

Review

Not peer-reviewed version

Power Transformer Prognostics and Health Management Using Machine Learning: A Review and Future Directions

[Ryad Zemouri](#) *

Posted Date: 15 January 2025

doi: 10.20944/preprints202501.1113.v1

Keywords: Power Transformer; Prognostics and Health Management; Machine Learning; Artificial Intelligence



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Power Transformer Prognostics and Health Management Using Machine Learning: A Review and Future Directions

Ryad Zemouri 

Hydro-Québec Research Center (CRHQ) 1800 Bd Lionel-Boulet Varennes J3X 1S1 Québec Canada;
zemouri.ryad@hydroquebec.com

Abstract: Power transformers (PT) play a vital role in the electrical power system. Assessing their health to predict their remaining useful life is essential to optimise maintenance. Scheduling the right maintenance for the right equipment at the right time is the ultimate goal of any power system utility. Optimal maintenance has a number of benefits. Human and social, by limiting sudden service interruptions, and economic, due to the direct and indirect costs of unscheduled downtime. PT now produces large amounts of easily accessible data due to the increasing use of IoT, sensors and connectivity between physical assets. As a result, Power Transformer Prognostics and Health Management (PT-PHM) methods are increasingly moving towards Artificial Intelligence (AI) techniques, with several hundred of scientific papers published on the topic of PT-PHM using AI techniques. On the other hand, the world of AI is undergoing a new evolution towards a third generation of AI models: Large scale foundation models. What is the current state of research in PT-PHM? What are the trends and challenges in AI and where do we need to go for power transformer prognostics and health management? This paper provides a comprehensive review of the state of the art in PT-PHM by analysing more than 200 papers, mostly published in scientific journals. Some elements to guide PT-PHM research are given at the end of the document.

Keywords: power transformer; prognostics and health management; machine learning; artificial intelligence

1. Introduction

Power transformers (PT) are essential links in the electrical power system and have a direct impact on the reliability of the entire network [1]. Being subject to continuous thermal, electrical and mechanical stresses, power transformers, especially aged ones, are prone to various types of failures that degrade their performance and reduce their lifetime [2]. It is therefore necessary to assess their health status, provide a strategic management plan for each transformer and determine the most economical management approach. Without effective management of these assets, it is difficult to make decisions about maintenance and replacement priorities [3]. Poor decisions can result in high maintenance costs and long periods of transformer downtime. The traditional concept of operation and maintenance strategies is often time-based, i.e. systematic maintenance strategies. The transformer is periodically inspected regardless of its condition. This type of maintenance is often very costly to the company and does not promote good asset management.

In general, maintenance strategies can be divided into three types: reactive maintenance, which is carried out after a failure has occurred; systematic maintenance, which is carried out periodically; and condition-based maintenance, where maintenance is conditioned by the state of health of the asset. The development of on-line monitoring sensors offers the possibility of implementing this condition-based maintenance. Real-time monitoring then increases the probability of detecting incipient failures while reducing the probability of failures occurring. It then helps to improve operational safety, control

unscheduled maintenance and prioritise the maintenance and replacement schedule based on the condition of each transformer.

The Prognostics and Health Management (PHM) process consists of three modules: Condition Monitoring, Health Assessment and Prognostics. The aim of the condition monitoring module is to compare the online data or extracted descriptors with certain expected or known values that define a threshold for generating alerts. This module is also known as the fault detection level, which aims to detect abnormalities or anomalies in system behavior. Then the Health Assessment module, usually known as the Diagnostics module, must identify the degradation states of the system and assess the causes of the degradation. Finally, the prognostics module aims to predict future trends in the health status of systems in order to estimate the time remaining before systems will be unable to perform their intended function.

In addition, with the increasing use of Internet of Things (IoT), sensors and connectivity between physical assets, PT are now producing large amounts of easily accessible data. The ability to assimilate this data to quickly extract useful information and descriptors for integration into the organizational knowledge chain is required. AI has become one of the most strategic technologies of the 21st century thanks to the growth in computing power, the availability of data and advances in algorithms. Thus, PHM methods for PT are increasingly moving towards artificial intelligence techniques, as evidenced by the number of publications in the literature [4–8]. The literature review conducted here consolidates the scientific production in the field of Power Transformer Prognostics and Health Management (PT-PHM). A comprehensive and detailed literature analysis is given. Subsequently, this literature review is divided into two parts: classical ML techniques and DL techniques. For each part, a formal description of the main ML models used and their application in the field of PT-PHM is detailed. For each part, a table summarizing the literature review is also provided. One of the main observations that can be drawn from this literature review is the conservative nature of the published work. It can be seen that some popular DL models, such as generative models (VAE, GAN), have been successful in other PHM applications, but are hardly used for PT-PHM. Other recently successful popular approaches such as Transformer-based Deep Neural Network [9] or Self-Supervised Learning [10] are completely missing in the PT-PHM domain. An important question for the community of researchers, engineers and students working in the field of PT-PHM is therefore *What are the trends in AI and where do we need to go for power transformer prognostics and health management?*

The section 2 gives a brief description of the evolution of ML techniques in order to give the reader some important elements to understand how the ongoing developments on the foundation models will affect the integration of ML techniques in PT-PHM. Then, the main part of this review paper is dedicated to sections 3 and 4, which give a complete literature review on the ML techniques for PT-PHM. Finally, sections 5 and 6 give some discussion and challenges about the way the research field in PT-PHM needs to go.

2. From Shallow Machine Learning to Foundation Models

2.1. Shallow Machine Learning

The rise of the first ML algorithms began in the 1990s, when the concept of learning from data was introduced. This was the first foundation of AI: a learning algorithm induces the way to solve a task from data, rather than defining how to solve it from prior knowledge. Various prediction, regression and classification applications were then performed on the data. However, for complex data, such as text or images, a step of feature engineering by domain experts is required.

In the first integration of shallow ML architectures in PHM applications, the raw measurements provided by the sensors cannot be easily linked to the health state of the systems. In fact, the data is often affected by a significant amount of noise or imperfect signal transmission. In addition, these data are often represented by complex time series, typically characterized by a high redundant information content, which tends to hide the relatively limited discriminative features of interest. For these reasons, once the data are acquired, a set of candidate features must be extracted and then only the most

informative among them must be properly selected. Once these steps are completed, the final set of extracted features can be used to train an ML algorithm to perform the desired diagnostic or prognostic task (Figure 1). The powerfulness of shallow ML techniques can be quite limited, and their input often consists of high-level features manually extracted from raw data by human experts.

2.2. Deep Machine Learning

The emergence of DML in the 2010s revitalized the field of AI. Several ingredients made this resurgence possible: 1) the availability of massive data, 2) the advent of GPU resources, and 3) the tenacity of many AI researchers. DNNs are then trained on the raw input data, and high-level features emerge through training. This has led to great performance gains by several DML architectures on standard benchmarks.

Thus, in the PHM process, DML architectures emerge as an extension of classical shallow ML architectures. Once DNNs are trained, their inputs pass through a nested series of successive computations, resulting in the extraction of a set of complex features across different engineering domains that are highly informative for the task of interest. This property is one of the hallmarks of DML and can be seen as one of its key success factors for automated *end-to-end* feature extraction from different data structures. The main power of DML is then the automated feature learning from low-level elementary features to high-level abstract features [11] (Figure 1). Nevertheless, DML models perform well when trained on specific data to solve a specific task in a specific context. This constraint limits their use for complex power systems, which require a high degree of model explainability, and robust generalization.

2.3. The Emergence of a New Concept, the Foundation Models

At the end of the 2020s, important AI concepts have been developed that form the basic ingredients of the foundation models. In fact, recent advances in AI research have led to the emergence of a new paradigm: foundation models. These recent advances are: Modular Learning, Transformer architecture, SSL, MMF, MTL, and graph-oriented approaches.

A foundation model is described as a general-purpose AI model that has been trained to solve a wide range of general-purpose tasks such as text synthesis, image manipulation, and audio generation. It can be a stand-alone system or can be used as a basis for many other applications or models. The idea of foundation models was first introduced by Bommasani et al, [12]. Some of the best known foundation models are BERT [13], GPT [14] and CLIP [15]. It is important to note that the Foundation Models define a new concept in AI rather than a new structure or algorithm of a ML model. The power of the foundation models lies in their scale, which requires improvements in computer hardware and the availability of much more training data. The modular DL is at the heart of the scalability of foundation models (Figure 1). Thus, foundation models overcome the limitations of DL models by providing better generalization performance on large multimodal datasets. However, building a foundation model is often very resource intensive. The most expensive models cost hundreds of millions of dollars for the underlying data and computation required. In contrast, it is much less expensive to adapt or directly use an existing foundation model for a specific use case.

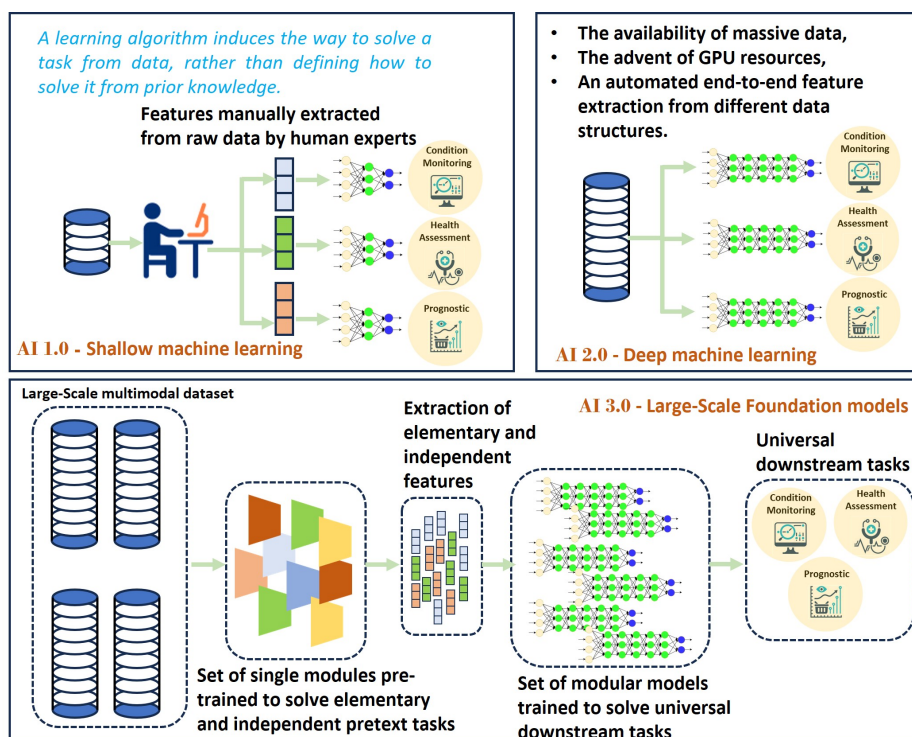


Figure 1. From shallow machine learning to foundation models

3. Classic ML Techniques

This section mainly discusses the use of classical ML methods for the PT-PHM. A total of 149 papers have been reviewed and analysed for this section. The figure 2 shows the density distribution of these papers in the year of publication. As can be seen from the figure 3, a first main group, representing more than 50% of the papers analyzed, concerns the Fuzzy Inference Systems, the SVM and the ANN. A second group, representing 19% of the papers, deals with other ML techniques such as the ANFIS, the Gaussian Process, the Feed-forward Wavelet Network and two clustering techniques, the Fuzzy c-means and the K-NN. In the third group, several ML techniques were used together for some comparative studies in 13% of the papers. For example, in [16–19] several ML techniques such as ANN, SVM and ANFIS were compared for HI assessment. Finally, a last group, representing almost 10% of the papers analyzed, concerns some marginal applications of the following ML techniques: Bayesian inference (2.70%), ELM (2.70%), ensemble learning (2%), decision tree (2%), random forest (1.35%) and hidden Markov models (0.68%).

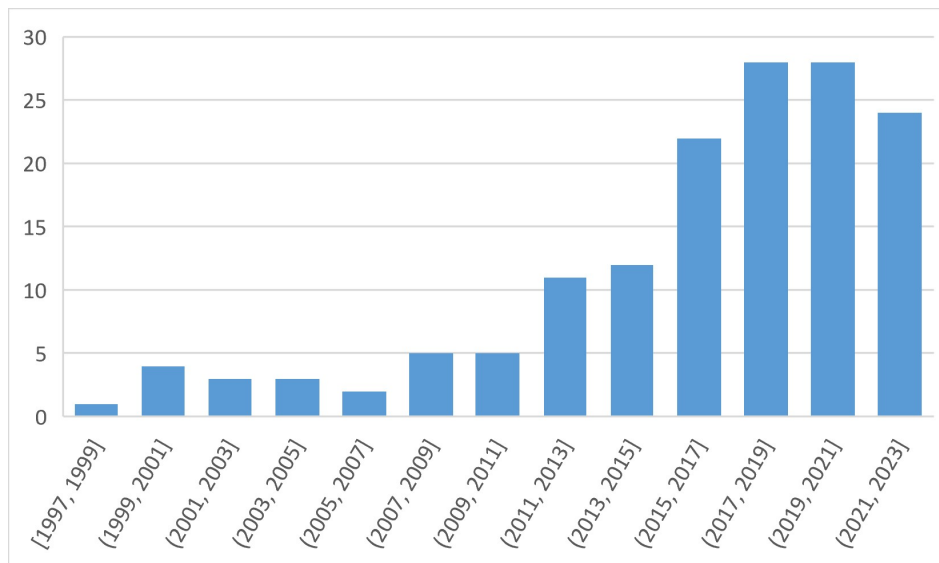


Figure 2. The density distribution of the whole conventional ML techniques papers in the year of publication.

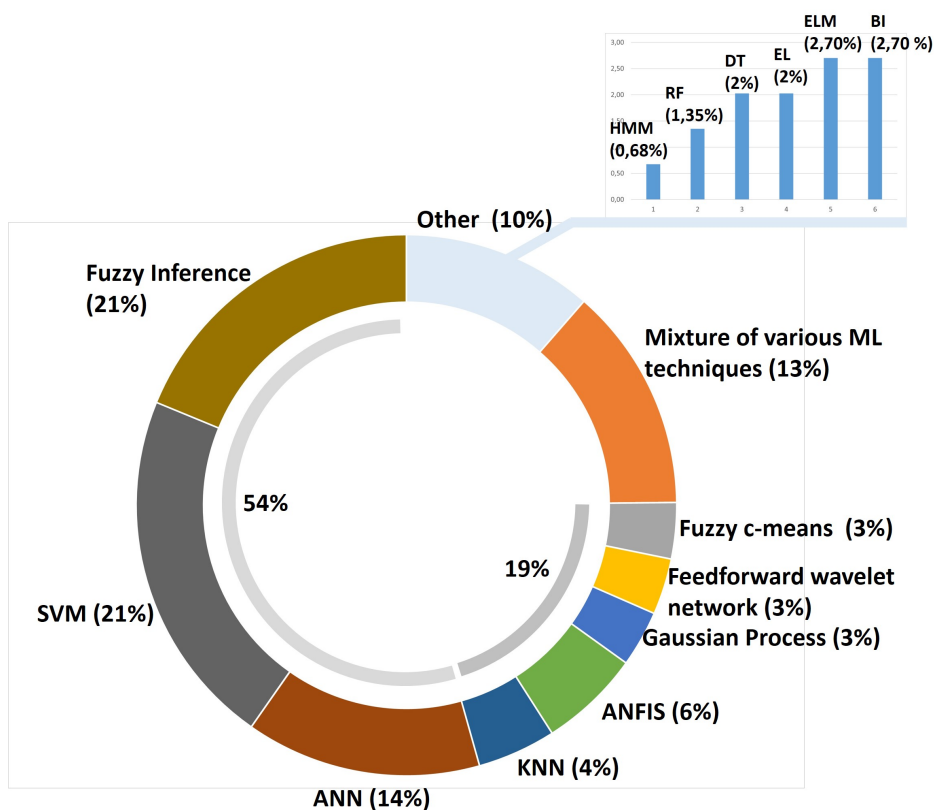


Figure 3. The density distribution of the various conventional ML techniques used for the health and asset management of the power transformers.

3.1. Artificial Neural Networks

Recent papers have used ANNs for monitoring and asset management of PT [28–30,34,42]. In these papers, shallow structures of the feed-forward MLP are trained by the back-propagation algorithm. In [42], a BPNN enhanced with the AdaBoost algorithm is used to estimate moisture content. The results show superior performance compared to other techniques such as SVM, RF and k-NN. Mousavi et al., [34] propose a method based on an Artificial Neural Network (ANN) integrated genetic algorithm to modify the effect of temperature variation in the results of polarisation and depolarisation current (PDC) tests. The purpose of ANN is to transfer parameters related to higher temperatures to target

parameters. In [30], an artificial neural network based approach is proposed to assign the weight for several independent DGA methods in the fusion procedure according to the fault type detection accuracy for the range of input gas concentrations.

3.2. Fuzzy Inference System

In [48], the risk index for power transformers is assessed from the aggregation of three fuzzy inference systems. The expert criteria are exploited to create the rules and the definitions of the input membership functions. The HI is predicted by the FIS based on the results of the physico-chemical tests normally performed on the insulating liquid of the units. This paper shows that the proposed methodology, based on the FIS, for assessing the condition of power allows the integration of values commonly available to asset managers in electrical systems. Through this integration, it is possible to rank the units in order to define maintenance strategies for the analysed units. With the same objective of decision making for PT asset management, a combined approach of FIS and fuzzy clustering means is used by [51] as an expert system for diagnosis and prognosis of incipient faults of power transformers along with critical parameters such as dissolved gas analysis, moisture content, furans, interfacial tension, degree of polymerisation.

