Dissolved Gas Analysis Principle-Based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil-Immersed Power Transformers: A Survey

Lefeng Cheng [1,2,*] and Tao Yu [1,2,*]

[1] School of Electric Power, South China University of Technology, Guangzhou 510640, China;

[2] Guangdong Key Laboratory of Clean Energy Technology, Guangzhou 510640, China

* Correspondence: chenglefeng_scut@163.com; taoyu1@suct.edu.cn; Tel.: +86-136-8223-6454，
   +86-130-0208-8518

Abstract: Compared with conventional methods of fault diagnosis for power transformers, which have defects such as imperfect encoding and too absolute encoding boundaries, this paper systematically discusses various intelligent approaches applied in fault diagnosis and decision making for large oil-immersed power transformers based on dissolved gas analysis (DGA), including expert system (EPS), artificial neural network (ANN), fuzzy theory, rough sets theory (RST), grey system theory (GST), swarm intelligence (SI) algorithms, data mining technology, machine learning (ML), and other intelligent diagnosis tools, and summarizes existing problems and solutions. From this survey, it is found that a single intelligent approach for fault diagnosis can only reflect operation status of the transformer in one particular aspect, causing various degrees of shortcomings that cannot be resolved effectively. Combined with the current research status in this field, the problems that must be addressed in DGA-based transformer fault diagnosis are identified, and the prospects for future development trends and research directions are outlined. This contribution presents a detailed and systematic survey on various intelligent approaches to faults diagnosing and decisions making of the power transformer, in which their merits and demerits are thoroughly investigated, as well as their improvement schemes and future development trends are proposed. Moreover, this paper concludes that a variety of intelligent algorithms should be combined for mutual complementation to form a hybrid fault diagnosis network, such that avoiding these algorithms falling into a local optimum. Moreover, it is necessary to improve the detection instruments so as to acquire reasonable characteristic gas data samples. The research summary, empirical generalization and analysis of predicament in this paper provide some thoughts and suggestions for the research of complex power grid in the new environment, as well as references and guidance for researchers to choose optimal approach to achieve DGA-based fault diagnosis and decision of the large oil-immersed power transformers in preventive electrical tests.

Keywords: power transformer; fault diagnosis and decision; dissolved gas analysis; intelligent algorithms; reliability assessment; hybrid network; preventive electrical tests

1. Introduction

Power transformers are one of the most crucial pieces of equipment in a power system, thus their safe and stable operation plays a significant role in the safe, stable and reliable operation of the whole power system [1]. During the operation of power transformers, various faults may happen

due to destruction of or inappropriate installation and other reasons [2]. These faults can seriously affect the normal operation of the transformer. Hence, in depth discussion of the different fault diagnosis methods of power transformers is a valuable research topic. As large power equipment, power transformers in general have a very long lifespan for the time they go into operation until their final decommissioning (the reference life given by the Southern China Power Grid Jiangmen Bureau is 20 years), thus they have many different requirements and differences in their overhauling process. In the whole life procedure of the transformer, it is rare to conduct hood adjustment and overhaul involving disassembly, which means that we have little chance to directly examine the internal insulation, especially the winding oil-immersed insulation. Hence, the internal conditions of the transformer can only be evaluated through a variety of preventive tests. In other words, we must assess the insulation ageing in transformers in some indirect way.

Generally speaking, various preventive tests can accurately reflect the performance and state of all aspects and parts of the power transformer to a certain extent. In these tests, the parameters that can really reflect the ageing failure of the transformer are often used to correct the original ageing assessment model in order to maximize the reliability evaluation value close to the real value and reduce the accumulation error with the time to decommissioning [3]. In China, preventive tests have been an important part of electric power production practice for a long time, and has which played a positive role in the safe operation of the power equipment [4]. Also in China, the Southern China Power Grid Corporation has issued an enterprise standard named Preventive Test Procedures for Electric Power Equipment, in which the prescribed preventive tests of insulation items as presented in Table 1 are given.

Table 1. Prescribed preventive test of insulation items.

No.	Test Items	No.	Test Items
1	Chromatogram analysis of dissolved gas in oil	17	Partial discharge measurement
2	DC resistance of winding	18	No-load closing under full voltage
3	Insulation resistance, absorption ration or (and) polarization index of winding	19	Temperature measuring device and its secondary circuit test
4	Tangent value of dielectric loss angle of winding	20	Gas relay and its secondary circuit test
5	Tangent value of condenser bushing tgδ and capacitance value	21	Checking and test of cooling device and its secondary circuit
6	Insulation oil checking test	22	Overall sealing inspection
7	High-voltage endurance test	23	Pressure releaser checking
8	Insulation resistance of iron core (with external grounding wire)	24	Insulation test of current transformer in casing
9	Insulation resistance of through bolts, iron yoke clamps, steel banding, iron core winding pressure ring and shielding	25	Degree of polymerization of insulated cardboard
10	Water content in oil	26	Content of furfural in oil
11	Gas content in oil	27	Test and check of OLTC device [1]
12	Leakage current of winding	28	Water content of insulated cardboard
13	Voltage ratio of all taps in windings	29	Impedance measurement
14	Checking of the group of three-phase transformer and the polarity of the single-phase transformer	30	Surface temperature measurement of oil tank
15	No-load current and no-load loss	31	Noise measurement
16	Off-impedance and load loss	32	Vibration measurement

[1] OLTC: On-Load Tap Changing.

As shown in Table 1, among the preventive test items, some are conducted after disintegration of the transformer, some are carried out in conjunction with or incidental to other items, some are routine checks and test items before or after the operation of the transformer, and some are implemented only in special circumstances. In these testing items, the chromatographic analysis of dissolved gas in oil, namely dissolved gas analysis (DGA) is an important means of transformer internal fault diagnosis. It provides an important basis for indirect discovering hidden faults in transformers. It is also proved by practice that the dissolved gas analysis of transformer oil technique is very effective to find latent faults in transformers as well as their development trends.

Hence, both in China and around the world, DGA technology is believed as an important approach for preventive test of power equipment. For a normal oil-immersed power transformer, the content limits of hydrogen-containing gases and hydrocarbon gases in transformer oil are as follows: the normal limits [3] of H_2, CH_4, C_2H_6, C_2H_4, C_2H_2 and total hydrocarbons are 150, 45, 35, 65, 5 and 150 ppm, respectively.

DGA is also a most important reference index in the model correction [4]. Here, the model correction is aimed at large oil-immersed power transformers, which all adopt oil-paper insulation structures, thus the electrical parts of the whole body are completely immersed in the transformer oil. By employing the DGA technique, the information of the dissolved gas in transformer oil such as their components and contents can be qualitatively and quantitatively analysed to find out the cause of gas production, so as to analyse and diagnose whether the internal state of the transformer during operation is normal, and finally find any potential faults inside the transformer in time. The DGA-based preventive test is a comprehensive test method involving transformer discharging and thermal issues, thus it has a larger monitoring scope than the partial discharge measurements under an induced voltage. Besides, it is easily realized online. Hence the DGA-based fault diagnosis and decision making is a significant approach in current insulation monitoring measures [5–9]. As previously stated, the enterprise standard developed by Southern China Power Grid Corporation named Q/CSG114002-2011 lists-the DGA based fault diagnosis test as the first test item for the oil-immersed power transformers. The relevant regulations in this enterprise standard and the standard DL/T722-2000 [10] named Guidelines for Analysis and Judgment of Dissolved Gases in Transformer Oil both demonstrate that there is a significant relationship between the type of transformer fault and the dissolved gas components in the transformer oil. For the three major transformer fault types, including overheating faults, electrical faults and partial discharges, the corresponding dissolved gas composition in the transformer oil may be briefly described as follows:

