, X., Wang, Z., Jarman, P., Fieldsend-Roxborough, A., 2019. Winding type recognition through supervised machine learning using frequency response analysis (FRA) data, in: 2019 2nd International Conference on Electrical Materials and Power Equipment (ICEMPE), IEEE. pp. 588–591.

Nurmanova, V., Akhmetov, Y., Bagheri, M., Zollanvari, A., Gharehpetian, G.B., Phung, T., 2020. A New Transformer FRA Test Setup for Advanced Interpretation and Winding Short-circuit Prediction, in: 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), IEEE. pp. 1–6.

Pasi, B., Mahajan, S., Rane, S., 2020. Redesigning of smart manufacturing system based on iot: perspective of disruptive innovations of industry 4.0 paradigm. International Journal of Mechanical and Production Engineering Research and Development 10, 727–746. doi:10.24247/ijmperdjun202067.

Raichura, M.B., Chothani, N.G., Patel, D.D., 2020. Identification of internal fault against external abnormalities in power transformer using hierarchical ensemble extreme learning machine technique. IET Science, Measurement & Technology , 111–121.

Rexhepi, V., 2017. An analysis of power transformer outages and reliability monitoring. Energy Procedia 141, 418–422. URL: https://www.sciencedirect.com/science/article/pii/S1876610217354590, doi:https://doi.org/10.1016/j.egypro.2017.11.053. power and Energy Systems Engineering.

Rogers, R., 1978. Ieee and iec codes to interpret incipient faults in transformers, using gas in oil analysis. IEEE transactions on electrical insulation , 349–354.

Shang, H., Xu, J., Zheng, Z., Qi, B., Zhang, L., 2019. A novel fault diagnosis method for power transformer based on dissolved gas analysis using hypersphere multiclass support vector machine and improved d–s evidence theory. Energies 12, 4017.

Soto, A.R., Lima, S.L., Saavedra, O.R., 2019. Incipient fault diagnosis in power transformers by dga using a machine learning ann-mean shift approach, in: 2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), IEEE. pp. 1–6.

Su, L., Cui, Y., Yu, Z., Chen, L., Xu, P., Hou, H., Sheng, G., 2019. A power transformer fault diagnosis method based on hierarchical classification and ensemble learning, in: Journal of Physics: Conference Series, IOP Publishing. p. 012045.

Sun, H.C., Huang, Y.C., Huang, C.M., 2012. A review of dissolved gas analysis in power transformers. Energy Procedia 14, 1220–1225.

Taha, I.B., Dessouky, S.S., Ghoneim, S.S., 2021. Transformer fault types and severity class prediction based on neural pattern-recognition techniques. Electric Power Systems Research 191, 106899.

Tamilselvan, P., Wang, P., 2013. Failure diagnosis using deep belief learning based health state classification. Reliability Engineering & System Safety 115, 124–135.

Viereck, K., Saveliev, A., 2019. Driving reliability with machine learning and improving operation by digitalization of power transformers, in: 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), IEEE. pp. 217–222.

Woon, W.L., Aung, Z., El-Hag, A., 2018. Intelligent monitoring of transformer insulation using convolutional neural networks, in: International Workshop on Data Analytics for Renewable Energy Integration, Springer. pp. 127–136.

Woon, W.L., El-Hag, A., Harbaji, M., 2016. Machine learning techniques for robust classification of partial discharges in oil–paper insulation systems. IET Science, Measurement & Technology 10, 221–227.

Wotzka, D., Cichoń, A., 2020. Study on the influence of measuring ae sensor type on the effectiveness of oltc defect classification. Sensors 20, 3095.

Xie, B., Zhao, D., Hong, T., 2020. Transformer monitoring and protection in dynamic power systems – a review. Frontiers in Energy Research 8. URL: https://www.frontiersin.org/article/10.3389/fenrg.2020.00150, doi:10.3389/fenrg.2020.00150.

Yadaiah, N., Ravi, N., 2011. Internal fault detection techniques for power transformers. Applied Soft Computing 11, 5259–5269.

Yang, X., Chen, W., Li, A., Yang, C., 2020. A hybrid machine-learning method for oil-immersed power transformer fault diagnosis. IEEJ Transactions on Electrical and Electronic Engineering 15, 501–507.

Zhao, Z., Tang, C., Zhou, Q., Xu, L., Gui, Y., Yao, C., 2017. Identification of power transformer winding mechanical fault types based on online IFRA by support vector machine. Energies 10, 2022.

Zheng, H., Shioya, R., 2020. A comparison between artificial intelligence method and standard diagnosis methods for power transformer dissolved gas analysis using two public databases. IEEJ Transactions on Electrical and Electronic Engineering 15, 1305–1311.