In the majority of the papers analyzed, the FIS have been applied mainly to FDD based on dissolved gas analysis, as detailed in the table 1. However, this type of method can also be found in other fields such as partial discharge analysis [44,107], Frequency Response Analysis [106,108] thermal analysis [106,111] and heat dissipation [106], Frequency spectroscopy [112], insulation resistance [106] and insulation paper analysis [51,109].

3.3. Support Vector Machine (SVM)

The SVM method is one of the most popular techniques used in power transformers for FDD, where the function $f(x)$ is used to classify conditions such as discharge faults ($f(x) = +1$) from thermal faults ($f(x) = -1$). SVMs have the advantage of using a small amount of training data, which reduces the training time [163,164]. The accuracy of the SVM technique depends on its parameters and the kernel function used [165].

As for the other ML techniques, SVM has mainly been used in power transformer diagnostics based on dissolved gas analysis, as shown in the table 1. The simplest way to use SVM for FDD based on a set of dissolved gas concentrations as input is binary classification. The most common are hydrogen (H₂), methane (CH₄), acetylene (C₂H₂), ethylene (C₂H₄), ethane (C₂H₆), carbon monoxide (CO) and carbon dioxide (CO₂). In [62] a multi-stage SVM classification was proposed for FDD. The first stage consists of separating a faulty condition from a non-faulty condition. The following stages are successively used to separate the faulty conditions, e.g. the thermal fault from the discharge fault. A method based on Kernel Principal Component Analysis (KPCA) and a hybrid improved Seagull Optimization Algorithm is proposed by [60] to optimise the parameters of the SVM. Furthermore, the KPCA technique was used by [58] for feature extraction and dimension reduction, and reduce the dimension of the feature vector. In order to obtain the optimal classification model in high-dimensional space, a Genetic Algorithm (GA) introduced in Whale Optimization Algorithm (WOA) was used to optimise two important parameters of SVM: Penalty Factor and Kernel Function Parameter.

Furthermore, the SVM techniques have been used in the detection of partial discharges [2,72,76], in frequency response analysis [71] in the analysis of vibro-acoustic signals [74], thermal analysis [21,161] and infrared analysis [70], in polarisation-depolarisation current analysis [65,82] in spectroscopy [75] in the prediction of the degree of polymerisation [73] and for the differential protection [77].

3.4. Summary of Work Using Classic ML methods

Table 1 summarizes the analysis of all the papers considered in this section in terms of the classic ML techniques used. Three tasks are then highlighted, the fault detection and diagnosis (FDD), the health index assessment and the prediction of some condition monitoring data such as the prediction of top oil temperature [20] or the thermal modeling for condition monitoring [21]. The main part of

the analyzed paper deals with dissolved gas analysis, alone or in combination with other monitoring parameters (marked with a + sign). In the next three subsections we will have a closer look at the main ML techniques used by the first group shown in the figure 3, i.e. ANNs, FIS and SVMs.

Table 1. Summary of work using Classic ML methods.

Classic ML	Task	Data	Ref
MLP	FDD	DGA	[22–32]
	FDD	DGA ⁺	[33,34]
	FDD		[35,36]
	HI	DGA	[37,38]
	HI	DGA ⁺	[39–41]
	Pred.	Other	[42]
ANFIS	FDD	DGA	[43–46]
	FDD	DGA ⁺	[47,48]
	FDD	Other	[49–51]
Clustering x SVM	FDD	DGA	[52–69]
	FDD	Other	[2,70–77]
	HI	DGA ⁺	[78–80]
	Pred.		[81]
x Fc-m	FDD	DGA	[82–84]
	HI	DGA	[85]
	Pred.	Other	[86]
x KNN	FDD	DGA	[87–91]
	FDD	Other	[92,93]
FIS	FDD	DGA	[51,94–105]
	FDD	Other	[106–112]
	HI	DGA ⁺	[48,113–119]
	HI	Other	[120,121]
BI	FDD	DGA	[122,123]
	HI	DGA ⁺	[124,125]
DT	FDD	DGA	[126,127]
	FDD	Other	[128]
WN	FDD	DGA	[129–131]
	FDD	Other	[132]
	HI	DGA ⁺	[133]
GP	FDD	DGA	[134–136]
	HI	DGA ⁺	[137]
	Pred.	Other	[138]
EL	FDD	DGA	[139–141]
RF	FDD	DGA	[142,143]
HMM	FDD	DGA	[144]
ELM	FDD	DGA	[145–148]
Mix. ML	FDD	DGA	[149–157]
	HI	DGA ⁺	[16–19,158,159]
	Pred.	Other	[20,21,160–162]

4. Deep Learning Architectures

Nowadays, due to the development of several heuristics for training large architectures and the use of GPU hardware [166], [167], [168], the size of ANNs can reach several hidden layers with more than 650 thousand neurons and 630 million trained parameters (e.g. alexNet [169]). In recent years, deep learning methods based on neural networks have attracted the attention of many researchers for dissolved gas analysis [170], infrared image analysis [171] for differential protection [172], or for the prediction of dibenzyl disulphide concentration [173]. In this section we discuss the use of Deep ML methods for PT-PHM. A total of 42 articles have been reviewed and analysed for the purpose of

this section. The figure 4 shows the density distribution of these papers in the year of publication. As shown in the figure 5, the most popular DML architecture is the convolutional neural network, which is used in more than 50% of the published papers. Recurrent Neural Networks and AE architectures are two other DL models that are also proving successful. Finally, two new concepts are emerging in the field of PT: attention mechanics and GNNs.

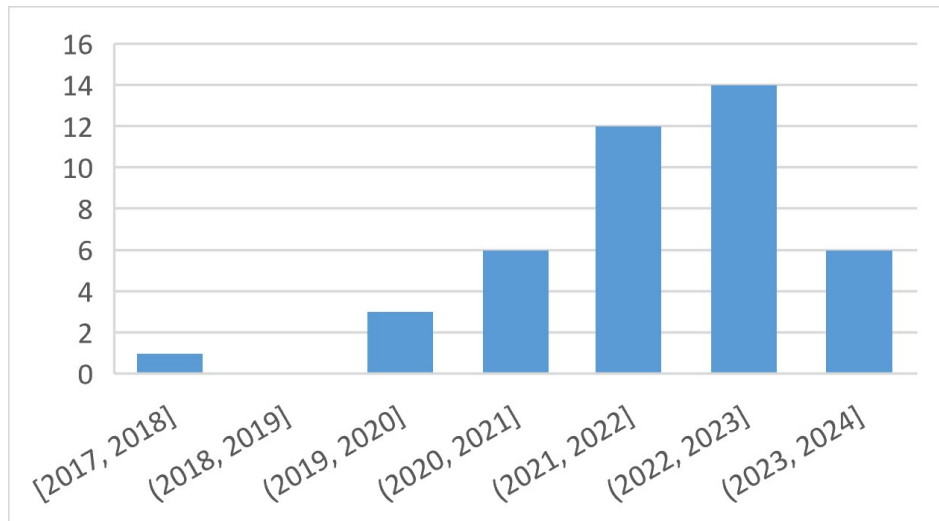


Figure 4. The density distribution of the whole conventional ML techniques papers in the year of publication.

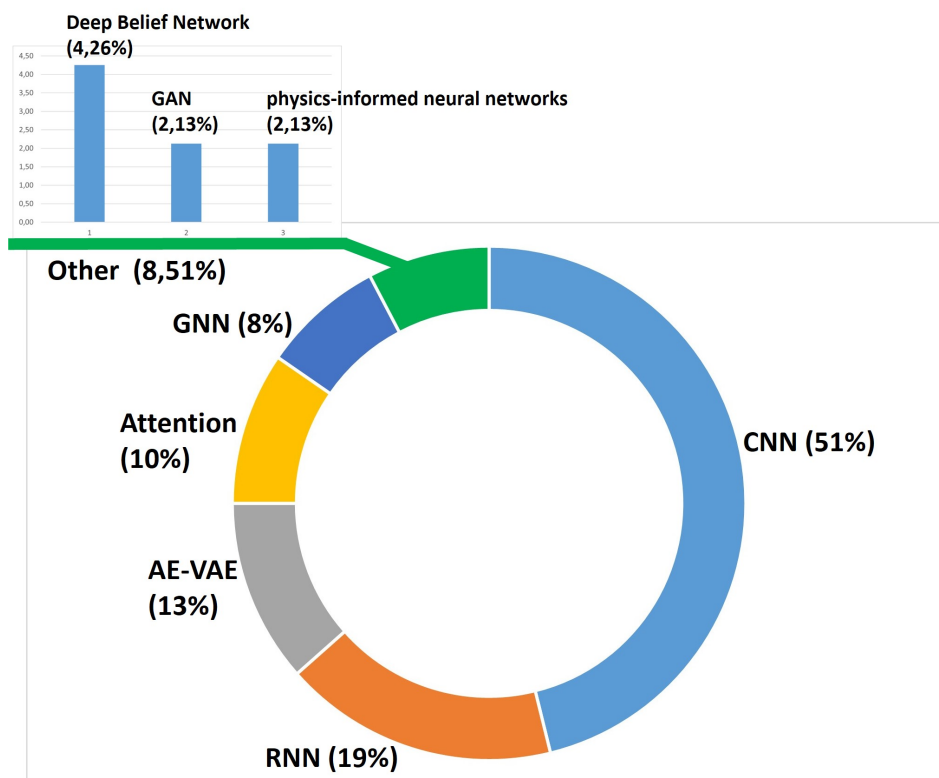


Figure 5. The density distribution of the various conventional ML techniques used for the health and asset management of the power transformers.

4.1. Convolutional Neural Networks

CNNs have been successfully applied to the detection and diagnosis of power transformers, as shown in the table 4, mainly for DGA, but also for partial discharge source detection [182–184] and vibration and acoustic signal analysis [181]. In [174], the difficulty of applying the CNN model to

a small vector size is addressed. In fact, existing dissolved gas analysis methods use five gases as features (H_2 , CH_4 , C_2H_6 , C_2H_4 , C_2H_2). Compared to image type data, such as thermal image data, only five features are too few to serve as input to the CNN model during training. A reduced number of input features combined with the deep layers of the CNN slows down the learning process and makes it too oscillatory, leading to overfitting. Features are then reconstructed to overcome this problem, as shown in Figure 6. Permutations and combinations of the five gases allow the features of the model input vector to be increased.

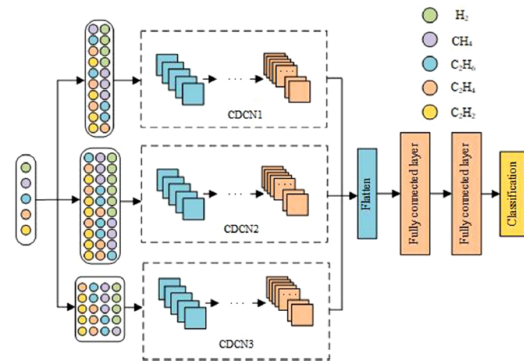


Figure 6. Permutations and combinations of the five dissolved gases are used to augment the characteristics of the CNN model input vector [174].

With the same objective of increasing the size of the input vector of dissolved gas concentration, the method proposed by [178] uses an input vector of 35 features containing three categories: 1) seven raw dissolved gas concentration data including H_2 , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , CO and CO_2 ; 2) the fraction of each gas relative to the total concentration of the seven gases; and 3) 21 proportions calculated from the ratio of two gas concentrations. All features are normalised to the interval $[0, 1]$ to reduce the scale differences of the input features.

A total of 28 dissolved gas fractions are considered in [176] (listed by table 2). To overcome the problem of over-fitting from a reduced set of data available for training, the *gcForest* algorithm is optimised by a combination of CNN and *CasXGBoost* (cascade extreme gradient boosting). A total of five output classes are used: LE-D (Low Energy Discharge), HE-D (High Energy Discharge), LM-T (Low Temperature Overheating), HT (High Temperature Overheating) and NC (Normal Condition).

Table 2. The 28 dissolved gas concentrations used by [176].

No	Ratios	No	Ratios	No	Ratios
1	H_2/CH_4	11	CH_4/CO	21	C_2H_4/CO_2
2	H_2/C_2H_2	12	CH_4/CO_2	22	C_2H_4/TH
3	H_2/C_2H_4	13	CH_4/TH	23	C_2H_6/CO
4	H_2/C_2H_6	14	C_2H_2/C_2H_4	24	C_2H_6/CO_2
5	H_2/CO	15	C_2H_2/C_2H_6	25	C_2H_6/TH
6	H_2/CO_2	16	C_2H_2/CO	26	CO/CO_2
7	H_2/TH	17	C_2H_2/CO_2	27	CO/TH
8	CH_4/C_2H_2	18	C_2H_2/TH	28	CO_2/TH
9	CH_4/C_2H_4	19	C_2H_4/C_2H_6	-	-
10	CH_4/C_2H_6	20	C_2H_4/CO	-	-

4.2. Recurrent Neural Network

In order to take into account the dynamics of the training data, a particular architecture of recurrent neural networks (RNNs) is specially adapted for the analysis of time signals. RNNs contain feedback loops to memorize information from previous units and are best suited to time series analysis. Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) cells are popular variants of RNNs that attempt to alleviate the vanishing gradient problem. LSTM has recently been used for dissolved gas analysis and power transformer prediction [192,197–199]. Thus, models have been developed and tested by [192] and [198] for predicting the concentration of the following gases H_2 , CH_4 , C_2H_6 , C_2H_4 , C_2H_2 from a concentration history. In [198] a simple structure of a LSTM neural network was tested

and compared with other models such as a *Support Vector Regression*, *Regression Tree* and *Gaussian Process Regression*. According to the results published by the authors, the *LSTM* model outperforms other models in predicting gas concentrations over a seven-week horizon. In addition, a more complex hybrid structure combining a *Highly Attentional Tracking Network (HATT)* model and a *Recurrent Long Short-Term Memory Network (RLSTM)* model has been proposed by [192] for predicting dissolved gas concentrations.

In order to improve the reliability of the diagnosis obtained from a fault classification model listed in Table 3, several configurations of *LSTM* and *bi-LSTM* models, illustrated by Figure 7, were tested by [197]. According to the results published by the authors, the performance of the *LSTM* model in terms of accuracy, recall and precision is 99.01%, 98.79% and 99.12% respectively.

Table 3. The types of failures used by [197].

Fault type	Fault code	Number of data
Thermal fault (<300 °C)	T1	320
Thermal fault (300 °C–700 °C)	T2	322
Thermal fault (>700 °C)	T3	325
Lower energy discharge	D1	323
Higher energy discharge	D2	319
Partial discharge	PD	318
Normal state condition	NS	315
Total		2242

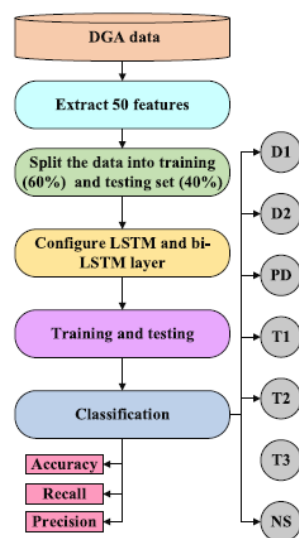


Figure 7. LSTM structure used by [197] to optimize diagnostic performance

4.3. Auto-Encoders

The AEs have recently been applied to dissolved gas analysis [180,200–203]. In [202], the AE is used to extract some relevant features from a set of several concentrations of dissolved gases such as H_2 , C_2H_2 , C_2H_4 , C_2H_6 , CH_4 and CO . Thus, the features extracted from the encoder are used to detect and identify two types of faults: thermal and electrical. As shown in Figure 8, health indicators are thus defined for the detection and identification of faults from the latent space obtained by the encoder. Four degradation sequences are illustrated by [202]: two thermal degradation sequences and two electrical sequences.

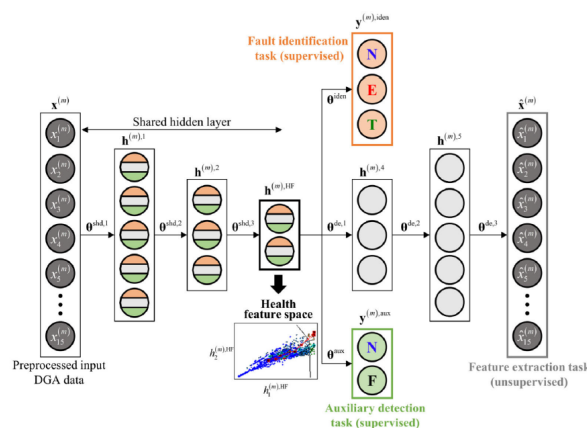


Figure 8. The Auto-encoder model used by [202] for the detection and identification of two types of faults: thermal and electrical.

Due to the complex operating condition of the transformer, its faults are with the characteristic of multi-class faults, class imbalance, and limited diagnostic data availability, the VAE has been used as a generative model for data augmentation [180,200,201]. The results presented in these papers show that the performance of the diagnostic model was improved after data augmentation by VAE.