For overheating faults, under the thermal and electrical effects, the transformer oil and organic insulating materials will gradually age and decompose, which produces a small amount of low molecular weight hydrocarbons and other gases, such as CO_2 and CO. Here, when the thermal stress only affects the decomposition of transformer oil at the source of heat not involving the solid insulation, the gases produced are mainly low molecular weight hydrocarbon gases, among which the characteristic gases are generally CH_4 and C_2H_4, and the sum of the two generally accounts for more than 80% of the total hydrocarbons. In this situation, acetylene is usually not generated due to overheating failures. Generally, the content of C_2H_2 will not exceed 2% of the total hydrocarbon when the overheating is below 500 °C; severe overheating (above 800 °C) also produces a small amount of C_2H_2, but the maximum content is not more than 6% of the total hydrocarbons; when it comes to the overheating faults of solid insulation, apart from the above low molecular weight hydrocarbon gases, more CO_2 and CO are also produced. Moreover, with the increase of temperature, the content of CO_2 and CO will increase gradually. For the overheating faulted which are limited to only partial oil blockages or poor heat dissipation, owing to the fact the overheating temperature is lower and the overheating area is larger, the pyrolysis effect of transformer oil is not obvious at this time, thus the content of low molecular weight hydrocarbon gases is not necessarily high.

Electrical faults refer to the deterioration of insulation caused by high electrical stress. Depending on the different energy densities, this type of fault can be divided into different types of fault, such as high energy-density discharges and low energy-density discharges (i.e., partial discharges and spark discharges). When an electric arc discharge occurs, the major characteristic gases produced of this type of fault are C_2H_2 and H_2, and then a large amount of C_2H_4 and CH_4. As the development of the arc discharge fault occurs rapidly, the gases are usually too late to be dissolved in transformer oil and then gather in the gas relay. Therefore, under this situation, the component and content of dissolved gases in oil are often highly related to the location of fault, the speed of oil flow and the duration of the fault. Under such a failure, C_2H_2 generally accounts for 20 to 70%, and H_2 accounts for 30 to 90% of the total hydrocarbons. In most cases, the content of C_2H_2 is higher than CH_4. When it involves the solid insulation, the content of gases in the gas relay and

the gas CO in oil are higher. In spark discharge faults, the major characteristic gases are C_2H_2 and H_2. In general, the total hydrocarbon content in this type of fault is not high due to the low fault energy. However, at this point, the proportion of C_2H_2 dissolved in oil in the total hydrocarbon can reach 25 to 90%, C_2H_4 content is less than 20% of the total hydrocarbons, and H_2 accounts for more than 30% of the total hydrocarbon.

As for partial discharge faults, they are a local and repetitive breakdown phenomenon occurring in the gas gap (or bubble) and the sharp points in the oil-paper insulating structure due to the weakness of insulation and the concentration of electric field. When a partial discharge occurs, the characteristic gas component content is different due to the difference of discharge energy density. Under normal circumstances, the total hydrocarbon content is not high, and the main component is H_2, which usually accounts for more than 90% of the total amount of gases; and the next is CH_4, which accounts for more than 90% of the total hydrocarbons. When the energy density of the discharge increases, the gas C_2H_2 will also be produced, but its proportion in the total hydrocarbon is generally no more than 2%.

Hence, on the whole, the gas components produced by different types of transformer faults are different according to the China standard DL/T722-2000 [10], as shown in Table 2. In Table 2, we find that the main gas components produced by different categories of transformer faults are also different.

Table 2. The characteristic gases produced in different types of transformer faults.

Fault Type	Main Gas Component	Minor Gas Component
Oil in overheating	CH_4, C_2H_2	H_2, C_2H_6
Oil and paper both in overheating	CH_4, C_2H_4, CO, CO_2,	H_2, C_2H_6
PD [1] in oil-paper insulation	H_2, CH_4, CO	C_2H_2, C_2H_6, CO_2
Spark discharge in oil	H_2, C_2H_2	/
Electric arc in oil	H_2, C_2H_2	CH_4, C_2H_4, C_2H_6
Electric arc both in oil and paper	H_2, C_2H_2, CO, CO_2	CH_4, C_2H_4, C_2H_6

[1] PD = partial discharge.

The DGA technicians both at home and abroad have conducted a lot of research work on how to determine the quantitative relationship between the content of these characteristic gases and the internal faults of power transformers. The China standard DL/T722-2000 [10] gives a recommended limit value of the gas content in the transformer oil, and it also gives the warning value of the absolute gas production rate of the transformer, as shown in Table 3. Therefore, the gas production rate can more accurately reflect the true state of the transformer than the characteristic gas content. However, in specific operation, if the test cycle of chromatographic analysis is longer, the rate of gas production will be inaccurate.

Table 3. The warning value of the content of dissolved gas in transformer oil (μL/L).

Components of Dissolved Gas in Transformer Oil	Content		Open Type	Diaphragm Type
	Above 330 kV	Below 220 kV		
Total hydrocarbon C_xH_y	150	150	6	12
C_2H_2	1	5	0.1	0.2
H_2	150	150	5	10
CO	/	/	50	100
CO_2	/	/	100	200

Given all that, the best method of DGA diagnosis is to combine the characteristic gas content with the gas production rate. For the content of characteristic gases, CH_4, C_2H_4, C_2H_6, C_2H_2 and H_2 are usually selected as five indexes in the characteristic gas. Typically, C_2H_2 is not generated in the normal transformer oil, thus in the chromatographic analysis, it should be only be paid attention once characteristic C_2H_2 gas appears. When corona discharge, water electrolysis or rust, serious overloads, high temperature overheating, and spark discharges and other failures occur in a transformer, it will generate H_2. Hence, H_2 is also a very important characteristic gas. At the same

time, according to the available data, the normal deterioration of solid insulation materials and the deterioration decomposition in the case of failure are manifested in the content of CO and CO_2. However, there is no unified method to determine the normal limit content of these characteristic gases in China. Therefore, considering test availability, CO and CO_2 are usually not considered.

According to the corresponding relationship between the fault of the transformer and the dissolved gas in the oil described above, the researchers at home and abroad have put forward many traditional approaches to judge the transformer faults via gas chromatography, in which the oil samples are extracted from the transformers in operation for further fractionation and analysis of dissolved gas in the oil. According to the test results, the operation status and fault types of the transformer can be judged and achieved. This gas chromatography methods for fault judgment are generally divided into three categories as follows:

The first one is the characteristic gas method [11–13], which is employed to analyze the content value of each component of the gas dissolved in transformer oil, as well as the total alkyne content and gas production rate. The gases produced inside the transformer have different characteristics in different types of faults. Hence, according to the gas chromatography of insulation oil test results, the features of gas production, and the warning values of characteristic gases, a preliminary and rough judgment on whether there is a failure and the failure properties can be achieved. Here, the characteristic gases include total hydrocarbon, hydrogen, methane, ethane, ethylene, acetylene, etc.

The second one is gas production rate method [14–17]. When the content of gas inside some transformers exceeds the warning value, we cannot judge whether there a failure has occurred in these transformers, while inside some other transformers, the content of gas is lower than the warning value but with a rapid increasing speed, attention should be paid at this point. Hence, the gas-production rate of the fault point can further reflect the existence, severity and development trends of the failures, which can be divided into absolute gas production rate and relative gas production rate. The former one should be used to judge the fault of the transformer.