4.4. Attention Mechanism

The attention mechanism has recently been used for PT asset management by the following papers [177,179,188,192,196]. The attention mechanism has been used to improve the prediction of the model by further mining the temporal relationship of the different time points in [192] and to improve the power transformer fault diagnosis based on DGA [177,179]. In transformer fault diagnosis, dissolved gas in oil data have a wide variety of types, and the main gas data of some faults are similar. Therefore, it is more difficult to obtain important characteristic data. To solve this problem, a channel space-time attention network is used by [177] to fully extract the significant features from the dissolved gas in oil. Furthermore, current transformer diagnostics methods focus on discrete dissolved gas analysis, neglecting deep feature extraction from multi-channel sequential data. The unused consecutive data contains the significant temporal information reflecting the transformer condition. To solve this problem, multichannel consecutive data cross-extraction is proposed by [179] to extract the significant information on time sequence and channel successively, as shown in Figure 9. The multichannel sequential data, including hydrogen (H₂), methane (CH₄), ethane (C₂H₆), ethylene (C₂H₄) and acetylene (C₂H₂). The output shows the probability distribution of the condition of the power transformer, with the highest being selected as the final diagnostic result. The transformer states include Normal Condition (NC), Low Overheating (LT), Medium Overheating (MT), High Overheating (HT), Partial Discharge (PD), Low Energy Discharge (LD) and High Energy Discharge (HD). The experimental tests carried out by the authors show that the MCDC model, based on the attention mechanism, outperforms other models such as HMM, GRU and DBN.

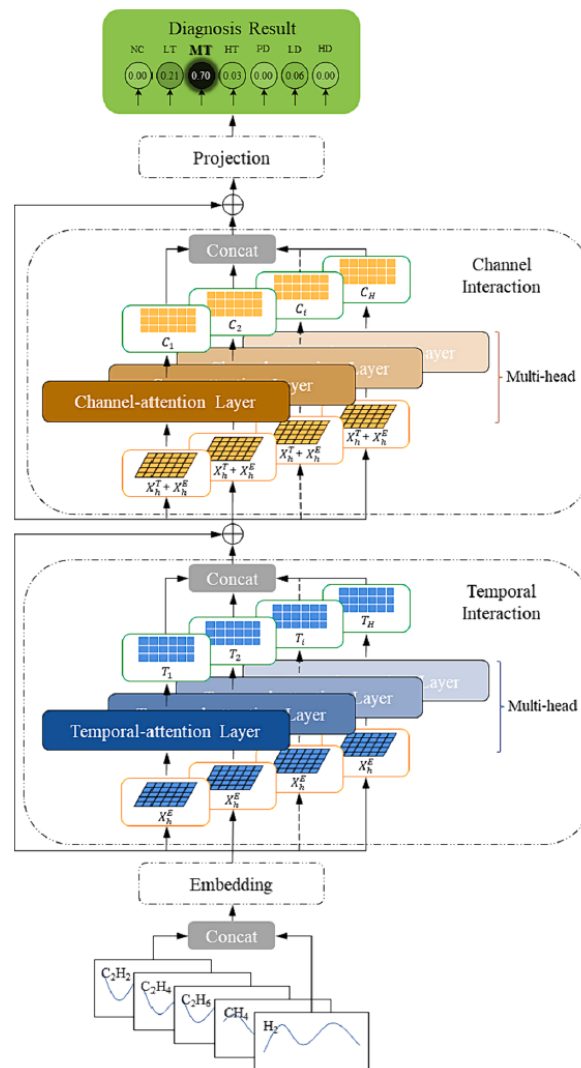


Figure 9. The multi-channel consecutive data cross-extraction proposed by [179] to extract significant information on time sequence and channel successively including hydrogen (H₂), methane (CH₄), ethane (C₂H₆), ethylene (C₂H₄) and acetylene (C₂H₂). The output diagnostic result is the highest value of transformer conditions as a probability distribution including Normal Condition (NC), Low Overheating (LT), Medium Overheating (MT), High Overheating (HT), Partial Discharge (PD), Low Energy Discharge (LD) and High Energy Discharge (HD).

Finally, another interesting application of the attention mechanism is multimodal fusion. This technique has been used by [188] to fuse two modalities: dissolved gas data and infrared images. The experimental diagnostic results presented by the authors show the superiority of multi-modal fusion over other techniques such as DBN, HMM, ANN.

4.5. Graph Neural Networks

In many scientific fields, some important objects and problems can be expressed naturally, or better, with a complex structure. In fact, structural and semantic information from the original data (images, sequential text or time series) can be used to incorporate domain-specific knowledge to capture finer relationships between data. Graph-based approaches, associated with the concept of ANN, represent a new paradigm in the field of ML that highlights semantic causal inference relationships [211,212], such as the interdependence between systems and components in a predictive maintenance [213]. Furthermore, these methods can achieve promising performances in reasoning tasks, while promoting their explainability and interpretability. In recent years, GNN [214–217] has attracted increasing interest from the scientific community [213,218–222], and several variants have been developed, such as convolutional GNN [223], recurrent GNN [224], and auto-encoder GNN [225]. The goal of a GNN

is to learn effective node representations by iteratively combining the structure of the graph with the representation of the node attributes.

The data processing capabilities of GNNs have been explored in recent publications for assessing the condition of power transformers [175,193,205,206]. A model based on GCN has been proposed by [206] to predict the dissolved gas concentration. As shown in Figure 10, the input vector of the prediction model is defined by $\mathbf{X}_t = (X_t^1, X_t^2, \dots, X_t^N)$, where each input variable $X_t^i = (x_{tT}^i, \dots, x_{t-1}^i, x_t^i)$ represents the history of the concentration of dissolved gas i^{th} , and each element x_j^i gives the concentration of dissolved gas i^{th} at time j . The output of the model is the prediction of the dissolved gas concentration at time $t + h$. $\hat{\mathbf{Y}}_{t+h} = (\hat{x}_{t+h}^1, \hat{x}_{t+h}^2, \dots, \hat{x}_{t+h}^N)$.

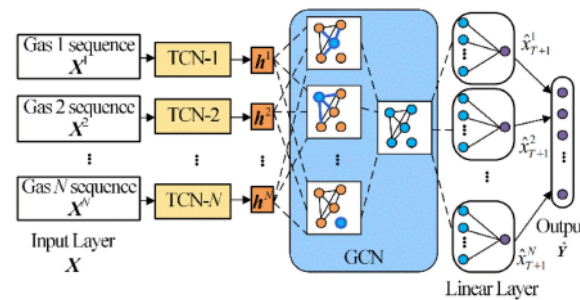


Figure 10. The dissolved gas concentration prediction model proposed by [206]. The model consists of three parts: a temporal convolutional neural network (TCN- N) layer, a graph neural network (GCN) layer, and a linear layer.

The concentration of the following five gases (H_2 , CH_4 , C_2H_6 , C_2H_4 , C_2H_2), is used for a 500 KV power transformer to predict one step ahead. The predictions made by GCN are better than those made by other architectures such as LSTM.

4.6. Summary of Work Using Deep Learning Architectures

Table 4 summarizes the analysis of all the papers considered in this section in terms of the Deep learning architectures used. Three tasks are then highlighted, the fault detection and diagnosis (FDD), the health index assessment and the prediction of some condition monitoring data.

Table 4. Summary of work using Deep learning architectures.

DL techniques	Task	Data	Ref
CNN	FDD	DGA	[174–180]
	FDD	Other	[172,181–188]
	HI	DGA	[189]
	HI	DGA+	[190,191]
	Pred.	DGA	[192,193]
	Pred.	Other	[194–196]
RNN	FDD	DGA	[197]
	Pred.	DGA	[192,193,198,199]
	FDD	Other	[172,188]
	Pred.	Other	[195,196]
AE-VAE	FDD	DGA	[180,200–203]
	Pred.	DGA	[204]
Attention	FDD	DGA	[177,179]
	FDD	Other	[188]
	Pred.	DGA	[192]
	Pred.	Other	[196]
GNN	FDD	DGA	[175]
	HI	DGA ⁺	[205]
	Pred.	DGA	[193,206]
DBN	FDD	DGA	[207,208]
GAN	FDD	DGA	[209]
PINNs	Pred.	Other	[210]

5. What are the trends in AI and where do we need to go for the prognostics and the health management?

5.1. Modular Learning

The development of a ML model that can perform multiple tasks without experiencing negative inter-task interference phenomena while maintaining good generalization performance on non-identically distributed data is less well controlled [226]. Negative interference is characterized by the phenomenon of catastrophic forgetting, well known in the field of ML. The ability to transfer knowledge to new tasks is enhanced by Modular learning. Modular neural networks (MNN) architectures are emerging solutions for positive learning transfer while avoiding negative interference [226]. Several MNN architectures have been developed in the literature, such as the modular architecture proposed by Rahaman et al. for an application to geospatial data [227], the Perceiver IO architecture proposed by Jaegle et al. [228], or the Neural Attentive Circuits architecture proposed by Rahaman [229]. In a Modular learning process, three main phases must be considered [226]:

1. The learning module management phase which must meet the training, management and storage objectives of the modules.
2. The routing phase which aims to define the way in which the modules are chosen and activated in order to meet a specific objective.
3. The aggregation phase whose objective is to construct the final response from the responses of the various modules requested by the routing.

5.2. Self-Supervised Learning

In most industrial processes monitored under real operating conditions, a large amount of data is collected daily by the various sensors. However, only a part of this data can be properly used, and a very negligible part of the data is labeled by experts. In fact, labeling is generally a very time-consuming process that is usually performed by human experts. In addition, most of the data comes from normal asset behavior. As a result, the data collected is generally not representative of the different degradation mechanisms.

Learning techniques can be divided into two main families: supervised and unsupervised learning. Learning deep neural networks with supervised techniques requires a large amount of labeled data to achieve an acceptable level of performance. For the reasons mentioned above, these techniques are not suitable for industrial applications. Furthermore, unsupervised techniques, which do not require labeled data, can be effective for fault detection applications, but lack knowledge of degradation to effectively perform more sophisticated tasks such as diagnosis or prognosis.

SSL is an unsupervised learning paradigm that explores effective feature representations from unlabeled data [10]. Unlike conventional supervised learning, which requires an abundance of labeled data, SSL exploits the underlying information of unlabeled data, reducing the dependence on the annotation phase, which is very time consuming. The general principle of SSL is to create pretext tasks to allow the model to acquire efficient and relevant representations during the task solving phase. For example, learning to reconstruct noisy or partially hidden input allows the model to extract relevant features that could be used by another classification model. SSL has the advantage of exploiting the inherent properties of the data to allow the model to learn to extract high-level global descriptors from a large amount of unlabeled data. Therefore, a significant number of research papers on SSL have been published recently [230–236].

Figure 11 shows a classic diagram of the use of SSL in industrial monitoring applications. In a first phase, SSL pre-trains learning models using simple pretext tasks. This phase allows the model to extract complex features from the data without the need for human expertise. In a second phase, the model is refined by supervised learning with a minimum of labeled data, reducing the cost of human intervention.

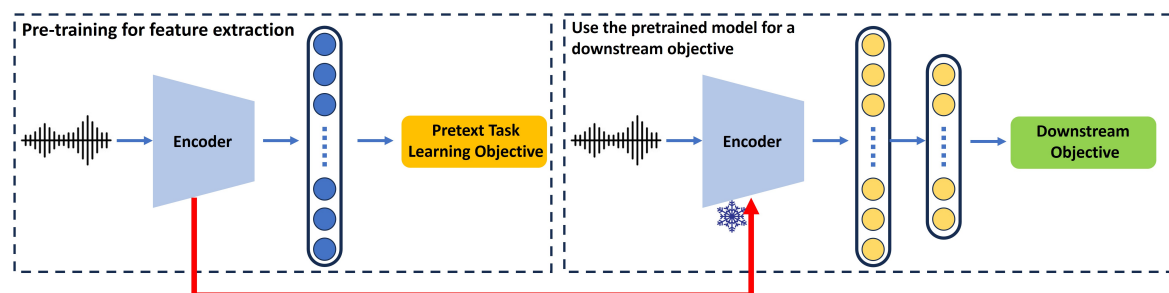


Figure 11. Typical application diagram of self-learning in a PHM process: pre-training of the model with SSL by defining pretext tasks, followed by knowledge transfer with supervised learning from a minimum of data labeled.

5.3. Multimodal Fusion

Industrial systems can generate many types of data from different sensors, such as time signals, images, videos, and also a significant amount of textual information, such as maintenance work orders, maintenance reports, and so on. Therefore, the field of PHM is gradually beginning to emphasize multi-sensor data fusion to better understand physical phenomena and various degradation mechanisms. This is reflected in the recent publication of several papers that have explored the principle of information fusion [237–261].

MMF requires some prior processing of each of the modalities before they can be fused. Indeed, two different modalities will have two different digital representations, e.g. an image and an acoustic signal. Their fusion cannot be done directly in the data space, but requires a descriptor extraction phase. The Figure 12 shows a typical example of fusion of two different modalities. Each modality has its own neural network that extracts the relevant features, which are then fused. The same classifier can then be trained directly on the shared feature space.

The MMF involves three main steps. The first is to train the neural networks that extract the features of each modality. This can be done using methods such as self-learning. Next, the obtained pre-trained models are frozen and used for the next step. The features of each modality are then extracted from one of the hidden layers before the output layer. Their fusion can be done either directly, e.g. by concatenation [256], or by using newer techniques such as transformers or attention

mechanisms [238,245,249,254,260]. Finally, the last step is to learn a classifier from the obtained feature space.

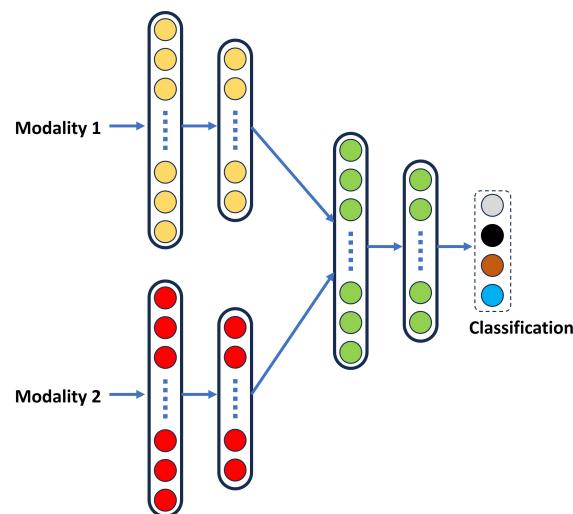


Figure 12. Principle of MMF by unique learning. Part of the model is common to both modalities [241].

5.4. Towards the Foundation Models

Foundation models offer a potentially effective solution to the PHM process for complex industrial systems, as their advanced cognitive abilities allow them to solve certain complex reasoning tasks [262]. Therefore, the success of these foundation models marks the transformation of the research paradigm in AI, where we move from a mono-modal, single-task research paradigm with limited data to a multimodal, MTL research paradigm with big data and large-scale foundation models [262]. Looking at the evolution of the last few years, there is a gradual move from a collection of monolithic architectures for narrow and unique tasks to a set of modular and reconfigurable architectures that can handle different types of tasks [262,263].

This new class of foundation models is composed of billions of parameters [263,264] and is trained on massive amounts of data. The concept of modular learning, which is a new paradigm in ML [226,265,266], is at the core of the foundation models. In addition, the recent development of Transformer architectures [266–268], has provided great opportunities for extracting complex features from foundation models. Furthermore, the concept of SSL has enabled neural networks to have a robust capacity for unsupervised representation of features and descriptors [10]. Finally, MMF algorithms, MTL combined with the attention mechanism, have allowed foundation models to interact between the different modalities of the learning data [269,270].

However, integrating an foundation model based solely on data-driven learning into a PHM process requires a high level of AI expertise and rigorous methodology. Three main aspects need to be considered: the modularity, reliability and explainability of the foundation models. In fact, a modular AI structure with knowledge distributed across multiple AI models is better suited to the PHM process of complex industrial system than a single monolithic model. Thanks to this modularity, the AI models will be more explainable, transferable from one asset to another and easily upgradeable without the constraint of catastrophic forgetting.

6. What Are the Challenges?

6.1. Modularity of the LSF Models

A major drawback of large-scale models is that they are based on monolithic structures. This leads to several undesirable consequences related to adaptability. First, it is very difficult to refine the model after the appearance of a new target task. Second, the explainability of the model is a real challenge when the justification of a given response is important for decision making. For several years, some research efforts have been directed towards the learning of disentangled representations [271,272], the

learning of causal representations with the hypothesis of independent causal mechanisms [273], and more explicitly, the learning of modular representations [265,274]. Modular architectures try to take advantage of this structure by allowing the learning of systems of sparse interacting neural modules. If we observe the developments of the last decade along this axis, we are gradually moving from a collection of monolithic architectures for narrow tasks to a modular and re-configurable architecture to handle different types of tasks.

Another expected benefit of modular deep learning is a better ability to transfer to new tasks. Transfer learning involves taking the "knowledge" learned from one specific task and applying it to another task in a different context. However, developing a model that can perform multiple tasks without suffering from negative interference between tasks and with good generalization performance on non-identically distributed data remains a significant challenge. Modular NN architectures represent emerging solutions for positive transfer learning while avoiding negative interference (catastrophic forgetting phenomenon).

6.2. Reliability of the LSF Models

When designing a ML model, it is common to focus on performance metrics based on the accuracy obtained on a test set drawn from the same distribution as the training set: this is called the Identical and Independent Distribution assumption. However, this does not take into account the deployment of ML systems in the real world, such as modern power grid systems, where the test environment is often very different from the learning environment. To improve the reliability of ML systems, three main conditions must be considered: models must represent their own uncertainty, they must generalize robustly to new scenarios and they must be able to adapt effectively to new data [275].