The last one is the three-ratio method, which is used to encode and classify the relative content of dissolved gases in transformer oil [18–22]. In this approach, five types of characteristic gases, including hydrogen, methane, ethane, ethylene and acetylene, are used to form three pairs of different ratios. For different ratio ranges, such three pairs of ratios are expressed by different codes for combinatorial analysis, so that the faults of the transformer can be judged via classifying the faults according to severity. In other words, we first judge the possible faults according to the attention value of content of each component or the attention value of gas production rate, and then use the three-ratio method to judge the type of faults. Based on this, the improved three-ratio method has been developed [23–25]. For example, Zhang et al. [23] proposed an improved three-ratio method as a calculation method for transformer fault basic probability assignment (BPA), which meets the requirements of BPA function, and its calculation result quantitatively reflects the probability of various faults. Zhang et al. [24] presented an improved three-ratio method based on the B-spline theory, which avoids the limit of the original three-ratio method with fixed boundary and is a new idea for solving fault diagnosis problems. This improved method can maintain the feature of identifying the majority of the samples, and can make the three-ratio method have learning ability.

In China, more than 50% of the transformer faults in the power system are found via DGA-based tests which are conducted for the diagnosis of transformer fault types and their level of severity according to the content, ratio to each other, and gas production rate of the dissolved gases in the transformer oil. Hence, besides the three main traditional ratio methods above, some improved methods have been investigated, including the Rogers method [26], Electric Association Research Society method and its improved version [27], improved/new three-ratio method (also called IEC three-ratio method) [28], Dornenburg two-ratio judgment method [29], basic triangular diagram method [30], gas-dominated diagram method [31], Germany's four-ratio method [26], hydrogen-acetylene-ethylene (HAE)-based triangular diagram method [26], thermal-discharge (TD) diagram method (also called TD graphic interpretation method) [32] and simplified Duval method [26].

The advantages and disadvantages of these ratio methods based on DGA are compared in Table 4. In actual application, these traditional methods are generally combined together for a comprehensive analysis in order to find the fault part of the transformer. As shown in Table 4, in the traditional transformer fault diagnosis, generally, the more detailed the classification of fault types, the lower the probability of correct judgment, and vice versa. Nevertheless, too rough a classification is not conducive to the accurate judgment of the fault. Due to the objective uncertainty of the cause-and-effect relationship of the transformer fault itself, as well as the uncertainty of the boundaries of the subjective judgment of the testing data, it is difficult to meet the requirements of engineering application with the above ratio methods, but in practice, the accuracy can be improved by using multiple hierarchical integrated diagnosis methods. Addressed concretely, first, we use the fuzzy judgment method to identify the possible fault types, such as discharge and overheating, which helps to identify the faults preliminarily, and is not easy to make a misjudgment. Secondly, we use those diagnosis methods which can realize more detailed fault classification to conduct careful judgment of the fault types. Finally, by implementing a comprehensive analysis, the correct fault type can be determined. By using this diagnosis methodology in traditional transformer fault diagnosis, on the one hand the misjudgment rate can be reduced, on the other hand the correct judgment rate can be improved.

Table 4. A comprehensive comparison of the traditional DGA based ratio methods in actual transformer fault diagnosis.

Traditional Methods	Characteristic Gases	Advantages	Disadvantages	Quality Grading
IEC three-ratio method [33–36]	CH_4/H_2 C_2H_4/C_2H_6 C_2H_2/C_2H_4	▪ The sequence of known faults is arranged more reasonable from incipient fault to severe fault based on the ratios; ▪ The most basic oil-filled power equipment fault diagnosis method based on the result of DGA; ▪ The fault types are reduced from eight in the past to six now, making the classification more flexible.	▪ More roughening classification; ▪ Accuracy is unsatisfactory for compound-faults; ▪ Incomplete coding, some cases cannot be diagnosed; ▪ The attention value and criteria specified for the characteristic gas content are too absolute; ▪ Cannot determine the exact location of the faults; ▪ Prone to misjudge with a high misjudgment rate; ▪ Poor dealing with mixed fault types.	★★★
Basic triangular diagram method [30]	CH_4, C_2H_4, C_2H_2 (relative content)	▪ A more intuitive diagram method to use DGA results for transformer fault analysis ▪ Can be widely used in the field fault diagnosis	▪ Limited to the scope of threshold diagnosis	★★★☆
Gas-dominated diagram method [31]	H_2, CH_4, C_2H_4, C_2H_6, C_2H_2 (relative concentration ratio, ppm)	▪ A more intuitive diagram method to use DGA results for transformer fault analysis ▪ Can be widely used in the field fault diagnosis	▪ Limited to the scope of threshold diagnosis	★★★☆
Characteristic gas method [11–13]	TH [1], H_2, CH_4, C_2H_4, C_2H_6, C_2H_2, etc.	▪ Can make a judgment of the nature of the fault according to the determination of the gas chromatography of the insulating oil, characteristics of gas production, and attention value of characteristic gas	▪ Make a preliminary and rough judgment of whether there is a fault and the nature of the fault	★★★
Gas production rate method [14–17]	Absolute and relative gas production rate	▪ Can further reflect existence, severity and development trend of the fault according to gas production rate of the fault point	▪ Cannot determine the exact location of the fault ▪ Be prone to misjudge the faults involving different types of faults with the same gas	★★★☆

		▪ Has a good diagnostic effect on overheating, electric arc and insulation breakdown faults	characteristic	
Electric Association Research Society method and its improved method [27]	/	▪ Fault category is simplified ▪ The upper and lower limits of the ratio range corresponding to the coding are more clearly defined ▪ A lower rate of false negative ▪ Can accurately judge the faults of overheating and discharge and has wide coverage	▪ The code combination of fault type superposition is not taken into account in practice ▪ Not in line with the actual situation to delete the code combination of 010 and 001 in the IEC method ▪ Still unable to deal with some faults	★★★★
Dornenburg two-ratio judgment method [29]	C_2H_2/C_2H_4, CH_4/H_2	▪ Determine the fault types according to the area in which the ratio is in a graph ▪ A higher rate of accurately judging overheating faults	▪ A preliminary and rough judgment ▪ The rate of misjudgment or false negative is higher	★★★
Germany's four-ratio method [26]	CH_4/H_2, C_2H_6/CH_4, C_2H_4/C_2H_6, C_2H_2/C_2H_4	▪ The classification of fault types is more specific ▪ Has a high accurate rate of judging the fault of high-temperature overheating	▪ Too many criteria which lead to a high rate of missed judgment ▪ Has a lower accurate rate of judging the low-energy discharge ▪ Cannot identify the partial discharge	★★★
HAE based triangular diagram method [26]	H_2, C_2H_4, C_2H_2 (relative content)	▪ Can be used as an empirical criterion and auxiliary reference ▪ Has a lower rate of misjudgment or false negative ▪ Has a wide coverage	▪ It is not convenient to consider the change in the proportion of alkenes and alkanes because of the removal of alkanes, and is unfavorable to estimating the temperature of local overheating	★★★☆
TD graphic interpretation method [32]	CH_4/H_2, C_2H_2/C_2H_4	▪ Can be better to distinguish the high-temperature overheating fault and discharge fault in inner part of the transformer ▪ Can quickly and correctly judge the nature of fault ▪ Can directly reflect the development trend of fault	▪ Cannot determine the exact location of the fault	★★★★
Rogers method [26]	/	▪ No blind spots exist in the coding ▪ Compound fault can be judged and the accuracy is satisfactory	▪ Cannot determine the exact location of the fault	★★★☆
Simplified Duval method [26]	CH_4, C_2H_2, C_2H_4	▪ Can be used as an auxiliary criterion ▪ More accurate judgment for the overheating fault	▪ Has a lower accurate judgment rate for the discharge fault	★★★☆

[1] TH = total hydrocarbon.

In addition, these mentioned traditional gas chromatography methods possess a good diagnostic power for the faults such as overheating and electrical arcing, and insulation-damaging failures. However, these methods, more or less, all have some defects, as shown in Table 4. For several examples, the characteristic gas method has low recognition precision and lower efficiency, meanwhile the three-ratio and improved three-ratio methods have disadvantages of incomplete coding and excessively absolute coding boundary. These shortcomings will undoubtedly be very harmful to the diagnosis of the latent faults in power transformers. Hence, the traditional methods cannot accurately determine the position of the fault. Moreover, for the different types of faults which have the same gas feature, it is easy to misjudge them when using traditional methods.

Therefore, due to complexity of transformer faults, a single method cannot be adopted in the diagnostic process, but rather a variety of methods should be employed. In other words, it is essential to explore the principles, methods and means from various disciplines that are helpful to the fault diagnosis of transformers, so as to make the fault diagnosis technology interdisciplinary. Aiming at the limitations of traditional methods above, with the rapid development of computer technology and artificial intelligence (AI) theory, multiple intelligence techniques, including artificial neural network (ANN) [37–46], expert system (EPS) [47–51], fuzzy theory [52–58], rough sets theory (RST) [36], grey system theory (GST) [59–66], and other intelligent diagnosis tools [5,67–92] such as swarm intelligence (SI) algorithm, data mining technology, machine learning (ML), mathematical statistics method, wavelet analysis (WA), optimized neural network, Bayesian network (BN), and evidential reasoning approach, have been introduced to the research field of transformer fault diagnosis based on the DGA approach. These intelligent methods make up for the deficiencies of the mentioned traditional DGA methods, and directly or indirectly improve the accuracy of transformer fault diagnosis, and provide a new train of thought for high-precision transformer fault diagnosis. For example, the EPS is considered one of the main forms of AI and the most active and extensive application fields in the application research of AI. Hence, in view of the professionalism, empiricism and complexity of transformer fault diagnosis, the application of EPS-based diagnosis methods has unique advantages [47–51]. Recently, several other approaches or techniques have been proposed for fault diagnosis of transformers, such as Rigatos and Siano's [82] proposed neural modeling and local statistical approach to fault diagnosis for the detection of incipient faults in power transformers, which can detect transformer failures at their early stages and consequently can deter critical conditions for the power grid; Shah and Bhalja [85] and Bacha et al. [5] both proposed support vector machine (SVM)-based intelligent fault classification approaches to power transformer DGA. Furthermore, the random forest technique-based fault discrimination scheme [84] for fault diagnosis of power transformers, as well as the multi-layer perceptron (MLP) neural network-based decision [46], vibration correlation-based winding condition assessment technique [86], and induced voltages ratio-based thermodynamic estimation algorithm [73] have been proposed consecutively. Besides, in order to develop more accurate diagnostic tools based on DGA, a large number of information processing-based algorithms have been extensively investigated, e.g., Abu-Siada and Hmood [88] proposed a new fuzzy logic algorithm to identify the power transformer criticality based on the dissolved gas-in-oil analysis; Illias et al. [89] developed a hybrid modified evolutionary particle swarm optimizer (PSO) time varying acceleration coefficient-ANN for power transformer fault diagnosis, which can obtain the highest accuracy than the previous methods; Pandya and Parekh [90] presented how interpretation of sweep frequency response analysis traces can be done for open circuit and short circuit winding faults on the power transformer. All of the above mentioned intelligent approaches have improved the conventional DGA-based transformer fault diagnosis methods, and directly or indirectly improved the accuracy of fault diagnosis for the oil-immersed power transformers [91,92]. In essence, the application of AI for transformer fault diagnosis is fundamentally still based on the analysis of the content of dissolved gas in transformer oil. Hence, these presented intelligent algorithms using DGA techniques have provided new ideas for high-precision transformer fault diagnosis. Based on these DGA principle-based intelligent algorithms, this paper conducts a detailed and thorough survey on the application of AI methods using DGA in the fault diagnosis of the oil-immersed power transformers. Finally, this paper summarizes and prospects the development direction of future transformer fault diagnosis methods.

The novel contributions of this paper can be summarized as follows: a detailed survey on various intelligent approaches and techniques, including EPS, ANN, fuzzy theory, RST, GST, SI algorithms, data mining technology, ML algorithms and other intelligent methods, applied in fault diagnosis and decision making of the power transformer, with the component content of the dissolved-gases in transformer oil as characteristic quantities, is conducted systematically. In this survey, drawing on the current research situation for this field, the advantages and existed issues of these intelligent approaches and techniques in the process of application have been described and

investigated thoroughly in the first, and then the problems that must be addressed in the fault diagnosis and decision making of the transformer based on DGA are identified in detail, and finally the prospects for their future development trends and research directions are outlined. It is concluded that future development of fault diagnosis and decision making of the transformer based on DGA should be combined with various intelligent algorithms and techniques, which complement each other to form a hybrid fault diagnosis network. The systematic survey in this paper provides references and guidance for researchers in choosing appropriate fault diagnosis and decision making methods for the oil-immersed power transformers in preventive tests.

The remainder of the paper is organized as follows: the application of EPS in DGA-based transformer fault diagnosis is summarized thoroughly in Section 2. Moreover, the applications of ANN, fuzzy theory, RST and GST in transformer fault diagnosis using DGA technique are comprehensively reviewed in Sections 3–6, respectively. Besides, the applications of other intelligent algorithms, including SI algorithms, data mining technology, ML algorithms, and other intelligent diagnosis tools such as mathematical statistics method, wavelet analysis (WA), optimized neural network, Bayesian Network (BN) and evidential reasoning approach, in DGA based transformer fault diagnosis are made a detailed review in Section 7. In Section 8, the future development direction of transformer fault diagnosis using DGA is discussed and prospected. Finally, Section 9 concludes the paper.

2. Application of EPS in DGA-Based Transformer Fault Diagnosis

2.1. Description of EPS-Based Transformer Fault Diagnosis Using DGA

EPS is a smart computer program system which contains a great deal of expertise and can accurately simulate experts' experience, skill and reasoning processes [47,93]. Here, EPS is focused on chromatographic analysis of dissolved gas in oil, in which the three-ratio method and the method of characteristic gases are employed to implement preliminary analysis of the operation condition of the transformer and judge the fault types of the transformer. At the same time, the knowledge-based program [94] is established by combining the data from external inspections, the characteristic tests of insulation oil, the preventive inspections of insulating oil, etc. Moreover, in the comprehensive analysis module, based on the analysis results of gas chromatography, external inspection, insulation oil characteristics and insulation preventive testing module, the operation status of the transformer is analysed and judged, and operational suggestions are provided to operators. Besides, the coordinator is the main module, which controls and coordinates the work of the gas module.