- **Quantifying uncertainty.** Quantifying prediction uncertainty allows practitioners to know when to trust model predictions. Various metrics can be used to quantify the quality of uncertainty, such as expected calibration error, which measures how well confidence in the model matches its accuracy. Quantifying uncertainty also helps improve decision making; a popular framework is selective prediction, where a model can refer its prediction to human experts when it is uncertain. Another popular task is open-set recognition, where the model encounters inputs from new classes at test time that were not seen during training, and the goal is to reliably detect that these inputs belong to one of the training classes.
- **Robust generalization.** Robust generalization involves making an estimate or prediction about something that is not seen. Prediction quality is typically measured in terms of accuracy (e.g., top-1 error for classification problems and root mean square error for regression problems) and appropriate scoring rules such as log-likelihood and Brier score. In the real world, we are interested not only in measurements on new data from the same distribution on which the model was trained, but also in robustness, measured by measurements on data subject to non-distributional changes, such as changes in covariates or subpopulations.
- **Adaptation.** Adaptation consists of testing the capabilities of the model during its learning process. Benchmarks typically evaluate static datasets with a predefined split between training and testing. In many applications, however, we are interested in models that can quickly adapt to new data and learn efficiently with as few labeled examples as possible. Examples include few-shot learning, where the model learns from a small set of examples; active learning, where the model not only learns but also participates in the acquisition of the data from which it learns; and lifelong learning, where the model learns during a sequence of tasks and must not forget information relevant to previous tasks.

6.3. Explainability of the LSF Models

ML models are considered black boxes because it is very difficult to understand how these models work in practice, despite their widespread use and exceptional performance. Therefore, it is difficult

for experts to trust and justify the decisions and recommendations made by these models in the field of power systems, where a high level of responsibility is required.

In recent years, eXplainable Artificial Intelligence (XAI) techniques have been developed to improve the explainability of ML models so that their results can be better understood [276–281]. There are several challenges and limitations that must be addressed when implementing XAI for power system applications. One of the main challenges is to use models that are both efficient and transparent. In general, accurate models are more complex and harder to understand. This general compromise is especially important in the field of electric power systems, where a typical user generally requires both high performance and precise exposition in order to enjoy a high level of trust.

In addition, the lack of standardization and clear definitions are one of the main limitations of XAI. Currently, while some works and studies define what explainability is, there is still no consensus on a specific definition of XAI and explainability. Some works focus on visualization methods, while others use the concept of feature importance or relevance. Another limitation of XAI techniques is the lack of metrics for evaluating the quality of the explanation. Although a clear definition of explainability can be provided, it is desirable to have a metric for evaluating the degree of explainability of a model. These metrics should measure an explainability score for each XAI technique, on each model.

7. Conclusion

This paper presents the integration of ML techniques in the field of PT-PHM. It provides a comprehensive review of the state of the art in PT-PHM by analysing more than 200 papers, mostly published in scientific journals. This analysis is divided into two parts: a part dealing with classical ML techniques, such as shallow ANN architectures, SVM, fuzzy inference systems, KNN; and a second part dealing with DL architectures, such as CNN, AE, attention mechanism and GNN.

After the first revolution in ML models, moving from simple ML models requiring significant feature extraction (AI 1.0) to deeper models with complex learning (AI 2.0), a new paradigm is emerging: Large-scale foundation models (AI 3.0). Despite the fact that scientific research in the field of PT-PHM is still very conservative, mainly in AI 1.0 and gradually moving towards AI 2.0, all the ingredients are in place for the transition to AI 3.0 techniques: more and more multimodal PT monitoring data are being collected such as DGA, infrared images, vibration analysis or thermal monitoring. New techniques such as Transformer-based Deep Neural Networks, Self-Supervised Learning or Multimodal Fusion should be used to optimize the PT-PHM process.

Abbreviations

The following abbreviations are used in this manuscript:

AE	Auto-Encoder
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference Systems
BI	Bayesian Inference
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DGA	Dissolved Gas Analysis
DML	Deep Machine Learning
DNN	Deep Neural Networks
DT	Decision Tree
EL	Ensemble Learning
ELM	Extreme Learning Machine
FDD	Fault Detection and diagnosis
FIS	Fuzzy Inference Systems
GAN	Generative Adversarial Network
GCN	Graph Convolutional Network
GNN	Graph Neural Network
GP	Gaussian Process
HI	Health Index
HMM	Hidden Markov model
KNN	K-Nearest Neighbors
LSF	Large-scale foundation models
ML	Machine Learning
MLP	Multi-Layer Perceptron
MMF	Multimodal Fusion
MNN	Modular neural networks
MTL	Multi-Task Learning
PCA	Principal Component Analysis
PHM	Prognostics and Health Management
PINN	Physics-Informed Neural Networks
PT	Power transformers
RF	Random Forest
RNN	Recurrent Neural Network
SSL	Self Supervised Learning
SVM	Support Vector Machine
VAE	Variational Auto-Encoder
WN	Wavelet Networks
XAI	eXplainable Artificial Intelligence

References

1. Dong, M.; Zheng, H.; Zhang, Y.; Shi, K.; Yao, S.; Kou, X.; Ding, G.; Guo, L. A Novel Maintenance Decision Making Model of Power Transformers Based on Reliability and Economy Assessment. *IEEE Access* **2019**, *7*, 28778–28790. <https://doi.org/10.1109/ACCESS.2019.2897606>.
2. Ma, H.; Saha, T.K.; Ekanayake, C.; Martin, D. Smart Transformer for Smart Grid—Intelligent Framework and Techniques for Power Transformer Asset Management. *IEEE Transactions on Smart Grid* **2015**, *6*, 1026–1034. <https://doi.org/10.1109/TSG.2014.2384501>.
3. Koziel, S.; Hilber, P.; Westerlund, P.; Shayesteh, E. Investments in data quality: Evaluating impacts of faulty data on asset management in power systems. *Applied Energy* **2021**, *281*, 116057. <https://doi.org/https://doi.org/10.1016/j.apenergy.2020.116057>.
4. Huang, Y.C.; Sun, H.C. Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic. *IEEE Transactions on Dielectrics and Electrical Insulation* **2013**, *20*, 974–981. <https://doi.org/10.1109/TDEI.2013.6518967>.

5. Sun, H.C.; Huang, Y.C.; Huang, C.M. Fault Diagnosis of Power Transformers Using Computational Intelligence: A Review. *Energy Procedia* **2012**, *14*, 1226–1231. 2011 2nd International Conference on Advances in Energy Engineering (ICAEE), <https://doi.org/https://doi.org/10.1016/j.egypro.2011.12.1080>.
6. Cheng, L.; Yu, T. Dissolved Gas Analysis Principle-Based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil-Immersed Power Transformers: A Survey. *Energies* **2018**, *11*. <https://doi.org/10.3390/en11040913>.
7. Zhang, Y.; Tang, Y.; Liu, Y.; Liang, Z. Fault diagnosis of transformer using artificial intelligence: A review. *Frontiers in Energy Research* **2022**, *10*. <https://doi.org/10.3389/fenrg.2022.1006474>.
8. Wang, L.; Littler, T.; Liu, X. Hybrid AI model for power transformer assessment using imbalanced DGA datasets. *IET Renewable Power Generation* **2023**, *17*, 1912–1922. [_eprint: https://doi.org/https://doi.org/10.1049/rpg2.12733](https://doi.org/https://doi.org/10.1049/rpg2.12733), <https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/rpg2.12733>, <https://doi.org/https://doi.org/10.1049/rpg2.12733>.
9. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need, 2017. <https://doi.org/10.48550/ARXIV.1706.03762>.
10. Jing, L.; Tian, Y. Self-Supervised Visual Feature Learning With Deep Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *43*, 4037–4058. <https://doi.org/10.1109/TPAMI.2020.2992393>.
11. Zhang, L.; Lin, J.; Liu, B.; Zhang, Z.; Yan, X.; Wei, M. A Review on Deep Learning Applications in Prognostics and Health Management. *IEEE Access* **2019**, *7*, 162415–162438. <https://doi.org/10.1109/ACCESS.2019.2950985>.
12. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.B.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the Opportunities and Risks of Foundation Models. *CoRR* **2021**, *abs/2108.07258*, [2108.07258].
13. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019, [arXiv:cs.CL/1810.04805].
14. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners, 2020, [arXiv:cs.CL/2005.14165].
15. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision, 2021, [arXiv:cs.CV/2103.00020].
16. Alqudsi, A.; El-Hag, A. Application of Machine Learning in Transformer Health Index Prediction. *Energies* **2019**, *12*. <https://doi.org/10.3390/en12142694>.
17. Mohmad, A.; Shapiyai, M.I.; Shamsudin, M.S.; Abu, M.A.; Hamid, A.A. Investigating performance of transformer health index in machine learning application using dominant features. *Journal of Physics: Conference Series* **2021**, *2128*, 012025. Publisher: IOP Publishing, <https://doi.org/10.1088/1742-6596/2128/1/012025>.
18. Zeinoddini-Meymand, H.; Kamel, S.; Khan, B. An Efficient Approach With Application of Linear and Nonlinear Models for Evaluation of Power Transformer Health Index. *IEEE Access* **2021**, *9*, 150172–150186. <https://doi.org/10.1109/ACCESS.2021.3124845>.
19. Ghoneim, S.S.M.; Taha, I.B.M. Comparative Study of Full and Reduced Feature Scenarios for Health Index Computation of Power Transformers. *IEEE Access* **2020**, *8*, 181326–181339. <https://doi.org/10.1109/ACCESS.2020.3028689>.
20. He, Q.; Si, J.; Tylavsky, D. Prediction of top-oil temperature for transformers using neural networks. *IEEE Transactions on Power Delivery* **2000**, *15*, 1205–1211. <https://doi.org/10.1109/61.891504>.
21. Doolgindachbaporn, A.; Callender, G.; Lewin, P.L.; Simonson, E.; Wilson, G. Data Driven Transformer Thermal Model for Condition Monitoring. *IEEE Transactions on Power Delivery* **2022**, *37*, 3133–3141. <https://doi.org/10.1109/TPWRD.2021.3123957>.
22. Guardado, J.; Naredo, J.; Moreno, P.; Fuerte, C. A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis. *IEEE Transactions on Power Delivery* **2001**, *16*, 643–647. <https://doi.org/10.1109/61.956751>.
23. Barbosa, F.R.; Almeida, O.M.; Braga, A.P.; Amora, M.A.; Cartaxo, S.J. Application of an artificial neural network in the use of physicochemical properties as a low cost proxy of power transformers DGA data. *IEEE Transactions on Dielectrics and Electrical Insulation* **2012**, *19*, 239–246. <https://doi.org/10.1109/TDEI.2012.6148524>.

24. Bhalla, D.; Bansal, R.K.; Gupta, H.O. Function analysis based rule extraction from artificial neural networks for transformer incipient fault diagnosis. *International Journal of Electrical Power & Energy Systems* **2012**, *43*, 1196–1203. <https://doi.org/https://doi.org/10.1016/j.ijepes.2012.06.042>.
25. Yang, M.T.; Hu, L.S. Intelligent fault types diagnostic system for dissolved gas analysis of oil-immersed power transformer. *IEEE Transactions on Dielectrics and Electrical Insulation* **2013**, *20*, 2317–2324. <https://doi.org/10.1109/TDEI.2013.6678885>.
26. Ghoneim, S.S.M.; Taha, I.B.M.; Elkalashy, N.I. Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* **2016**, *23*, 1838–1845. <https://doi.org/10.1109/TDEI.2016.005301>.
27. Illias, H.A.; Chai, X.R.; Abu Bakar, A.H. Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. *Measurement: Journal of the International Measurement Confederation* **2016**, *90*, 94 – 102. Cited by: 78, <https://doi.org/10.1016/j.measurement.2016.04.052>.
28. Velásquez, R.M.A.; Lara, J.V.M. Root cause analysis improved with machine learning for failure analysis in power transformers. *Engineering Failure Analysis* **2020**, *115*, 104684. <https://doi.org/https://doi.org/10.1016/j.engfailanal.2020.104684>.
29. Taha, I.B.M.; Dessouky, S.S.; Ghoneim, S.S.M. Transformer fault types and severity class prediction based on neural pattern-recognition techniques. *Electric Power Systems Research* **2021**, *191*, 106899. <https://doi.org/https://doi.org/10.1016/j.epsr.2020.106899>.
30. Ahmadi, S.A.; Sanaye-Pasand, M. A Robust Multi-Layer Framework for Online Condition Assessment of Power Transformers. *IEEE Transactions on Power Delivery* **2022**, *37*, 947–954. <https://doi.org/10.1109/TPWRD.2021.3074545>.
31. Souahlia, S.; Bacha, K.; Chaari, A. MLP neural network-based decision for power transformers fault diagnosis using an improved combination of Rogers and Doernenburg ratios DGA. *International Journal of Electrical Power & Energy Systems* **2012**, *43*, 1346–1353. <https://doi.org/https://doi.org/10.1016/j.ijepes.2012.05.067>.
32. Thango, B.A. On the Application of Artificial Neural Network for Classification of Incipient Faults in Dissolved Gas Analysis of Power Transformers. *Machine Learning and Knowledge Extraction* **2022**, *4*, 839–851. <https://doi.org/10.3390/make4040042>.
33. Trappey, A.J.C.; Trappey, C.V.; Ma, L.; Chang, J.C.M. Intelligent engineering asset management system for power transformer maintenance decision supports under various operating conditions. *Computers & Industrial Engineering* **2015**, *84*, 3–11. <https://doi.org/https://doi.org/10.1016/j.cie.2014.12.033>.
34. Amidedin Mousavi, S.; Hekmati, A.; Sedighzadeh, M.; Bigdeli, M.; Bazargan, A. ANN based temperature compensation for variations in polarization and depolarization current measurements in transformer. *Thermal Science and Engineering Progress* **2020**, *20*, 100671. <https://doi.org/https://doi.org/10.1016/j.tsep.2020.100671>.
35. Setayeshmehr, A.; Akbari, A.; Borsi, H.; Gockenbach, E. On-line monitoring and diagnoses of power transformer bushings. *IEEE Transactions on Dielectrics and Electrical Insulation* **2006**, *13*, 608–615. <https://doi.org/10.1109/TDEI.2006.1657975>.
36. Rigatos, G.; Siano, P. Power transformers' condition monitoring using neural modeling and the local statistical approach to fault diagnosis. *International Journal of Electrical Power & Energy Systems* **2016**, *80*, 150–159. <https://doi.org/https://doi.org/10.1016/j.ijepes.2016.01.019>.
37. Islam, M.M.; Lee, G.; Hettiwatte, S.N. Application of a general regression neural network for health index calculation of power transformers. *International Journal of Electrical Power & Energy Systems* **2017**, *93*, 308–315. <https://doi.org/https://doi.org/10.1016/j.ijepes.2017.06.008>.
38. Islam, N.; Khan, R.; Das, S.K.; Sarker, S.K.; Islam, M.M.; Akter, M.; Muyeen, S.M. Power transformer health condition evaluation: A deep generative model aided intelligent framework. *Electric Power Systems Research* **2023**, *218*, 109201. <https://doi.org/https://doi.org/10.1016/j.epsr.2023.109201>.
39. Trappey, A.J.C.; Trappey, C.V.; Ma, L.; Chang, J.C. Integrating Real-Time Monitoring and Asset Health Prediction for Power Transformer Intelligent Maintenance and Decision Support. In Proceedings of the Engineering Asset Management - Systems, Professional Practices and Certification; Tse, P.W.; Mathew, J.; Wong, K.; Lam, R.; Ko, C., Eds., Cham, 2015; pp. 533–543. https://doi.org/https://doi.org/10.1007/978-3-319-09507-3_46.
40. Benhmed, K.; Mooman, A.; Younes, A.; Shaban, K.; El-Hag, A. Feature Selection for Effective Health Index Diagnoses of Power Transformers. *IEEE Transactions on Power Delivery* **2018**, *33*, 3223–3226. <https://doi.org/10.1109/TPWRD.2017.2762920>.