EPS is good at logic reasoning and symbol processing. It has an explicit knowledge representation form and can explain the reasoning behaviour, and use deep knowledge to diagnose faults. The biggest merit of EPS is to achieve a comprehensive analysis of a large number of testing data and monitoring information. In this analysis process, EPS is employed to combine with expert experience to make a diagnosis comprehensively, accurately and quickly, which provides reasonable advice for the maintenance personnel as well as scientific information for further maintenance. Recently, researchers have carried out a lot of research in the field of transformer fault diagnosis using the EPS, and developed a series of expert systems with fault detection and diagnosis functions [47,49]. Moreover, these expert systems are integrated with a rich knowledge base which is developed based on fault phenomena, gas analysis in oil, and electrical and insulation testing results, as well as based on case diagnosis. In aspect of reasoning, these expert systems are combined with ANN [48], fuzzy mathematics [50], etc. and have shown the potential practical value and broad application prospect in practice [51]. A DGA-based EPS for transformer fault diagnosis is generally composed of seven parts [95] as introduced as follows:

(a) Transformer fault diagnosis knowledge base: it is established as a modular structure and the core of the whole diagnosis system. As introduced, usually, this knowledge base is established by focusing on gas chromatography analysis, and at the same time, it combines some testing

means, such as external inspections, insulation oil characteristic tests, and insulation preventive inspections and tests.

(b)   Comprehensive database: it is composed of two parts, among them, one part is gas analysis module, and the other part is an insulation damage prevention database and dynamic database. The two parts are used to perform the dynamic and static calls of the data. In the former part, all kinds of gas data and insulation prevention data can be archived as historical data so that users can inquire and manage it at any time. This part draws the final conclusion, carries on the longitudinal analysis according to the current input data and the integration of the trend of historical change, and carries on the transverse analysis with the related test data. The latter part is a context tree that stores intermediate reasoning results and final judgment conclusions so that they can be invoked by the interpretation mechanism when the user needs to explain.

(c)   Reasoning engine: its role is mainly to solve some fuzzy and uncertain issues. In this process, the goal-driven reverse reasoning is achieved, as well as the fuzzy logic is introduced, so that it can successfully handle some fuzzy problems.

(d)   Learning system: it is the interface with the experts in the practical field, through which, the knowledge of the experts in the field can be extracted, classified and summarized, such that the knowledge is formalized and encoded in the diagnostic knowledge base formed by the computer system.

(e)   System context: it is a place where intermediate results are stored. A notebook is provided by the system context for the reasoning engine to record and guide the work of the reasoning engine, so that the reasoning engine can work smoothly.

(f)   Sign extractor: it is a typical human-computer interaction interface [96,97]. Here, the sign is sent into the system via this interface using the man-machine interactive mode.

(g)   Interpreter: it is also a typical human-machine interaction interface. It can answer all the questions that the user has put forward at any time.

Based on the description of the EPS-based transformer fault diagnosis using DGA, and according to [98], the interrelationship of each component introduced above is shown in Figure 1.
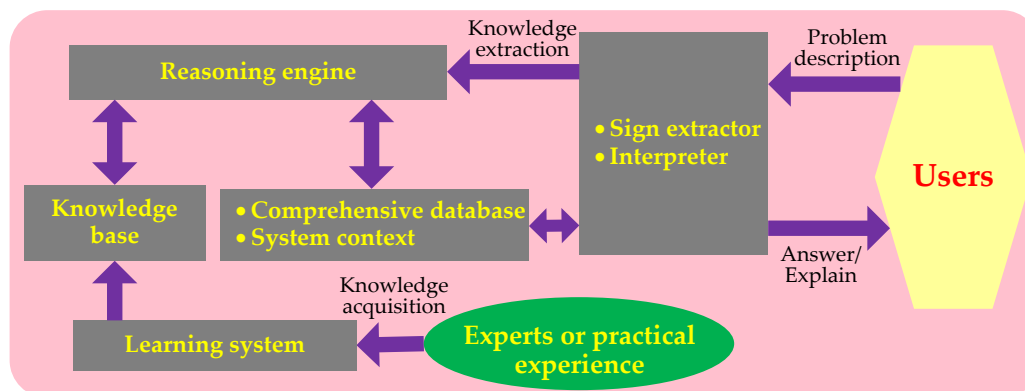
Figure 1. The interrelationship of each component in the EPS.

2.2. EPS-Based Transformer Fault Diagnosis Using DGA: A Survey

Power transformers are complex systems. In DGA-based fault diagnosis systems, incomplete information and uncertain factors always exist, such that it is often difficult to obtain complete test data in practice. Therefore, EPS has been widely used in DGA-based transformer fault diagnosis systems. Lin et al. [47] developed a prototype of an EPS based on the DGA technique for diagnosis of suspected transformer faults and their maintenance actions. In this system, not only a synthetic method is proposed to assist the popular gas ratio, but also the uncertainty of key gas analysis, norms threshold and gas ratio boundaries are managed by using a fuzzy set concept, so this designed EPS finally shows effectiveness in transformer diagnosis by via testing it for Taiwan

Power Company's transformers gas records. Saha and Purkait [49] developed an EPS in order to address the issue that insulation condition assessment is usually performed by experts with special knowledge and experience due to the complexity of the transformer insulation structure and various degradation mechanisms under multiple stresses, which can imitate the performance of a human experts, to make the complicated insulation condition assessment procedure accessible to plant maintenance engineers. The application examples show that this designed EPS can provide accurate insulation diagnosis. Chen and Li [96] developed an EPS for power transformer insulation fault diagnosis, which takes DGA as the characteristic parameter. The diagnosis results from practical application show that this designed EPS can comprehensively analyse the insulation status of transformer, identify the type of fault correctly, and determine the location, severity and development trend of the fault. However, for some specific faults, this system cannot achieve an accurate diagnosis. In view of this situation, Jain et al. [97] used the fuzzy technique to find out the association matrix between fault causes and phenomena based on the sample, which overcomes the issue of knowledge acquisition by EPS to some extent. Shu et al. [98] used the RST with strong data analysis ability and error tolerance to realize the establishment of a complete knowledge base for the transformer fault diagnosis EPS. Du [99] designed an EPS based on information integration and multi-layer distributed reasoning mechanism, in which the chromatographic data collected from 221 fault transformers are used as an original fault sample set to conduct transformer fault diagnosis. The diagnosis results show that the accuracy of comprehensive diagnosis is 89%. In addition, Wang et al. [48] developed a combined ANN and EPS tool for transformer fault diagnosis using dissolved gas-in-oil analysis. In this system, the combination of the ANN and EPS outputs has an optimization mechanism to ensure high diagnosis accuracy for all general fault types. The test results show that this developed system has better performance than ANN or EPS used individually. Apart from the combination of ANN, EPS can be combined with fuzzy theory [50], comprehensive relational grade theory [51], etc. Here, due to the limitation of training data and non-linearity, Mani and Jerome [50] presented an intuitionistic fuzzy EPS to diagnose several faults in a power transformer, such that the estimation of key-gas ratio in the transformer oil can become simpler. This proposed method can identify the type of fault developing within a transformer even if there is conflict in the results of AI technique applied to DGA data. In addition, Li et al. [51] proposed a new comprehensive relational grade theory which is applied to EPS of transformer fault diagnosis and improves effectively the running and maintenance of power transformer. The database and repository in this EPS is an open system, which guarantees that new fault sample can be added into the system and repository can be classed and modified by experts.