41. Alqudsi, A.; El-Hag, A. Assessing the power transformer insulation health condition using a feature-reduced predictor mode. *IEEE Transactions on Dielectrics and Electrical Insulation* **2018**, *25*, 853–862. <https://doi.org/10.1109/TDEI.2018.006630>.
42. Liu, J.; Ding, Z.; Fan, X.; Geng, C.; Song, B.; Wang, Q.; Zhang, Y. A BPNN Model-Based AdaBoost Algorithm for Estimating Inside Moisture of Oil–Paper Insulation of Power Transformer. *IEEE Transactions on Dielectrics and Electrical Insulation* **2022**, *29*, 614–622. <https://doi.org/10.1109/TDEI.2022.3157909>.
43. Hooshmand, R.A.; Parastegari, M.; Forghani, Z. Adaptive neuro-fuzzy inference system approach for simultaneous diagnosis of the type and location of faults in power transformers. *IEEE Electrical Insulation Magazine* **2012**, *28*, 32–42. <https://doi.org/10.1109/MEI.2012.6268440>.
44. Khan, S.A.; Equbal, M.D.; Islam, T. A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models. *IEEE Transactions on Dielectrics and Electrical Insulation* **2015**, *22*, 590 – 596. Cited by: 146, <https://doi.org/10.1109/TDEI.2014.004478>.
45. Kari, T.; Gao, W.; Zhao, D.; Zhang, Z.; Mo, W.; Wang, Y.; Luan, L. An integrated method of ANFIS and Dempster-Shafer theory for fault diagnosis of power transformer. *IEEE Transactions on Dielectrics and Electrical Insulation* **2018**, *25*, 360–371. <https://doi.org/10.1109/TDEI.2018.006746>.
46. Fan, J.; Wang, F.; Sun, Q.; Bin, F.; Liang, F.; Xiao, X. Hybrid RVM–ANFIS algorithm for transformer fault diagnosis. *IET Generation, Transmission & Distribution* **2017**, *11*, 3637–3643, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2017.0547>]. <https://doi.org/https://doi.org/10.1049/iet-gtd.2017.0547>.
47. Nezami, M.M.; Equbal, M.D.; Khan, S.A.; Sohail, S. An ANFIS Based Comprehensive Correlation Between Diagnostic and Destructive Parameters of Transformer’s Paper Insulation. *Arabian Journal for Science and Engineering* **2021**, *46*, 1541–1547. <https://doi.org/10.1007/s13369-020-05180-4>.
48. Medina, R.D.; Zaldivar, D.A.; Romero, A.A.; Zuñiga, J.; Mombello, E.E. A fuzzy inference-based approach for estimating power transformers risk index. *Electric Power Systems Research* **2022**, *209*, 108004. <https://doi.org/https://doi.org/10.1016/j.epsr.2022.108004>.
49. Prasojo, R.A.; Diwyacitta, K.; Suwarno.; Gumilang, H. Transformer Paper Expected Life Estimation Using ANFIS Based on Oil Characteristics and Dissolved Gases (Case Study: Indonesian Transformers). *Energies* **2017**, *10*. <https://doi.org/10.3390/en10081135>.
50. Thango, B.A.; Bokoro, P.N. A Technique for Transformer Remnant Cellulose Life Cycle Prediction Using Adaptive Neuro-Fuzzy Inference System. *Processes* **2023**, *11*. <https://doi.org/10.3390/pr11020440>.
51. Soni, R.; Mehta, B. Diagnosis and prognosis of incipient faults and insulation status for asset management of power transformer using fuzzy logic controller & fuzzy clustering means. *Electric Power Systems Research* **2023**, *220*, 109256. <https://doi.org/https://doi.org/10.1016/j.epsr.2023.109256>.
52. Ganyun, L.; Haozhong, C.; Haibao, Z.; Lixin, D. Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electric Power Systems Research* **2005**, *74*, 1–7. <https://doi.org/https://doi.org/10.1016/j.epsr.2004.07.008>.
53. Liao, R.; Zheng, H.; Grzybowski, S.; Yang, L. Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electric Power Systems Research* **2011**, *81*, 2074–2080. <https://doi.org/https://doi.org/10.1016/j.epsr.2011.07.020>.
54. Bacha, K.; Souahlia, S.; Gossa, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electric Power Systems Research* **2012**, *83*, 73–79. <https://doi.org/https://doi.org/10.1016/j.epsr.2011.09.012>.
55. Wei, C.; Tang, W.; Wu, Q. Dissolved gas analysis method based on novel feature prioritisation and support vector machine. *IET Electric Power Applications* **2014**, *8*, 320–328, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-epa.2014.0085>]. <https://doi.org/https://doi.org/10.1049/iet-epa.2014.0085>.
56. Wu, J.; Li, K.; Sun, J.; Xie, L. A Novel Integrated Method to Diagnose Faults in Power Transformers. *Energies* **2018**, *11*. <https://doi.org/10.3390/en11113041>.
57. Zheng, H.; Zhang, Y.; Liu, J.; Wei, H.; Zhao, J.; Liao, R. A novel model based on wavelet LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers. *Electric Power Systems Research* **2018**, *155*, 196–205. <https://doi.org/https://doi.org/10.1016/j.epsr.2017.10.010>.
58. Fan, Q.; Yu, F.; Xuan, M. Transformer fault diagnosis method based on improved whale optimization algorithm to optimize support vector machine. *Energy Reports* **2021**, *7*, 856–866. <https://doi.org/https://doi.org/10.1016/j.egyr.2021.09.188>.

59. Benmahamed, Y.; Kherif, O.; Tegar, M.; Boubakeur, A.; Ghoneim, S.S.M. Accuracy Improvement of Transformer Faults Diagnostic Based on DGA Data Using SVM-BA Classifier. *Energies* **2021**, *14*. <https://doi.org/10.3390/en14102970>.
60. Wu, Y.; Sun, X.; Zhang, Y.; Zhong, X.; Cheng, L. A Power Transformer Fault Diagnosis Method-Based Hybrid Improved Seagull Optimization Algorithm and Support Vector Machine. *IEEE Access* **2022**, *10*, 17268–17286. <https://doi.org/10.1109/ACCESS.2021.3127164>.
61. Benmahamed, Y.; Tegar, M.; Boubakeur, A. Application of SVM and KNN to Duval Pentagon 1 for transformer oil diagnosis. *IEEE Transactions on Dielectrics and Electrical Insulation* **2017**, *24*, 3443–3451. <https://doi.org/10.1109/TDEI.2017.006841>.
62. Dhini, A.; Surjandari, I.; Kusumoputro, B.; Faqih, A.; Kusiak, A. Data-driven Fault Diagnosis of Power Transformers using Dissolved Gas Analysis (DGA). *International Journal of Technology* **2020**, *11*, 388–399. <https://doi.org/https://doi.org/10.14716/ijtech.v11i2.3625>.
63. Kari, T.; Gao, W.; Zhao, D.; Abiderexiti, K.; Mo, W.; Wang, Y.; Luan, L. Hybrid feature selection approach for power transformer fault diagnosis based on support vector machine and genetic algorithm. *IET Generation, Transmission & Distribution* **2018**, *12*, 5672–5680, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2018.5482>]. <https://doi.org/https://doi.org/10.1049/iet-gtd.2018.5482>.
64. Ilias, H.A.; Zhao Liang, W. Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimisation. *PLOS ONE* **2018**, *13*, 1–15. <https://doi.org/10.1371/journal.pone.0191366>.
65. Ma, H.; Saha, T.K.; Ekanayake, C. Predictive learning and information fusion for condition assessment of power transformer. In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, 2011, pp. 1–8. ISSN: 1944-9925, <https://doi.org/10.1109/PES.2011.6039069>.
66. Thango, B.A. Dissolved Gas Analysis and Application of Artificial Intelligence Technique for Fault Diagnosis in Power Transformers: A South African Case Study. *Energies* **2022**, *15*. <https://doi.org/10.3390/en15239030>.
67. Hua, Y.; Sun, Y.; Xu, G.; Sun, S.; Wang, E.; Pang, Y. A fault diagnostic method for oil-immersed transformer based on multiple probabilistic output algorithms and improved DS evidence theory. *International Journal of Electrical Power & Energy Systems* **2022**, *137*, 107828. <https://doi.org/https://doi.org/10.1016/j.ijepes.2021.107828>.
68. Hong, L.; Chen, Z.; Wang, Y.; Shahidehpour, M.; Wu, M. A novel SVM-based decision framework considering feature distribution for Power Transformer Fault Diagnosis. *Energy Reports* **2022**, *8*, 9392–9401. <https://doi.org/https://doi.org/10.1016/j.egy.2022.07.062>.
69. Das, S.; Paramane, A.; Chatterjee, S.; Rao, U.M. Accurate Identification of Transformer Faults From Dissolved Gas Data Using Recursive Feature Elimination Method. *IEEE Transactions on Dielectrics and Electrical Insulation* **2023**, *30*, 466–473. <https://doi.org/10.1109/TDEI.2022.3215936>.
70. Zou, H.; Huang, F. A novel intelligent fault diagnosis method for electrical equipment using infrared thermography. *Infrared Physics & Technology* **2015**, *73*, 29–35. <https://doi.org/https://doi.org/10.1016/j.infrared.2015.08.019>.
71. Zhao, Z.; Tang, C.; Zhou, Q.; Xu, L.; Gui, Y.; Yao, C. Identification of Power Transformer Winding Mechanical Fault Types Based on Online IFRA by Support Vector Machine. *Energies* **2017**, *10*. <https://doi.org/10.3390/en10122022>.
72. Ahmed, A. Power Transformer Condition Monitoring and Diagnosis. *The Institution of Engineering and Technology: Stevenage, UK* **2018**.
73. Prasojo, R.A.; et al. Power transformer paper insulation assessment based on oil measurement data using SVM-classifier. *International Journal on Electrical Engineering and Informatics* **2018**, *10*, 661–673.
74. Tavakoli, A.; Maria, L.D.; Valecillos, B.; Bartalesi, D.; Garatti, S.; Bittanti, S. A Machine Learning approach to fault detection in transformers by using vibration data. *IFAC-PapersOnLine* **2020**, *53*, 13656–13661. <https://doi.org/https://doi.org/10.1016/j.ifacol.2020.12.866>.
75. Zhang, Y.; Li, J.; Fan, X.; Liu, J.; Zhang, H. Moisture Prediction of Transformer Oil-Immersed Polymer Insulation by Applying a Support Vector Machine Combined with a Genetic Algorithm. *Polymers* **2020**, *12*. <https://doi.org/10.3390/polym12071579>.
76. Arias Velásquez, R.M. Support vector machine and tree models for oil and Kraft degradation in power transformers. *Engineering Failure Analysis* **2021**, *127*, 105488. <https://doi.org/https://doi.org/10.1016/j.engfailanal.2021.105488>.

77. Kazemi, Z.; Naseri, F.; Yazdi, M.; Farjah, E. An EKF-SVM machine learning-based approach for fault detection and classification in three-phase power transformers. *IET Science, Measurement & Technology* **2021**, *15*, 130–142, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/smt2.12015>]. <https://doi.org/https://doi.org/10.1049/smt2.12015>.
78. Ashkezari, A.D.; Ma, H.; Saha, T.K.; Cui, Y. Investigation of feature selection techniques for improving efficiency of power transformer condition assessment. *IEEE Transactions on Dielectrics and Electrical Insulation* **2014**, *21*, 836–844. <https://doi.org/10.1109/TDEI.2013.004090>.
79. Ibrahim, K.; Sharkawy, R.; Temraz, H.; Salama, M. Selection criteria for oil transformer measurements to calculate the Health Index. *IEEE Transactions on Dielectrics and Electrical Insulation* **2016**, *23*, 3397–3404. <https://doi.org/10.1109/TDEI.2016.006058>.
80. Panwar, R.; Meena, V.S.; Negi, A.S.; Jarial, R. Ranking of power transformers on the basis of their health index and fault detection on the basis of DGA results using support vector machine (SVM). *Int. J. Eng. Technol. Manag. Appl. Sci* **2017**, *5*, 393–397.
81. Wijethunge, T.; Tharkana, P.; Wimalaweera, A.; Wijayakulasooriya, J.; Kumara, S.; Bandara, K.; Fernando, M. A Machine Learning Approach for FDS Based Power Transformer Moisture Estimation. In Proceedings of the 2021 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), 2021, pp. 539–542. ISSN: 2576-2397, <https://doi.org/10.1109/CEIDP50766.2021.9705366>.
82. Ma, H.; Saha, T.K.; Ekanayake, C. Statistical learning techniques and their applications for condition assessment of power transformer. *IEEE Transactions on Dielectrics and Electrical Insulation* **2012**, *19*, 481–489. <https://doi.org/10.1109/TDEI.2012.6180241>.
83. Ma, H.; Ekanayake, C.; Saha, T.K. Power transformer fault diagnosis under measurement originated uncertainties. *IEEE Transactions on Dielectrics and Electrical Insulation* **2012**, *19*, 1982–1990. <https://doi.org/10.1109/TDEI.2012.6396956>.
84. Sahri, Z.; Yusof, R.; Watada, J. FINNIM: Iterative Imputation of Missing Values in Dissolved Gas Analysis Dataset. *IEEE Transactions on Industrial Informatics* **2014**, *10*, 2093–2102. <https://doi.org/10.1109/TII.2014.2350837>.
85. Ashkezari, A.D.; Ma, H.; Saha, T.K.; Ekanayake, C. Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers. *IEEE Transactions on Dielectrics and Electrical Insulation* **2013**, *20*, 965–973. <https://doi.org/10.1109/TDEI.2013.6518966>.
86. Li, S.; Ge, Z.; Abu-Siada, A.; Yang, L.; Li, S.; Wakimoto, K. A New Technique to Estimate the Degree of Polymerization of Insulation Paper Using Multiple Aging Parameters of Transformer Oil. *IEEE Access* **2019**, *7*, 157471–157479. <https://doi.org/10.1109/ACCESS.2019.2949580>.
87. Dias, L.; Ribeiro, M.; Leitão, A.; Guimarães, L.; Carvalho, L.; Matos, M.A.; Bessa, R.J. An unsupervised approach for fault diagnosis of power transformers. *Quality and Reliability Engineering International* **2021**, *37*, 2834–2852, [<https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.2892>]. <https://doi.org/https://doi.org/10.1002/qre.2892>.
88. Kim, Y.; Park, T.; Kim, S.; Kwak, N.; Kweon, D. Artificial Intelligent Fault Diagnostic Method for Power Transformers using a New Classification System of Faults. *Journal of Electrical Engineering & Technology* **2019**, *14*, 825–831. <https://doi.org/10.1007/s42835-019-00105-0>.
89. Tanfilyeva, D.V.; Tanfyev, O.V.; Kazantsev, Y.V. K-nearest neighbor method for power transformers condition assessment. *IOP Conference Series: Materials Science and Engineering* **2019**, *643*, 012016. <https://doi.org/10.1088/1757-899X/643/1/012016>.
90. Kherif, O.; Benmahamed, Y.; Tegar, M.; Boubakeur, A.; Ghoneim, S.S.M. Accuracy Improvement of Power Transformer Faults Diagnostic Using KNN Classifier With Decision Tree Principle. *IEEE Access* **2021**, *9*, 81693–81701. <https://doi.org/10.1109/ACCESS.2021.3086135>.
91. Nanfak, A.; Eke, S.; Meghnefi, F.; Fofana, I.; Ngaleu, G.M.; Kom, C.H. Hybrid DGA Method for Power Transformer Faults Diagnosis Based on Evolutionary k-Means Clustering and Dissolved Gas Subsets Analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* **2023**, *30*, 2421–2428. <https://doi.org/10.1109/TDEI.2023.3275119>.
92. Harbaji, M.; Shaban, K.; El-Hag, A. Classification of common partial discharge types in oil-paper insulation system using acoustic signals. *IEEE Transactions on Dielectrics and Electrical Insulation* **2015**, *22*, 1674–1683. <https://doi.org/10.1109/TDEI.2015.7116364>.
93. Kunicki, M.; Wotzka, D. A Classification Method for Select Defects in Power Transformers Based on the Acoustic Signals. *Sensors* **2019**, *19*. <https://doi.org/10.3390/s19235212>.

94. Huang, Y.C.; Yang, H.T.; Huang, C.L. Developing a new transformer fault diagnosis system through evolutionary fuzzy logic. *IEEE Transactions on Power Delivery* **1997**, *12*, 761–767. <https://doi.org/10.1109/61.584363>.
95. Mofizul Islam, S.; Wu, T.; Ledwich, G. A novel fuzzy logic approach to transformer fault diagnosis. *IEEE Transactions on Dielectrics and Electrical Insulation* **2000**, *7*, 177–186. <https://doi.org/10.1109/94.841806>.
96. Su, Q.; Mi, C.; Lai, L.; Austin, P. A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer. *IEEE Transactions on Power Systems* **2000**, *15*, 593–598. <https://doi.org/10.1109/59.867146>.
97. Wang, M.H. A novel extension method for transformer fault diagnosis. *IEEE Transactions on Power Delivery* **2003**, *18*, 164–169. <https://doi.org/10.1109/TPWRD.2002.803838>.
98. Naresh, R.; Sharma, V.; Vashisth, M. An Integrated Neural Fuzzy Approach for Fault Diagnosis of Transformers. *IEEE Transactions on Power Delivery* **2008**, *23*, 2017–2024. <https://doi.org/10.1109/TPWRD.2008.2002652>.
99. Aghaei, J.; Gholami, A.; Shayanfar, H.A.; Dezhmakhoo, A. Dissolved gas analysis of transformers using fuzzy logic approach. *European Transactions on Electrical Power* **2010**, *20*, 630–638, [<https://onlinelibrary.wiley.com/doi/pdf/10.1002/etep.343>]. <https://doi.org/https://doi.org/10.1002/etep.343>.
100. Abu-Siada, A.; Hmood, S. A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis. *International Journal of Electrical Power & Energy Systems* **2015**, *67*, 401–408. <https://doi.org/https://doi.org/10.1016/j.ijepes.2014.12.017>.
101. Irungu, G.K.; Akumu, A.O.; Munda, J.L. A new fault diagnostic technique in oil-filled electrical equipment; the dual of Duval triangle. *IEEE Transactions on Dielectrics and Electrical Insulation* **2016**, *23*, 3405–3410. <https://doi.org/10.1109/TDEI.2016.005927>.
102. Noori, M.; Effatnejad, R.; Hajihosseini, P. Using dissolved gas analysis results to detect and isolate the internal faults of power transformers by applying a fuzzy logic method. *IET Generation, Transmission & Distribution* **2017**, *11*, 2721–2729, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2017.0028>]. <https://doi.org/https://doi.org/10.1049/iet-gtd.2017.0028>.
103. Velásquez, R.M.A.; Lara, J.V.M. Principal Components Analysis and Adaptive Decision System Based on Fuzzy Logic for Power Transformer. *Fuzzy Information and Engineering* **2017**, *9*, 493–514. <https://doi.org/https://doi.org/10.1016/j.fiae.2017.12.005>.
104. Mahmoudi, N.; Samimi, M.H.; Mohseni, H. Experiences with transformer diagnosis by DGA: case studies. *IET Generation, Transmission & Distribution* **2019**, *13*, 5431–5439, <https://doi.org/https://doi.org/10.1049/iet-gtd.2019.1056>.
105. Wani, S.A.; Gupta, D.; Farooque, M.U.; Khan, S.A. Multiple incipient fault classification approach for enhancing the accuracy of dissolved gas analysis (DGA). *IET Science, Measurement & Technology* **2019**, *13*, 959–967, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-smt.2018.5135>]. <https://doi.org/https://doi.org/10.1049/iet-smt.2018.5135>.
106. Žarković, M.; Stojković, Z. Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics. *Electric Power Systems Research* **2017**, *149*, 125–136. <https://doi.org/https://doi.org/10.1016/j.epsr.2017.04.025>.
107. Abdel-Galil, T.; Sharkawy, R.; Salama, M.; Bartnikas, R. Partial discharge pattern classification using the fuzzy decision tree approach. *IEEE Transactions on Instrumentation and Measurement* **2005**, *54*, 2258–2263. <https://doi.org/10.1109/TIM.2005.858143>.
108. Secue, J.; Mombello, E. Sweep frequency response analysis (SFRA) for the assessment of winding displacements and deformation in power transformers. *Electric Power Systems Research* **2008**, *78*, 1119–1128. <https://doi.org/https://doi.org/10.1016/j.epsr.2007.08.005>.
109. Abu-Siada, A.; Lai, S.P.; Islam, S.M. A Novel Fuzzy-Logic Approach for Furan Estimation in Transformer Oil. *IEEE Transactions on Power Delivery* **2012**, *27*, 469–474. <https://doi.org/10.1109/TPWRD.2012.2186986>.
110. Bejmert, D.; Rebizant, W.; Schiel, L. Transformer differential protection with fuzzy logic based inrush stabilization. *International Journal of Electrical Power & Energy Systems* **2014**, *63*, 51–63. <https://doi.org/https://doi.org/10.1016/j.ijepes.2014.05.056>.
111. dos Santos, G.M.; de Aquino, R.R.B.; Lira, M.M.S. Thermography and artificial intelligence in transformer fault detection. *Electrical Engineering* **2018**, *100*, 1317–1325. <https://doi.org/10.1007/s00202-017-0595-2>.
112. Zhang, Y.; Li, S.; Fan, X.; Liu, J.; Li, J. Prediction of Moisture and Aging Conditions of Oil-Immersed Cellulose Insulation Based on Fingerprints Database of Dielectric Modulus. *Polymers* **2020**, *12*. <https://doi.org/10.3390/polym12081722>.

113. Jaiswal, G.C.; Ballal, M.S.; Tutakne, D. Health index based condition monitoring of distribution transformer. In Proceedings of the 2016 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Dec 2016, pp. 1–5. <https://doi.org/10.1109/PEDES.2016.7914222>.
114. Ranga, C.; Chandel, A.K.; Chandel, R. Fuzzy Logic Expert System for Optimum Maintenance of Power Transformers. *International Journal on Electrical Engineering & Informatics* **2016**, *8*.
115. Ranga, C.; Chandel, A.K.; Chandel, R. Expert system for condition monitoring of power transformer using fuzzy logic. *Journal of Renewable and Sustainable Energy* **2017**, *9*.
116. Ranga, C.; Chandel, A.K.; Chandel, R. Condition assessment of power transformers based on multi-attributes using fuzzy logic. *IET Science, Measurement & Technology* **2017**, *11*, 983–990, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-smt.2016.0497>]. <https://doi.org/https://doi.org/10.1049/iet-smt.2016.0497>.
117. Romero-Quete, A.A.; Gómez, H.D.; Molina, J.D.; Moreno, G. A Practical method for risk assessment in power transformer fleets. *Dyna* **2017**, *84*, 11–18.
118. Kadim, E.J.; Azis, N.; Jasni, J.; Ahmad, S.A.; Talib, M.A. Transformers Health Index Assessment Based on Neural-Fuzzy Network. *Energies* **2018**, *11*. <https://doi.org/10.3390/en11040710>.
119. Sharma, J.P. Regression Approach to Power Transformer Health Assessment Using Health Index. In Proceedings of the Advances in Electromechanical Technologies; Pandey, V.C.; Pandey, P.M.; Garg, S.K., Eds., Singapore, 2021; pp. 603–616. https://doi.org/https://doi.org/10.1007/978-981-15-5463-6_54.
120. Liao, R.; Zheng, H.; Grzybowski, S.; Yang, L.; Zhang, Y.; Liao, Y. An Integrated Decision-Making Model for Condition Assessment of Power Transformers Using Fuzzy Approach and Evidential Reasoning. *IEEE Transactions on Power Delivery* **2011**, *26*, 1111–1118. <https://doi.org/10.1109/TPWRD.2010.2096482>.
121. Abu-Elanien, A.E.B.; Salama, M.M.A.; Ibrahim, M. Calculation of a Health Index for Oil-Immersed Transformers Rated Under 69 kV Using Fuzzy Logic. *IEEE Transactions on Power Delivery* **2012**, *27*, 2029–2036. <https://doi.org/10.1109/TPWRD.2012.2205165>.
122. Tang, W.H.; Goulermas, J.Y.; Wu, Q.H.; Richardson, Z.J.; Fitch, J. A Probabilistic Classifier for Transformer Dissolved Gas Analysis With a Particle Swarm Optimizer. *IEEE Transactions on Power Delivery* **2008**, *23*, 751–759. <https://doi.org/10.1109/TPWRD.2008.915812>.
123. Aizpurua, J.I.; Catterson, V.M.; Stewart, B.G.; McArthur, S.D.J.; Lambert, B.; Ampofo, B.; Pereira, G.; Cross, J.G. Power transformer dissolved gas analysis through Bayesian networks and hypothesis testing. *IEEE Transactions on Dielectrics and Electrical Insulation* **2018**, *25*, 494–506. <https://doi.org/10.1109/TDEI.2018.006766>.
124. Li, S.; Ma, H.; Saha, T.; Wu, G. Bayesian information fusion for probabilistic health index of power transformer. *IET Generation, Transmission & Distribution* **2018**, *12*, 279–287. [_eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2017.0582](https://doi.org/https://doi.org/10.1049/iet-gtd.2017.0582), <https://doi.org/https://doi.org/10.1049/iet-gtd.2017.0582>.
125. Sarajcev, P.; Jakus, D.; Vasilj, J.; Nikolic, M. Analysis of Transformer Health Index Using Bayesian Statistical Models. In Proceedings of the 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), 2018, pp. 1–7.
126. Odongo, G.; Musabe, R.; Hanyurwimfura, D. A Multinomial DGA Classifier for Incipient Fault Detection in Oil-Impregnated Power Transformers. *Algorithms* **2021**, *14*. <https://doi.org/10.3390/a14040128>.
127. Menezes, A.G.C.; Araujo, M.M.; Almeida, O.M.; Barbosa, F.R.; Braga, A.P.S. Induction of Decision Trees to Diagnose Incipient Faults in Power Transformers. *IEEE Transactions on Dielectrics and Electrical Insulation* **2022**, *29*, 279–286. <https://doi.org/10.1109/TDEI.2022.3148453>.
128. Azarakhsh, J. The power transformer differential protection using decision tree. *Bulletin de la Société Royale des Sciences de Liège* **2017**, *86*, 726–738.
129. Chen, W.; Pan, C.; Yun, Y.; Liu, Y. Wavelet Networks in Power Transformers Diagnosis Using Dissolved Gas Analysis. *IEEE Transactions on Power Delivery* **2009**, *24*, 187–194. <https://doi.org/10.1109/TPWRD.2008.2002974>.
130. Huang, Y.C.; Huang, C.M. Evolving wavelet networks for power transformer condition monitoring. *IEEE Transactions on Power Delivery* **2002**, *17*, 412–416. <https://doi.org/10.1109/61.997908>.
131. Huang, Y.C. A new data mining approach to dissolved gas analysis of oil-insulated power apparatus. *IEEE Transactions on Power Delivery* **2003**, *18*, 1257–1261. <https://doi.org/10.1109/TPWRD.2003.817736>.
132. Medeiros, R.P.; Costa, F.B. A Wavelet-Based Transformer Differential Protection With Differential Current Transformer Saturation and Cross-Country Fault Detection. *IEEE Transactions on Power Delivery* **2018**, *33*, 789–799. <https://doi.org/10.1109/TPWRD.2017.2764062>.

133. Ahmed, M.; Elkhatib, M.; Salama, M.; Shaban, K.B. Transformer Health Index estimation using Orthogonal Wavelet Network. In Proceedings of the 2015 IEEE Electrical Power and Energy Conference (EPEC), Oct 2015, pp. 120–124. <https://doi.org/10.1109/EPEC.2015.7379937>.
134. Tao, L.; Yang, X.; Zhou, Y.; Yang, L. A Novel Transformers Fault Diagnosis Method Based on Probabilistic Neural Network and Bio-Inspired Optimizer. *Sensors* **2021**, *21*. <https://doi.org/10.3390/s21113623>.
135. Wang, L.; Littler, T.; Liu, X. Gaussian Process Multi-Class Classification for Transformer Fault Diagnosis Using Dissolved Gas Analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* **2021**, *28*, 1703–1712. <https://doi.org/10.1109/TDEI.2021.009470>.
136. Meng, K.; Dong, Z.Y.; Wang, D.H.; Wong, K.P. A Self-Adaptive RBF Neural Network Classifier for Transformer Fault Analysis. *IEEE Transactions on Power Systems* **2010**, *25*, 1350–1360. <https://doi.org/10.1109/TPWRS.2010.2040491>.
137. Islam, M.; Lee, G.; Hettiwatte, S.N.; Williams, K. Calculating a Health Index for Power Transformers Using a Subsystem-Based GRNN Approach. *IEEE Transactions on Power Delivery* **2018**, *33*, 1903–1912. <https://doi.org/10.1109/TPWRD.2017.2770166>.
138. Dong, M. A Data-driven Long-Term Dynamic Rating Estimating Method for Power Transformers. *IEEE Transactions on Power Delivery* **2021**, *36*, 686–697. <https://doi.org/10.1109/TPWRD.2020.2988921>.
139. Ghoneim, S.S.M.; Farrag, T.A.; Rashed, A.A.; El-Kenawy, E.S.M.; Ibrahim, A. Adaptive Dynamic Meta-Heuristics for Feature Selection and Classification in Diagnostic Accuracy of Transformer Faults. *IEEE Access* **2021**, *9*, 78324–78340. <https://doi.org/10.1109/ACCESS.2021.3083593>.
140. Zhang, M.; Chen, W.; Zhang, Y.; Liu, F.; Yu, D.; Zhang, C.; Gao, L. Fault Diagnosis of Oil-Immersed Power Transformer Based on Difference-Mutation Brain Storm Optimized Catboost Model. *IEEE Access* **2021**, *9*, 168767–168782. <https://doi.org/10.1109/ACCESS.2021.3135283>.
141. Chen, H.C.; Zhang, Y.; Chen, M. Transformer Dissolved Gas Analysis for Highly-Imbalanced Dataset Using Multiclass Sequential Ensembled ELM. *IEEE Transactions on Dielectrics and Electrical Insulation* **2023**, *30*, 2353–2361. <https://doi.org/10.1109/TDEI.2023.3280436>.
142. Haque, N.; Jamshed, A.; Chatterjee, K.; Chatterjee, S. Accurate Sensing of Power Transformer Faults From Dissolved Gas Data Using Random Forest Classifier Aided by Data Clustering Method. *IEEE Sensors Journal* **2022**, *22*, 5902–5910. <https://doi.org/10.1109/JSEN.2022.3149409>.
143. Wang, T.; Li, Q.; Yang, J.; Xie, T.; Wu, P.; Liang, J. Transformer Fault Diagnosis Method Based on Incomplete Data and TPE-XGBoost. *Applied Sciences* **2023**, *13*. <https://doi.org/10.3390/app13137539>.
144. Jiang, J.; Chen, R.; Chen, M.; Wang, W.; Zhang, C. Dynamic Fault Prediction of Power Transformers Based on Hidden Markov Model of Dissolved Gases Analysis. *IEEE Transactions on Power Delivery* **2019**, *34*, 1393–1400. <https://doi.org/10.1109/TPWRD.2019.2900543>.
145. Zhang, L.; Zhai, J. Fault diagnosis for oil-filled transformers using voting based extreme learning machine. *Cluster Computing* **2019**, *22*, 8363–8370. <https://doi.org/10.1007/s10586-018-1804-0>.
146. Han, X.; Ma, S.; Shi, Z.; An, G.; Du, Z.; Zhao, C. Transformer Fault Diagnosis Technology Based on Maximally Collapsing Metric Learning and Parameter Optimization Kernel Extreme Learning Machine. *IEEE Transactions on Electrical and Electronic Engineering* **2022**, *17*, 665–673. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/tee.23554](https://onlinelibrary.wiley.com/doi/pdf/10.1002/tee.23554), <https://doi.org/10.1002/tee.23554>.
147. Han, X.; Ma, S.; Shi, Z.; An, G.; Du, Z.; Zhao, C. A Novel Power Transformer Fault Diagnosis Model Based on Harris-Hawks-Optimization Algorithm Optimized Kernel Extreme Learning Machine. *Journal of Electrical Engineering & Technology* **2022**, *17*, 1993–2001. <https://doi.org/10.1007/s42835-022-01000-x>.
148. Han, X.; Huang, S.; Ma, S.; An, G.; An, Q.; Du, Z.; He, P. Fault diagnosis method for transformer based on NCA and CapSA-RELM. *Electrical Engineering* **2023**. <https://doi.org/10.1007/s00202-023-01974-5>.
149. Illias, H.A.; Chan, K.C.; Mokhlis, H. Hybrid feature selection–artificial intelligence–gravitational search algorithm technique for automated transformer fault determination based on dissolved gas analysis. *IET Generation, Transmission & Distribution* **2020**, *14*, 1575–1582. [_eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2019.1189](https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2019.1189), <https://doi.org/10.1049/iet-gtd.2019.1189>.
150. Shintemirov, A.; Tang, W.; Wu, Q.H. Power Transformer Fault Classification Based on Dissolved Gas Analysis by Implementing Bootstrap and Genetic Programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2009**, *39*, 69–79. <https://doi.org/10.1109/TSMCC.2008.2007253>.

151. Miranda, V.; Castro, A. Improving the IEC table for transformer failure diagnosis with knowledge extraction from neural networks. *IEEE Transactions on Power Delivery* **2005**, *20*, 2509–2516. <https://doi.org/10.1109/TPWRD.2005.855423>.
152. Duraisamy, V.; Devarajan, N.; Somasundareswari, D.; Vasanth, A.A.M.; Sivanandam, S. Neuro fuzzy schemes for fault detection in power transformer. *Applied Soft Computing* **2007**, *7*, 534–539. <https://doi.org/https://doi.org/10.1016/j.asoc.2006.10.001>.
153. Faiz, J.; Soleimani, M. Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis. *IEEE Transactions on Dielectrics and Electrical Insulation* **2018**, *25*, 1798–1806. <https://doi.org/10.1109/TDEI.2018.007191>.
154. Rao, U.M.; Fofana, I.; Rajesh, K.N.V.P.S.; Picher, P. Identification and Application of Machine Learning Algorithms for Transformer Dissolved Gas Analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* **2021**, *28*, 1828–1835. <https://doi.org/10.1109/TDEI.2021.009770>.
155. Senoussaoui, M.E.A.; Brahami, M.; Fofana, I. Combining and comparing various machine-learning algorithms to improve dissolved gas analysis interpretation. *IET Generation, Transmission & Distribution* **2018**, *12*, 3673–3679, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2018.0059>]. <https://doi.org/https://doi.org/10.1049/iet-gtd.2018.0059>.
156. Lu, W.; Shi, C.; Fu, H.; Xu, Y. Research on transformer fault diagnosis based on ISOMAP and IChOA-LSSVM. *IET Electric Power Applications* **2023**, *17*, 773–787.
157. Rajesh, K.N.V.P.S.; Rao, U.M.; Fofana, I.; Rozga, P.; Paramane, A. Influence of Data Balancing on Transformer DGA Fault Classification With Machine Learning Algorithms. *IEEE Transactions on Dielectrics and Electrical Insulation* **2023**, *30*, 385–392. <https://doi.org/10.1109/TDEI.2022.3230377>.
158. Rediansyah, D.; Prasajo, R.A.; Suwarno. Study on Artificial Intelligence Approaches for Power Transformer Health Index Assessment. In Proceedings of the 2021 International Conference on Electrical Engineering and Informatics (ICEEI), 2021, pp. 1–4. ISSN: 2155-6830, <https://doi.org/10.1109/ICEEI52609.2021.9611109>.
159. Zeinoddini-Meymand, H.; Vahidi, B. Health index calculation for power transformers using technical and economical parameters. *IET Science, Measurement & Technology* **2016**, *10*, 823–830, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-smt.2016.0184>]. <https://doi.org/https://doi.org/10.1049/iet-smt.2016.0184>.
160. Aizpurua, J.I.; McArthur, S.D.J.; Stewart, B.G.; Lambert, B.; Cross, J.G.; Catterson, V.M. Adaptive Power Transformer Lifetime Predictions Through Machine Learning and Uncertainty Modeling in Nuclear Power Plants. *IEEE Transactions on Industrial Electronics* **2019**, *66*, 4726–4737. <https://doi.org/10.1109/TIE.2018.2860532>.
161. Kunicki, M.; Borucki, S.; Cichoń, A.; Frymus, J. Modeling of the Winding Hot-Spot Temperature in Power Transformers: Case Study of the Low-Loaded Fleet. *Energies* **2019**, *12*. <https://doi.org/10.3390/en12183561>.
162. Valencia, F.; Arcos, H.; Quilumba, F. Comparison of Machine Learning Algorithms for the Prediction of Mechanical Stress in Three-Phase Power Transformer Winding Conductors. *Journal of Electrical and Computer Engineering* **2021**, *2021*, 4657696. Publisher: Hindawi, <https://doi.org/10.1155/2021/4657696>.
163. C. H. Wei, W.H.T.; Wu, Q.H. A Hybrid Least-square Support Vector Machine Approach to Incipient Fault Detection for Oil-immersed Power Transformer. *Electric Power Components and Systems* **2014**, *42*, 453–463, [<https://doi.org/10.1080/15325008.2013.857180>]. <https://doi.org/10.1080/15325008.2013.857180>.
164. Wani, S.A.; Rana, A.S.; Sohail, S.; Rahman, O.; Parveen, S.; Khan, S.A. Advances in DGA based condition monitoring of transformers: A review. *Renewable and Sustainable Energy Reviews* **2021**, *149*, 111347. <https://doi.org/https://doi.org/10.1016/j.rser.2021.111347>.
165. Kari, T.; Gao, W.; Tuluhong, A.; Yaermaimaiti, Y.; Zhang, Z. Mixed Kernel Function Support Vector Regression with Genetic Algorithm for Forecasting Dissolved Gas Content in Power Transformers. *Energies* **2018**, *11*. <https://doi.org/10.3390/en11092437>.
166. Kalaiselvi, T.; Sriramakrishnan, P.; Somasundaram, K. Survey of using GPU CUDA programming model in medical image analysis. *Informatics in Medicine Unlocked* **2017**, *9*, 133 – 144. <https://doi.org/https://doi.org/10.1016/j.imu.2017.08.001>.
167. Smistad, E.; Falch, T.L.; Bozorgi, M.; Elster, A.C.; Lindseth, F. Medical image segmentation on GPUs – A comprehensive review. *Medical Image Analysis* **2015**, *20*, 1 – 18. <https://doi.org/https://doi.org/10.1016/j.media.2014.10.012>.
168. Eklund, A.; Dufort, P.; Forsberg, D.; LaConte, S.M. Medical image processing on the GPU – Past, present and future. *Medical Image Analysis* **2013**, *17*, 1073 – 1094. <https://doi.org/https://doi.org/10.1016/j.media.2013.05.008>.

169. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, USA, 2012; NIPS'12, pp. 1097–1105.
170. Lopes, S.M.d.A.; Flauzino, R.A.; Altafim, R.A.C. Incipient fault diagnosis in power transformers by data-driven models with over-sampled dataset. *Electric Power Systems Research* **2021**, *201*, 107519. <https://doi.org/https://doi.org/10.1016/j.epsr.2021.107519>.
171. Mlakić, D.; Nikolovski, S.; Majdandžić, L. Deep learning method and infrared imaging as a tool for transformer faults detection. *Journal of Electrical Engineering* **2018**, *6*, 98–106.
172. Afrasiabi, S.; Afrasiabi, M.; Parang, B.; Mohammadi, M. Designing a composite deep learning based differential protection scheme of power transformers. *Applied Soft Computing* **2020**, *87*, 105975. <https://doi.org/https://doi.org/10.1016/j.asoc.2019.105975>.
173. Das, S.; Paramane, A.; Rao, U.M.; Chatterjee, S.; Kumar, K.S. Corrosive Dibenzyl Disulfide Concentration Prediction in Transformer Oil Using Deep Neural Network. *IEEE Transactions on Dielectrics and Electrical Insulation* **2023**, *30*, 1608–1615. <https://doi.org/10.1109/TDEI.2023.3289774>.
174. Li, Z.; He, Y.; Xing, Z.; Duan, J. Transformer fault diagnosis based on improved deep coupled dense convolutional neural network. *Electric Power Systems Research* **2022**, *209*, 107969. <https://doi.org/https://doi.org/10.1016/j.epsr.2022.107969>.
175. Zhai, X.; Tian, J.; Li, J. A Semi-Supervised Fault Diagnosis Method for Transformers Based on Discriminative Feature Enhancement and Adaptive Weight Adjustment. *IEEE Transactions on Instrumentation and Measurement* **2024**, *73*, 1–10. <https://doi.org/10.1109/TIM.2024.3398079>.
176. Zhang, Y.; Wang, Y.; Liu, J.; Zhang, H.; Fan, X.; Zhang, D. Improved multi-grained cascade forest model for transformer fault diagnosis. *CSEE Journal of Power and Energy Systems* **2022**, pp. 1–9. <https://doi.org/10.17775/CSEEJPES.2021.04670>.
177. Lei, L.; He, Y.; Xing, Z. Dissolved Gas Analysis for Power Transformer Fault Diagnosis Based on Deep Zero-shot Learning. *IEEE Transactions on Dielectrics and Electrical Insulation* **2024**, pp. 1–1. <https://doi.org/10.1109/TDEI.2024.3469913>.
178. Lin, J.; Ma, J.; Zhu, J. Hierarchical Federated Learning for Power Transformer Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*, 1–11. <https://doi.org/10.1109/TIM.2022.3196736>.
179. Zheng, W.; Zhang, G.; Zhao, C.; Zhu, Q. Multichannel consecutive data cross-extraction with 1DCNN-attention for diagnosis of power transformer. *International Journal of Electrical Power and Energy Systems* **2024**, *158*, 109951. <https://doi.org/https://doi.org/10.1016/j.ijepes.2024.109951>.
180. Li, K.; Li, J.; Huang, Q.; Chen, Y. Data augmentation for fault diagnosis of oil-immersed power transformer. *Energy Reports* **2023**, *9*, 1211–1219. <https://doi.org/https://doi.org/10.1016/j.egyr.2023.05.110>.
181. Hong, K.; Jin, M.; Huang, H. Transformer Winding Fault Diagnosis Using Vibration Image and Deep Learning. *IEEE Transactions on Power Delivery* **2021**, *36*, 676–685. <https://doi.org/10.1109/TPWRD.2020.2988820>.
182. Sun, Y.; Ma, S.; Sun, S.; Liu, P.; Zhang, L.; Ouyang, J.; Ni, X. Partial Discharge Pattern Recognition of Transformers Based on MobileNets Convolutional Neural Network. *Applied Sciences* **2021**, *11*. <https://doi.org/10.3390/app11156984>.
183. Xi, Y.; Yu, L.; Chen, B.; Chen, G.; Chen, Y. Research on Pattern Recognition Method of Transformer Partial Discharge Based on Artificial Neural Network. *Security and Communication Networks* **2022**, *2022*, 5154649. Publisher: Hindawi, <https://doi.org/10.1155/2022/5154649>.
184. Do, T.D.; Tuyet-Doan, V.N.; Cho, Y.S.; Sun, J.H.; Kim, Y.H. Convolutional-Neural-Network-Based Partial Discharge Diagnosis for Power Transformer Using UHF Sensor. *IEEE Access* **2020**, *8*, 207377–207388. <https://doi.org/10.1109/ACCESS.2020.3038386>.
185. Zhou, Y.; He, Y.; Xing, Z.; Wang, L.; Shao, K.; Lei, L.; Li, Z. Vibration Signal-Based Fusion Residual Attention Model for Power Transformer Fault Diagnosis. *IEEE Sensors Journal* **2024**, *24*, 17231–17242. <https://doi.org/10.1109/JSEN.2024.3382811>.
186. Xing, Z.; He, Y.; Wang, X.; Chen, J.; Du, B.; He, L.; Liu, X. Vibration-Signal-Based Deep Noisy Filtering Model for Online Transformer Diagnosis. *IEEE Transactions on Industrial Informatics* **2023**, *19*, 11239–11251. <https://doi.org/10.1109/TII.2023.3245193>.
187. Luo, Z.; Wang, C.; Qi, Z.; Luo, C. LA YOLOv8s: A lightweight-attention YOLOv8s for oil leakage detection in power transformers. *Alexandria Engineering Journal* **2024**, *92*, 82–91. <https://doi.org/https://doi.org/10.1016/j.aej.2024.02.054>.

188. Xing, Z.; He, Y. Multi-modal information analysis for fault diagnosis with time-series data from power transformer. *International Journal of Electrical Power and Energy Systems* **2023**, *144*, 108567. <https://doi.org/10.1016/j.ijepes.2022.108567>.
189. Jin, L.; Kim, D.; Chan, K.Y.; Abu-Siada, A. Deep Machine Learning-Based Asset Management Approach for Oil- Immersed Power Transformers Using Dissolved Gas Analysis. *IEEE Access* **2024**, *12*, 27794–27809. <https://doi.org/10.1109/ACCESS.2024.3366905>.
190. Xing, Z.; He, Y. Multimodal Mutual Neural Network for Health Assessment of Power Transformer. *IEEE Systems Journal* **2023**, *17*, 2664–2673. <https://doi.org/10.1109/JSYST.2023.3237225>.
191. Xing, Z.; He, Y.; Chen, J.; Wang, X.; Du, B. Health evaluation of power transformer using deep learning neural network. *Electric Power Systems Research* **2023**, *215*, 109016. <https://doi.org/10.1016/j.epsr.2022.109016>.
192. Zhong, M.; Cao, Y.; He, G.; Feng, L.; Tan, Z.; Mo, W.; Fan, J. Dissolved gas in transformer oil forecasting for transformer fault evaluation based on HATT-RLSTM. *Electric Power Systems Research* **2023**, *221*, 109431. <https://doi.org/10.1016/j.epsr.2023.109431>.
193. Luo, D.; Chen, W.; Fang, J.; Liu, J.; Yang, J.; Zhang, K. GRU-AGCN model for the content prediction of gases in power transformer oil. *Frontiers in Energy Research* **2023**, *11*. <https://doi.org/10.3389/fenrg.2023.1135330>.
194. He, L.; Li, L.; Li, M.; Li, Z.; Wang, X. A Deep Learning Approach to the Transformer Life Prediction Considering Diverse Aging Factors. *Frontiers in Energy Research* **2022**, *10*. <https://doi.org/10.3389/fenrg.2022.930093>.
195. Lin, W.; Miao, X.; Chen, J.; Xiao, S.; Lu, Y.; Jiang, H. Forecasting thermal parameters for ultra-high voltage transformers using long- and short-term time-series network with conditional mutual information. *IET Electric Power Applications* **2022**, *16*, 548–564. [_eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/elp2.12175, https://doi.org/10.1049/elp2.12175](https://doi.org/10.1049/elp2.12175).
196. Chen, Q.; Li, Z. A Transformer Vibration Amplitude Prediction Method Via Fusion Of Multi-Signals. In Proceedings of the 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2), 2021, pp. 3268–3273. <https://doi.org/10.1109/EI252483.2021.9713413>.
197. Das, S.; Paramane, A.; Chatterjee, S.; Rao, U.M. Sensing Incipient Faults in Power Transformers Using Bi-Directional Long Short-Term Memory Network. *IEEE Sensors Letters* **2023**, *7*, 1–4. <https://doi.org/10.1109/LSENS.2022.3233135>.
198. Wang, L.; Littler, T.; Liu, X. Dynamic Incipient Fault Forecasting for Power Transformers Using an LSTM Model. *IEEE Transactions on Dielectrics and Electrical Insulation* **2023**, *30*, 1353–1361. <https://doi.org/10.1109/TDEI.2023.3253463>.
199. Ma, X.; Hu, H.; Shang, Y. A New Method for Transformer Fault Prediction Based on Multifeature Enhancement and Refined Long Short-Term Memory. *IEEE Transactions on Instrumentation and Measurement* **2021**, *70*, 1–11. <https://doi.org/10.1109/TIM.2021.3098383>.
200. Zhong, M.; Yi, S.; Fan, J.; Zhang, Y.; He, G.; Cao, Y.; Feng, L.; Tan, Z.; Mo, W. Power transformer fault diagnosis based on a self-strengthening offline pre-training model. *Engineering Applications of Artificial Intelligence* **2023**, *126*, 107142. <https://doi.org/10.1016/j.engappai.2023.107142>.
201. Vidal, J.F.; Castro, A.R.G. Diagnosing Faults in Power Transformers With Variational Autoencoder, Genetic Programming, and Neural Network. *IEEE Access* **2023**, *11*, 30529–30545. <https://doi.org/10.1109/ACCESS.2023.3258544>.
202. Kim, S.; Jo, S.H.; Kim, W.; Park, J.; Jeong, J.; Han, Y.; Kim, D.; Youn, B.D. A Semi-Supervised Autoencoder With an Auxiliary Task (SAAT) for Power Transformer Fault Diagnosis Using Dissolved Gas Analysis. *IEEE Access* **2020**, *8*, 178295–178310. <https://doi.org/10.1109/ACCESS.2020.3027830>.
203. Xu, C.; Li, X.; Wang, Z.; Xie, J.; Yang, B.; Zhao, B. Fault Diagnosis of Power Transformer Based on Stacked Sparse Auto-Encoders and Broad Learning System. In Proceedings of the 2021 6th International Conference on Robotics and Automation Engineering (ICRAE), 2021, pp. 217–222. <https://doi.org/10.1109/ICRAE53653.2021.9657760>.
204. Seo, B.; Shin, J.; Kim, T.; Youn, B.D. Missing data imputation using an iterative denoising autoencoder (IDAE) for dissolved gas analysis. *Electric Power Systems Research* **2022**, *212*, 108642. <https://doi.org/10.1016/j.epsr.2022.108642>.
205. Luo, D.; Xi, R.; Che, L.; He, H. Health condition assessment of transformers based on cross message passing graph neural networks. *Frontiers in Energy Research* **2022**, *10*. <https://doi.org/10.3389/fenrg.2022.973736>.

206. Luo, D.; Fang, J.; He, H.; Lee, W.J.; Zhang, Z.; Zai, H.; Chen, W.; Zhang, K. Prediction for Dissolved Gas in Power Transformer Oil Based On TCN and GCN. *IEEE Transactions on Industry Applications* **2022**, *58*, 7818–7826. <https://doi.org/10.1109/TIA.2022.3197565>.
207. Dai, J.; Song, H.; Sheng, G.; Jiang, X. Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network. *IEEE Transactions on Dielectrics and Electrical Insulation* **2017**, *24*, 2828–2835. <https://doi.org/10.1109/TDEI.2017.006727>.
208. Zou, D.; Li, Z.; Quan, H.; Peng, Q.; Wang, S.; Hong, Z.; Dai, W.; Zhou, T.; Yin, J. Transformer fault classification for diagnosis based on DGA and deep belief network. *Energy Reports* **2023**, *9*, 250–256. <https://doi.org/https://doi.org/10.1016/j.egy.2023.09.183>.
209. Bai, X.; Zang, Y.; Li, J.; Song, Z.; Zhao, K. Transformer fault diagnosis method based on two-dimensional cloud model under the condition of defective data. *Electrical Engineering* **2023**. <https://doi.org/10.1007/s00202-023-01964-7>.
210. Bragone, F.; Morozovska, K.; Hilber, P.; Laneryd, T.; Luvisotto, M. Physics-informed neural networks for modelling power transformer's dynamic thermal behaviour. *Electric Power Systems Research* **2022**, *211*, 108447. <https://doi.org/https://doi.org/10.1016/j.epsr.2022.108447>.
211. Chen, M.; Herrera, F.; Hwang, K. Cognitive Computing: Architecture, Technologies and Intelligent Applications. *IEEE Access* **2018**, *6*, 19774–19783. <https://doi.org/10.1109/ACCESS.2018.2791469>.
212. M., S.; Murugappan, A.; T., M. Cognitive computing technological trends and future research directions in healthcare — A systematic literature review. *Artificial Intelligence in Medicine* **2023**, *138*, 102513. <https://doi.org/https://doi.org/10.1016/j.artmed.2023.102513>.
213. Xia, L.; Zheng, P.; Li, X.; Gao, R.; Wang, L. Toward cognitive predictive maintenance: A survey of graph-based approaches. *Journal of Manufacturing Systems* **2022**, *64*, 107–120. <https://doi.org/https://doi.org/10.1016/j.jmsy.2022.06.002>.
214. Zhang, Z.; Cui, P.; Zhu, W. Deep Learning on Graphs: A Survey. *CoRR* **2018**, *abs/1812.04202*, [1812.04202].
215. Mesgaran, M.; Hamza, A.B. A graph encoder–decoder network for unsupervised anomaly detection. *Neural Computing and Applications* **2023**, *35*, 23521–23535. <https://doi.org/10.1007/s00521-023-08964-5>.
216. Zhu, Y.; Lyu, F.; Hu, C.; Chen, X.; Liu, X. Encoder-Decoder Architecture for Supervised Dynamic Graph Learning: A Survey, 2022, [arXiv:cs.LG/2203.10480].
217. Bacciu, D.; Errica, F.; Micheli, A.; Podda, M. A gentle introduction to deep learning for graphs. *Neural Networks* **2020**, *129*, 203–221. <https://doi.org/https://doi.org/10.1016/j.neunet.2020.06.006>.
218. Khemani, B.; Patil, S.; Kotecha, K.; Tanwar, S. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data* **2024**, *11*, 18. <https://doi.org/10.1186/s40537-023-00876-4>.
219. Chen, Z.; Xu, J.; Alippi, C.; Ding, S.X.; Shardt, Y.; Peng, T.; Yang, C. Graph neural network-based fault diagnosis: a review, 2021. <https://doi.org/10.48550/ARXIV.2111.08185>.
220. Li, T.; Zhou, Z.; Li, S.; Sun, C.; Yan, R.; Chen, X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mechanical Systems and Signal Processing* **2022**, *168*, 108653. <https://doi.org/https://doi.org/10.1016/j.ymsp.2021.108653>.
221. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* **2021**, *32*, 4–24. <https://doi.org/10.1109/tnnls.2020.2978386>.
222. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI Open* **2020**, *1*, 57–81. <https://doi.org/https://doi.org/10.1016/j.aiopen.2021.01.001>.
223. Chen, M.; Wei, Z.; Huang, Z.; Ding, B.; Li, Y. Simple and Deep Graph Convolutional Networks. In Proceedings of the Proceedings of the 37th International Conference on Machine Learning; III, H.D.; Singh, A., Eds. PMLR, 13–18 Jul 2020, Vol. 119, *Proceedings of Machine Learning Research*, pp. 1725–1735.
224. Ruiz, L.; Gama, F.; Ribeiro, A. Gated Graph Recurrent Neural Networks. *IEEE Transactions on Signal Processing* **2020**, *68*, 6303–6318. <https://doi.org/10.1109/TSP.2020.3033962>.
225. Rennard, V.; Nikolentzos, G.; Vazirgiannis, M. Graph Auto-Encoders for Learning Edge Representations. In Proceedings of the Complex Networks & Their Applications IX; Benito, R.M.; Cherifi, C.; Cherifi, H.; Moro, E.; Rocha, L.M.; Sales-Pardo, M., Eds., Cham, 2021; pp. 117–129.
226. Pfeiffer, J.; Ruder, S.; Vulić, I.; Ponti, E.M. Modular Deep Learning, 2023, [arXiv:cs.LG/2302.11529].
227. Rahaman, N.; Weiss, M.; Träuble, F.; Locatello, F.; Lacoste, A.; Bengio, Y.; Pal, C.; Li, L.E.; Schölkopf, B. A General Purpose Neural Architecture for Geospatial Systems, 2022, [arXiv:cs.LG/2211.02348].