Although some research results of the EPS in the DGA-based transformer fault diagnosis have been achieved, there still some urgent issues to be addressed, which are mainly presented in the following three aspects:

- Completeness is difficult to achieve in the establishment of the fault diagnosis knowledge base. When s a fault symptom that does not exist in the knowledge base occurs, the EPS cannot identify the type of this fault due to the fact no corresponding fault rule is established in the knowledge base.
- The accuracy is difficult to be grasped when diagnosing some fault symptoms with indeterminate mathematical correlation.
- The knowledge management is rather difficult because the establishment of the adopted knowledge-based rule-based system. Moreover, due to the complexity of construction algorithms, it is rather troublesome when the knowledge base has to being maintained.

In recent years, Flores et al. [100] presented an efficient EPS for power transformer condition assessment, in which a knowledge mining procedure is performed as an important step, by conducting surveys whose results are fed into a first Type-2 Fuzzy Logic System (T2-FLS). In this step, the condition of the transformer taking only the results of DGA into account can be initially evaluated. The use of T2-FLS can allow the inclusion of other factors as inputs of the diagnostic algorithm, which could be either new influence factors or a combination of the ones used in the designed EPS. In addition, Ranga et al. [101] proposed a fuzzy logic-based EPS for condition

monitoring power transformers, in which the fuzzy logic model utilizes the data gathered from various diagnostic tests to determine the overall health index of the transformers. This proposed model on one hand can determine the individual health index of transformer oil and paper insulation, and on the other hand it can identify the incipient faults present within the transformers and handle all situations corresponding to single or multiple faults. Ranga et al. have tested 30 transformer oil samples from Indian railways which were collected from different traction substations. The test results proved the efficacy and reliability of the proposed technique. Žarković and Stojković [102] also presented a methodology for power transformer condition monitoring and diagnostics based on the analysis of AI expert systems. The possibility of the presented monitoring methodology is to assist the operator's engineers in decision making about urgency of intervention and type of maintenance of power transformer. They have analysed the application of the Mamdani-model and Sugeno-model in fuzzy EPS for fault diagnosis based on the current state of the power transformer. The testing results show acceptable effectiveness of this proposed fuzzy EPS in detecting different faults and might serve as a good orientation in the power transformer condition monitoring.

Overall, for the EPS applied in the DGA-based transformer fault diagnosis, there are two urgent issues to be solved in the future. The first one is the bottleneck of knowledge acquisition. This is because on the one hand, the knowledge of experts is incomplete, and on the other hand, it is difficult to achieve rule-based expert knowledge representation. The second is the uncertainty of diagnostic reasoning, especially for some fault phenomena which are not very definite in mathematical correlation, the accuracy of the diagnosis is difficult to be guaranteed. Therefore, the two above burning problems substantially affect the accuracy of transformer fault diagnosis when using the DGA techniques. A summary for the application of EPS in DGA based transformer fault diagnosis is presented in Table 5 as follows.

Table 5. A summary for the application of EPS in DGA based transformer fault diagnosis.

Advantages and Disadvantages	Main Components	Primary Means
▪ good at logic reasoning and symbol processing ▪ has an explicit knowledge representation form ▪ can explain the reasoning behaviour ▪ use deep knowledge to diagnose faults ▪ can achieve a comprehensive analysis of a large number of testing data and monitoring information ▪ incomplete fault diagnosis knowledge base ▪ accuracy is not high when diagnosing some fault symptoms ▪ knowledge management and maintenance is rather difficult ▪ weak ability of knowledge acquisition ▪ uncertainty of diagnostic reasoning	▪ transformer fault diagnosis knowledge base ▪ comprehensive database ▪ reasoning engine ▪ learning system ▪ system context ▪ sign extractor ▪ interpreter	▪ combined with ANN [48] ▪ combined with fuzzy mathematics [50,102] ▪ combined with fuzzy set [47,97] ▪ combined with rough sets theory [98] ▪ combined with information integration and reasoning [99] ▪ combined with comprehensive relational grade theory [51] ▪ combined with knowledge mining technology [100] ▪ combined with fuzzy logic model [101]

3. Application of ANN in DGA Based Transformer Fault Diagnosis

As reviewed in Section 2, it is essential to combine the EPS with other AI techniques so that the EPS can play a better role in transformer fault diagnosis based on DGA. Therefore, when the development of EPS in transformer fault diagnosis using DGA meets with some technical obstacles, the research and application of ANN is developing rapidly, especially the new AI techniques, such as improved probabilistic neural network [41], self-adaptive radial basis function (RBF) neural network [42], knowledge discovery-based neural network [43], knowledge extraction-based neural network [44], fuzzy reasoning-based neural network [45], MLP neural network-based decision [46], back propagation (BP) neural network [103], recurrent ANN [104], deep learning (DL) based ANN [105], hybrid ANN and EPS [106], and generalized regression neural network (GRNN) [40,107].

Besides, the combination of ANN and mathematical morphology has been applied for the transformer fault diagnosis [108]. Hence, recently, combining with DGA, the development of ANN theory, which is based on non-linear parallel processing technique, provides a new way for transformer fault diagnosis.

Here, the ANN is a type of non-linear dynamic network system that simulates the structure of human brain neurons. It has abilities of large-scale parallel information processing, strong fault tolerance, robustness and self-learning function [109]. It can map the input and output relationships of highly non-linear and unascertained systems [110]. Hence, ANN is very suitable for solving the issues of transformer fault diagnosis [111–113].

3.1. Basic Idea of Transformer Fault Diagnosis System Based on ANN

The basic idea of an ANN-based transformer fault diagnosis system can be stated as follows. First, the input and ideal output of the system are used as the type of characteristic gas dissolved in transformer oil and the type of fault corresponding to the characteristic gas, respectively. Second, the input variable produces the actual outputs through the ANN. Lastly, the deviation between the ideal output and the actual output is employed to dynamically adjust the connection weights of ANN, thus forming a network structure with transformer fault decision classification function.

Hence, the working process of the ANN-based transformer diagnosis system consists of two stages as follows [114]:

- Learning stage. In the process of learning, gas analysis data and other various testing data which come from the calculation results of historical data of the transformer will be treated as data sets to be read into the neural network, and then the weights and thresholds will be calculated via the BP learning calculation method.
- Working stage. During the fault diagnosis, the testing samples from different power transformers will be calculated to obtain actual outputs of the network, and finally these outputs will be compared with expected outputs of the network. In general, the ANN-based transformer fault diagnosis system uses a modular structure, in which the sample training of each module is conducted independently. In the main module of ANN, horizontal and longitudinal, historical and current comprehensive analysis and judgment will be conducted according to the analysis result of each module. Then, the result of analysis and judgment propagates through the forward channel to each hidden layer node of the main module. After that, the result is propagated to each node of the output layer via the action of activation function. Finally, the diagnosis conclusion can be output through the activity function of the output point.

Hence, for a given training sample, ANN has the following functional advantages:

(a) ANN can better implement the failure mode representation and then form the required decision classification areas.
(b) ANN can simplify the process of sample training.
(c) The nodes, hidden-layer nodes, and activation function of the network are tended to be simple, which accelerates the speed of diagnosing.
(d) The fuzzy logic theory has been introduced into ANN, which can better address some issues with data uncertainty.

3.2. ANN-Based Transformer Fault Diagnosis Using DGA: A Survey

In addition to the above basic operation stages, generally the first step is to normalize the input variables of the network, such as when a fuzzy technique has been used to conduct data pre-processing [111], in order to reduce the impact of different order of magnitude of the input variables in the network on the network convergence performance. Furthermore, the number of hidden-layer nodes of the network will also affect the network convergence performance; accordingly, Wang et al. [115] took the application of single hidden-layer neural network in the

DGA-based transformer fault diagnosis as an example, and based on which, the influence of the number of hidden-layer nodes on the training effect and generalization ability of the network has been elaborated. On the basis of [115], Zhang et al. [37] investigated the application of double hidden-layer neural network in DGA-based transformer fault diagnosis, in which the convergence speed and training error of the network with different numbers of hidden layers and same numbers of input and output nodes are compared, and the results show that the proposed method has a better effect in fault diagnosis.

The training algorithm of neural network usually adopts a BP algorithm, hence Liu [116], based on collected 105 learning samples, adopted a supervised learning BP neural network for diagnosis and the accuracy of diagnosis was over 83%; Zhang et al. [37] deemed that the neural network with a single hidden layer has the best classification effect after investigating the influence of double-layer BP network structure on chromatographic diagnosis results, and it has the minimum amount of computation and at the same time it fully satisfies the requirements of the non-linear mapping between the failure phenomenon and the cause. However, there are some defects in the BP algorithm, such as the fact it easily falls into local convergence (i.e., easily falls into local minima), the accuracy of the solution is not high, and higher requirements for initial values. To address this concern, various improved algorithms have been proposed, such as the BP neural network for variable learning rate [106], the homotopic BP algorithm [117], and the BP algorithm with momentum term [118]. Apart from the common BP neural network structure, there are some other types of network structure, such as probabilistic neural network structure [119], combined genetic algorithm (GA) multi-layer feedforward network [120], competitive learning theory based self-organized network [121], RBF network [122,123], and WNN [67,124–127]. These improved ANN-based models have enhanced the accuracy of transformer fault diagnosis to varying degrees, which can be seen a new exploration of transformer fault diagnosis.

In the 1990s Zhang et al. [37] proposed an ANN approach to the diagnosis and detection of faults in oil-filled power transformers based on DGA, in which a two-step ANN method is employed to detect faults with or without involving cellulose that obtains a good diagnosis accuracy; Castro and Miranda [43] described a new methodology for mapping a neural network into a rule-based fuzzy inference system, in order to make explicit the knowledge captured during the learning stage. The proposed method is applied in transformer fault diagnosis using DGA and illustrates the good results obtained and the knowledge discovery made possible. In order to extract knowledge from trained ANN so that the user can gain a better understanding of the solution arrived by the neural network, Bhalla et al. [39] applied a pedagogical approach for rule extraction from functions approximating ANN with application to incipient fault diagnosis using the concentrations of the gases dissolved in transformer oil as the inputs. This proposed methodology has been successfully applied in transformer incipient fault diagnosis. Lin et al. [40] proposed a combined predicting model based on kernel principal component analysis and a GRNN using an improved fruit fly optimization algorithm to select the smooth factor. This method shows a better data fitting ability and more accurate prediction ability compared with SVM and grey model (GM) methods. In order to improve the accuracy of ANN applied in the transformer fault diagnosis, Yi et al. [41] proposed a variant of probabilistic neural network with self-adaptive strategy, called self-adaptive probabilistic neural network, which can solve the transformer fault diagnosis problem and shows a more accurate prediction and better generalization performance when compared with other neural networks. Moreover, Meng et al. [42] presented a novel hybrid self-adaptive training approach-based RBFNN for power transformer fault diagnosis, which clearly demonstrates the improved classification accuracy compared with other alternatives and shows that it can be employed as a reliable transformer fault diagnosis tool. In addition, Souahlia et al. [46] used an improved combination of Rogers and Doernenburg ratios DGA to make MLP neural network-based decisions for power transformers fault diagnosis. This developed pre-processing approach can significantly improve the diagnosis accuracies for power transformer fault classification. Besides, Dong et al. [124] proposed a least squares weighted fusion algorithm integrated with rough set and fuzzy WNN (FWNN) for transformer fault diagnosis using DGA. In this method, on the one hand it

can better improve the diagnosis accuracy, when the output vector of single FWNN has the similar element. On the other hand, its diagnosis accuracy cannot be limited by the neural network hidden layer number and correlated training parameter. This proposed mechanism shows good diagnosis classification ability.

Hence, the brief overview above shows that ANN has been widely used in transformer fault diagnosis based on the DGA technique. Notwithstanding, although ANN can deal with very complicated classification problems, and it has achieved good results in DGA-based transformer fault diagnosis, there are still some shortcomings in ANN diagnosis technique as follows:

(a) Its performance is limited by the number of selected training samples, thus its diagnostic performance generally depends on the completeness of the training sample.
(b) Users can only see the inputs and outputs (it operates like a black box) so the process of intermediate analysis and deduction cannot be understand.
(c) The representation and utilization of knowledge is generally single, imperfect and incomplete.
(d) The phenomenon of oscillation easily occurs in the identification and affects the application of ANN in high-accuracy transformer fault diagnosis.

As a result, more and more researchers tend to combine the ANN diagnosis techniques with other intelligent algorithms, which is expected to become a rapid development direction of transformer fault diagnosis based on DGA in the future. For example, a RBFNN-based transformer fault diagnosis model was developed in [128], but the process of modelling is more complicated. Among most neural network models, GRNN is a neural network with a high parallelism, thus it just needs a small sampling of data while the output results of the network can still be converged to the optimal regression surface with a simple algorithm structure, high approximation accuracy, and better non-linear convergence performance [129]. Based on GRNN, Ding et al. [107] developed a transformer fault diagnosis model based on the DGA method and GRNN, in which the input eigenvector of the GRNN-based fault diagnosis model is achieved via the DGA method. This model is employed to conduct simulation experiment based on four typical fault diagnosis cases of a main transformer in a certain substation, and at the same time it is compared with the diagnosis results of the typical BP neural network (BPNN), and Levenberg Marguardt algorithm (LM)-improved BPNN (called LM-BPNN). The simulation showed that this combined DGA and GRNN transformer fault diagnosis model has faster diagnosis speed, higher classification accuracy, stronger generalization ability and the establishment of the model is simple. Here, according to [107], the principle of GRNN algorithm is briefly introduced as follows: the GRNN is composed of four layers, including input layer, model layer, summation layer and output layer. Based on non-linear regression analysis, GRNN uses sampling data as a post-condition for Parzen non-parametric estimation. Note that GRNN does not need to know the exact equation form, but just needs to calculate the probability density function so as to obtain original equation form. Hence, GRNN obtains the joint probability density function between independent and dependent variables from the sample data sets. As elaborated in [107], assume that two random variables are x and y, and the joint probability density is $f(x, y)$, and then the regression expression of y for x is shown as:

$$E[y \mid x_0] = \frac{\int_{-\infty}^{0} y f(x_0, y)\mathrm{d}y}{\int_{-\infty}^{0} f(x_0, y)\mathrm{d}y} \tag{1}$$

By using the Parzen non-parametric estimation theory, the probability density function $f(x_0, y)$ of the sample sets (x_i, y_i) ($i = 1, 2, 3, \ldots, n$) can be obtained as shown in (2), where n is the content of sample sets. p is the number of dimensions. σ is the distribution density of the RBF:

$$f(x_0, y) = \frac{1}{n(2\pi)^{\frac{p+1}{2}} \sigma^{p+1}} \sum_{i=1}^{n} e^{-\mathrm{d}(x_0, x_i)} e^{-\mathrm{d}(y_0, y_i)} \tag{2}$$

where $\ d(x_0, x_i) = \sum_{j=1}^{p} \left[\dfrac{(x_{0j} - x_{ij})}{\sigma} \right]^2\ $ and $\ d(y, y_i) = (y - y_i)^2$.