228. Jaegle, A.; Borgeaud, S.; Alayrac, J.; Doersch, C.; Ionescu, C.; Ding, D.; Koppula, S.; Zoran, D.; Brock, A.; Shelhamer, E.; et al. Perceiver IO: A General Architecture for Structured Inputs & Outputs. *CoRR* **2021**, *abs/2107.14795*, [2107.14795].
229. Rahaman, N.; Weiss, M.; Locatello, F.; Pal, C.; Bengio, Y.; Schölkopf, B.; Li, L.E.; Ballas, N. Neural Attentive Circuits, 2022, [arXiv:cs.LG/2210.08031].
230. Ding, Y.; Zhuang, J.; Ding, P.; Jia, M. Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings. *Reliability Engineering and System Safety* **2022**, *218*, 108126. <https://doi.org/https://doi.org/10.1016/j.res.2021.108126>.
231. Li, J.; Huang, R.; Chen, J.; Xia, J.; Chen, Z.; Li, W. Deep Self-Supervised Domain Adaptation Network for Fault Diagnosis of Rotating Machine With Unlabeled Data. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*, 1–9. <https://doi.org/10.1109/TIM.2022.3164136>.
232. Li, G.; Wu, J.; Deng, C.; Wei, M.; Xu, X. Self-supervised learning for intelligent fault diagnosis of rotating machinery with limited labeled data. *Applied Acoustics* **2022**, *191*, 108663. <https://doi.org/https://doi.org/10.1016/j.apacoust.2022.108663>.
233. Mao, W.; Chen, J.; Liu, J.; Liang, X. Self-Supervised Deep Domain-Adversarial Regression Adaptation for Online Remaining Useful Life Prediction of Rolling Bearing Under Unknown Working Condition. *IEEE Transactions on Industrial Informatics* **2023**, *19*, 1227–1237. <https://doi.org/10.1109/TII.2022.3172704>.
234. Qiao, Y.; Lü, J.; Wang, T.; Liu, K.; Zhang, B.; Snoussi, H. A Multihead Attention Self-Supervised Representation Model for Industrial Sensors Anomaly Detection. *IEEE Transactions on Industrial Informatics* **2024**, *20*, 2190–2199. <https://doi.org/10.1109/TII.2023.3280337>.
235. Mao, W.; Liu, K.; Zhang, Y.; Liang, X.; Wang, Z. Self-Supervised Deep Tensor Domain-Adversarial Regression Adaptation for Online Remaining Useful Life Prediction Across Machines. *IEEE Transactions on Instrumentation and Measurement* **2023**, *72*, 1–16. <https://doi.org/10.1109/TIM.2023.3265109>.
236. Xu, Y.; Wang, H.; Liu, Z.; Zuo, M. Self-Supervised Defect Representation Learning for Label-Limited Rail Surface Defect Detection. *IEEE Sensors Journal* **2023**, *23*, 29235–29246. <https://doi.org/10.1109/JSEN.2023.324668>.
237. Guan, Y.; Meng, Z.; Sun, D.; Liu, J.; Fan, F. 2MNet: Multi-sensor and multi-scale model toward accurate fault diagnosis of rolling bearing. *Reliability Engineering and System Safety* **2021**, *216*, 108017. <https://doi.org/https://doi.org/10.1016/j.res.2021.108017>.
238. Feng, J.; Su, J.; Feng, X. A Residual Multihead Self-Attention Network Using Multimodal Shallow Feature Fusion for Motor Fault Diagnosis. *IEEE Sensors Journal* **2023**, *23*, 29131–29142. <https://doi.org/10.1109/JSEN.2023.3322151>.
239. Zhang, X.; Feng, Y.; Chen, J.; Liu, Z.; Wang, J.; Huang, H. Knowledge distillation-optimized two-stage anomaly detection for liquid rocket engine with missing multimodal data. *Reliability Engineering and System Safety* **2024**, *241*, 109676. <https://doi.org/https://doi.org/10.1016/j.res.2023.109676>.
240. Sun, D.; Li, Y.; Liu, Z.; Jia, S.; Noman, K. Physics-inspired multimodal machine learning for adaptive correlation fusion based rotating machinery fault diagnosis. *Information Fusion* **2024**, *108*, 102394. <https://doi.org/https://doi.org/10.1016/j.inffus.2024.102394>.
241. Huang, Y.; Tao, J.; Sun, G.; Wu, T.; Yu, L.; Zhao, X. A novel digital twin approach based on deep multimodal information fusion for aero-engine fault diagnosis. *Energy* **2023**, *270*, 126894. <https://doi.org/https://doi.org/10.1016/j.energy.2023.126894>.
242. Kounta, C.A.K.A.; Kamsu-Foguem, B.; Nouredine, F.; Tangara, F. Multimodal deep learning for predicting the choice of cut parameters in the milling process. *Intelligent Systems with Applications* **2022**, *16*, 200112. <https://doi.org/https://doi.org/10.1016/j.iswa.2022.200112>.
243. Zhao, Y.; Zhang, Y.; Li, Z.; Bu, L.; Han, S. AI-enabled and multimodal data driven smart health monitoring of wind power systems: A case study. *Advanced Engineering Informatics* **2023**, *56*, 102018. <https://doi.org/https://doi.org/10.1016/j.aei.2023.102018>.
244. Li, X.; Zhong, X.; Shao, H.; Han, T.; Shen, C. Multi-sensor gearbox fault diagnosis by using feature-fusion covariance matrix and multi-Riemannian kernel ridge regression. *Reliability Engineering and System Safety* **2021**, *216*, 108018. <https://doi.org/https://doi.org/10.1016/j.res.2021.108018>.
245. Long, Z.; Zhang, X.; Zhang, L.; Qin, G.; Huang, S.; Song, D.; Shao, H.; Wu, G. Motor fault diagnosis using attention mechanism and improved adaboost driven by multi-sensor information. *Measurement* **2021**, *170*, 108718. <https://doi.org/https://doi.org/10.1016/j.measurement.2020.108718>.

246. Xu, X.; Bao, S.; Shao, H.; Shi, P. A multi-sensor fused incremental broad learning with D-S theory for online fault diagnosis of rotating machinery. *Advanced Engineering Informatics* **2024**, *60*, 102419. <https://doi.org/https://doi.org/10.1016/j.aei.2024.102419>.
247. Ye, M.; Yan, X.; Jiang, D.; Xiang, L.; Chen, N. MIFDELN: A multi-sensor information fusion deep ensemble learning network for diagnosing bearing faults in noisy scenarios. *Knowledge-Based Systems* **2024**, *284*, 111294. <https://doi.org/https://doi.org/10.1016/j.knosys.2023.111294>.
248. Du, H.; Wang, Q.; Zhang, X.; Qian, W.; Wang, J. A novel multi-sensor hybrid fusion framework. *Measurement Science and Technology* **2024**, *35*, 086105. <https://doi.org/10.1088/1361-6501/ad42c4>.
249. Li, Y.; Luo, X.; Xie, Y.; Zhao, W. Multi-head spatio-temporal attention based parallel GRU architecture: a novel multi-sensor fusion method for mechanical fault diagnosis. *Measurement Science and Technology* **2023**, *35*, 015111. <https://doi.org/10.1088/1361-6501/acfe29>.
250. Ma, T.; Shen, J.; Song, D.; Xu, F. A vibro-acoustic signals hybrid fusion model for blade crack detection. *Mechanical Systems and Signal Processing* **2023**, *204*, 110815. <https://doi.org/https://doi.org/10.1016/j.ymsp.2023.110815>.
251. Ma, T.; Shen, J.; Song, D.; Xu, F. Multi-sensor and multi-level information fusion model for compressor blade crack detection. *Measurement* **2023**, *222*, 113622. <https://doi.org/https://doi.org/10.1016/j.measurement.2023.113622>.
252. Liang, J.; Mao, Z.; Liu, F.; Kong, X.; Zhang, J.; Jiang, Z. Multi-sensor signals multi-scale fusion method for fault detection of high-speed and high-power diesel engine under variable operating conditions. *Engineering Applications of Artificial Intelligence* **2023**, *126*, 106912. <https://doi.org/https://doi.org/10.1016/j.engappai.2023.106912>.
253. Tong, J.; Liu, C.; Zheng, J.; Pan, H. Multi-sensor information fusion and coordinate attention-based fault diagnosis method and its interpretability research. *Engineering Applications of Artificial Intelligence* **2023**, *124*, 106614. <https://doi.org/https://doi.org/10.1016/j.engappai.2023.106614>.
254. Zhang, Y.; Ji, J.; Ren, Z.; Ni, Q.; Wen, B. Multi-sensor open-set cross-domain intelligent diagnostics for rotating machinery under variable operating conditions. *Mechanical Systems and Signal Processing* **2023**, *191*, 110172. <https://doi.org/https://doi.org/10.1016/j.ymsp.2023.110172>.
255. Guo, J.; He, Q.; Zhen, D.; Gu, F.; Ball, A.D. Multi-sensor data fusion for rotating machinery fault detection using improved cyclic spectral covariance matrix and motor current signal analysis. *Reliability Engineering and System Safety* **2023**, *230*, 108969. <https://doi.org/https://doi.org/10.1016/j.res.2022.108969>.
256. Zhu, J.; Wang, Y.; Xia, M.; Williams, D.; de Silva, C.W. A New Multisensor Partial Domain Adaptation Method for Machinery Fault Diagnosis Under Different Working Conditions. *IEEE Transactions on Instrumentation and Measurement* **2023**, *72*, 1–10. <https://doi.org/10.1109/TIM.2023.3318679>.
257. Liu, J.; Xie, F.; Zhang, Q.; Lyu, Q.; Wang, X.; Wu, S. A multisensory time-frequency features fusion method for rotating machinery fault diagnosis under nonstationary case. *Journal of Intelligent Manufacturing* **2023**. <https://doi.org/10.1007/s10845-023-02198-x>.
258. Meng, Z.; Zhu, J.; Cao, S.; Li, P.; Xu, C. Bearing Fault Diagnosis Under Multisensor Fusion Based on Modal Analysis and Graph Attention Network. *IEEE Transactions on Instrumentation and Measurement* **2023**, *72*, 1–10. <https://doi.org/10.1109/TIM.2023.3301895>.
259. Wan, S.; Li, T.; Fang, B.; Yan, K.; Hong, J.; Li, X. Bearing Fault Diagnosis Based on Multisensor Information Coupling and Attentional Feature Fusion. *IEEE Transactions on Instrumentation and Measurement* **2023**, *72*, 1–12. <https://doi.org/10.1109/TIM.2023.3269115>.
260. Zhang, Y.; Feng, K.; Ma, H.; Yu, K.; Ren, Z.; Liu, Z. MMFNet: Multisensor Data and Multiscale Feature Fusion Model for Intelligent Cross-Domain Machinery Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*, 1–11. <https://doi.org/10.1109/TIM.2022.3213016>.
261. Guo, J.; He, Q.; Zhen, D.; Gu, F.; Ball, A.D. Multiscale cyclic frequency demodulation-based feature fusion framework for multi-sensor driven gearbox intelligent fault detection. *Knowledge-Based Systems* **2024**, *283*, 111203. <https://doi.org/https://doi.org/10.1016/j.knosys.2023.111203>.
262. Li, Y.F.; Wang, H.; Sun, M. ChatGPT-like large-scale foundation models for prognostics and health management: A survey and roadmaps. *Reliability Engineering and System Safety* **2024**, *243*, 109850. <https://doi.org/https://doi.org/10.1016/j.res.2023.109850>.
263. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.B.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the Opportunities and Risks of Foundation Models. *CoRR* **2021**, *abs/2108.07258*, [2108.07258].

264. Zhou, C.; Li, Q.; Li, C.; Yu, J.; Liu, Y.; Wang, G.; Zhang, K.; Ji, C.; Yan, Q.; He, L.; et al. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT, 2023, [arXiv:cs.AI/2302.09419].
265. Sun, H. Modularity in deep learning. Theses, Université Paris-Saclay, 2023.
266. Goyal, A.; Bengio, Y. Inductive Biases for Deep Learning of Higher-Level Cognition. *CoRR* **2020**, *abs/2011.15091*, [2011.15091].
267. Chitty-Venkata, K.T.; Emani, M.; Vishwanath, V.; Somani, A.K. Neural Architecture Search for Transformers: A Survey. *IEEE Access* **2022**, *10*, 108374–108412. <https://doi.org/10.1109/ACCESS.2022.3212767>.
268. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2023**, *45*, 87–110. <https://doi.org/10.1109/TPAMI.2022.3152247>.
269. Fei, N.; Lu, Z.; Gao, Y.; Yang, G.; Huo, Y.; Wen, J.; Lu, H.; Song, R.; Gao, X.; Xiang, T.; et al. Towards artificial general intelligence via a multimodal foundation model. *Nature Communications* **2022**, *13*, 3094. <https://doi.org/10.1038/s41467-022-30761-2>.
270. Swamy, V.; Satayeva, M.; Frej, J.; Bossy, T.; Vogels, T.; Jaggi, M.; Käser, T.; Hartley, M.A. Multi-ModN—Multimodal, Multi-Task, Interpretable Modular Networks. In Proceedings of the Advances in Neural Information Processing Systems; Oh, A.; Neumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; Levine, S., Eds. Curran Associates, Inc., 2023, Vol. 36, pp. 28115–28138.
271. Siddharth, N.; Paige, B.; van de Meent, J.W.; Desmaison, A.; Goodman, N.D.; Kohli, P.; Wood, F.; Torr, P.H.S. Learning Disentangled Representations with Semi-Supervised Deep Generative Models, 2017, [arXiv:stat.ML/1706.00400].
272. Hristov, Y.; Angelov, D.; Burke, M.; Lascarides, A.; Ramamoorthy, S. Disentangled Relational Representations for Explaining and Learning from Demonstration. *CoRR* **2019**, *abs/1907.13627*, [1907.13627].
273. Berrevoets, J.; Kacprzyk, K.; Qian, Z.; van der Schaar, M. Causal Deep Learning, 2023, [arXiv:cs.LG/2303.02186].
274. Pfeiffer, J.; Ruder, S.; Vulić, I.; Ponti, E.M. Modular Deep Learning, 2023, [arXiv:cs.LG/2302.11529].
275. Tran, D.; Liu, J.; Dusenberry, M.W.; Phan, D.; Collier, M.; Ren, J.; Han, K.; Wang, Z.; Mariet, Z.; Hu, H.; et al. Plex: Towards Reliability using Pretrained Large Model Extensions, 2022, [arXiv:cs.LG/2207.07411].
276. Akhtar, S.; Adeel, M.; Iqbal, M.; Namoun, A.; Tufail, A.; Kim, K.H. Deep learning methods utilization in electric power systems. *Energy Reports* **2023**, *10*, 2138–2151. <https://doi.org/https://doi.org/10.1016/j.egy.2023.09.028>.
277. Heymann, F.; Quest, H.; Lopez Garcia, T.; Ballif, C.; Galus, M. Reviewing 40 years of artificial intelligence applied to power systems – A taxonomic perspective. *Energy and AI* **2024**, *15*, 100322. <https://doi.org/https://doi.org/10.1016/j.egyai.2023.100322>.
278. Longo, L.; Brcic, M.; Cabitza, F.; Choi, J.; Confalonieri, R.; Ser, J.D.; Guidotti, R.; Hayashi, Y.; Herrera, F.; Holzinger, A.; et al. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion* **2024**, *106*, 102301. <https://doi.org/https://doi.org/10.1016/j.inffus.2024.102301>.
279. Machlev, R.; Heistrene, L.; Perl, M.; Levy, K.Y.; Belikov, J.; Mannor, S.; Levron, Y. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI* **2022**, *9*, 100169. <https://doi.org/https://doi.org/10.1016/j.egyai.2022.100169>.
280. Shukla, V.; Sant, A.; Sharma, P.; Nayak, M.; Khatri, H. An explainable artificial intelligence based approach for the prediction of key performance indicators for 1 megawatt solar plant under local steppe climate conditions. *Engineering Applications of Artificial Intelligence* **2024**, *131*, 107809. <https://doi.org/https://doi.org/10.1016/j.engappai.2023.107809>.
281. Titz, M.; Pütz, S.; Witthaut, D. Identifying drivers and mitigators for congestion and redispatch in the German electric power system with explainable AI. *Applied Energy* **2024**, *356*, 122351. <https://doi.org/https://doi.org/10.1016/j.apenergy.2023.122351>.

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