Based on (2), the predictive output of y can be obtained as shown in (3) and its final simplified form is shown in (4) as follows:

$$y(x_0) = \frac{\sum\limits_{i=1}^{n} [e^{-d(x_0, x_i)} \int\limits_{-\infty}^{+\infty} y e^{-d(y_0, y_i)} \mathrm{d}y]}{\sum\limits_{i=1}^{n} [e^{-d(x_0, x_i)} \int\limits_{-\infty}^{+\infty} e^{-d(y_0, y_i)} \mathrm{d}y]} \tag{3}$$

$$y(x_0) = \frac{\sum\limits_{i=1}^{n} y e^{-d(y_0, y_i)}}{\sum\limits_{i=1}^{n} e^{-d(x_0, x_i)}} \tag{4}$$

Apart from GRNN in [107], ANN can be combined with other intelligent algorithms for transformer fault diagnosis based on the DGA technique. Ghanizadeh and Gharehpetian [130] proposed a new method which combines ANN with cross-correlation-based features. This developed model is employed to discriminate between mechanical defects and electrical faults, which are as two major faults in power transformer windings. The principle of this model is shown in Figure 2. This proposed method can precisely discriminate among disc-to-disc short circuit faults, radial deformation and axial displacement defects and determine their location or extent with a good accuracy.
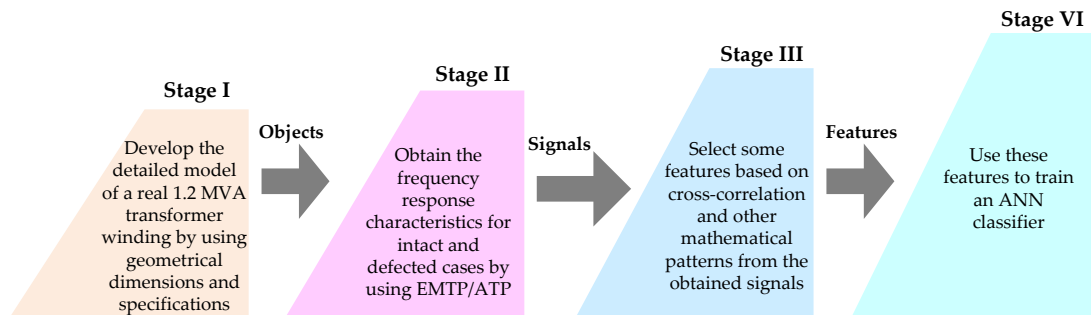
Figure 2. The principle of the model developed in [130].

Besides, a model combined estimation of distribution algorithm (EDA) with ANN is developed in [131], called EDA-ANN method, which is employed to realize the fault recognition with dissolved gas data. This EDA is a new population evolutionary algorithm based on a probabilistic model. In this EDA-ANN model, the outcomes can be put out with continuous inputs, thus the model can realize the transformer fault recognition with the continuous value of the inputs. The case based on some real fault data shows that this proposed method is feasible and accurate. The ANN can be trained by using adaptive back-propagation learning algorithm that converges much faster than the conventional back-propagation algorithm, based on which, Patel and Khubchandani [132] presented an improved ANN-based model to recognize the incipient faults of power transformers, which can improve the diagnosis accuracy of the conventional DGA approaches. In [108], Shi et al. proposed a new method which is based on mathematical morphology and ANN, in order to solve the discrimination between the magnetizing inrush and the internal fault of a power transformer when designing differential transformer protection. The ANN can also be combined with wavelet transform, and on this basis, Vanamadevi et al. [133] aimed at describing a method for the detection and classification of impulse faults in a transformer winding using the wavelet transform and an ANN, which is proved to be satisfactory in detection and classification of faults. In addition, Ying et al. [134] demonstrated a risk assessment method based on the combination of FAHP and ANN, in which the FAHP is employed to analyze the hierarchy structure of power

transformer and construct a fuzzy matrix. The results show that this FAHP-ANN method can overcome the disadvantage of ANN model structure in traditional risk assessment method, and it also shortens the time of assessment, increases the precision of the data and achieves the pre-set target. Swarm intelligence algorithms also have been combined with ANN in recent years, for example, Nashruladin [135] presented an application of ANN and GA for transformer winding/insulation faults diagnosis using DGA. In this model, a back-propagation training method is applied in ANN to detect the faults without cellulose involvement. At the same time, the GA is used to locate the optimal values to enhance the accuracy of fault detection. Besides, the DGA is chosen to diagnose the transformer faults and enables to carry out during online operation of the transformer. For another example, Zhang [136] proposed an evolutionary ANN programming based on Super SAB algorithm, which can improve diagnostic accuracy of conventional DGA methodologies. In this model, the Super SAB algorithm can provide both higher learning efficiency and stronger generalization capacity versus standard BP and Bold-Driver algorithm used in DGA, thus the author deemed that this algorithm possesses a promising future in the diagnostic field for power transformer equipment.

To sum up this section, we can conclude that ANN has been widely used in current transformer fault diagnosis based on DGA techniques. To overcome the defects of ANN, many improved ANN structures have been proposed by researchers, which can improve the accuracy of the fault diagnoses to a certain degree. In the future, the development of ANN in transformer fault diagnosis based on DGA will tend to be combined with more and more intelligent tools and algorithms, such as fuzzy logic, grey theory, EPS, SI algorithm, DL, reinforcement learning (RL), and other ML methods. This will be a promising development direction for the DGA-based transformer fault diagnosis in the future. A summary of the application of ANN in DGA-based transformer fault diagnosis is presented in Table 6.

Table 6. A summary for the application of ANN in DGA based transformer fault diagnosis.

Advantages and Disadvantages	Working Process	Primary Means
- large-scale parallel information processing ability - strong fault tolerance - robustness and self-learning function - can map the input and output relationships of highly non-linear and unascertained systems - better address some issues with data uncertainty - performance is limited by training samples - cannot understand the process of intermediate analysis and deduction - oscillation can easily occur	- learning stage: read various testing data sets into the neural network, and then calculate the weights and thresholds via BP learning - working stage: calculate the testing samples to obtain actual outputs which are then compared with expected outputs, and finally output diagnosis conclusion via the activity function of the output point	- improved probabilistic neural network [41,119] - RBF neural network [42,122,123,128,129] - knowledge discovery-based neural network [43] - knowledge extraction-based neural network [44] - fuzzy reasoning-based neural network [45] - MLP neural network-based decision [46] - BP neural network [103] - recurrent ANN [104] - DL based ANN [105] - hybrid ANN and EPS [106] - GRNN [40,107] - combined with mathematical morphology [108] - combined GA multi-layer feedforward network [120,135] - combined with competitive learning theory [121] - WNN and FWNN [67,124–127] - EDA-ANN [131] - combined with FAHP [134]

4. Application of Fuzzy Theory in DGA-Based Transformer Fault Diagnosis

4.1. Fuzzy Theory Description

The fuzziness introduced here refers to the uncertainty of the objective things in the real world in terms of state, property, etc. The most fundamental reason for this phenomenon is that the state of a thing is not unique, which means for between the states of right and wrong, there may be many intermediate and transitional states, and many states may even coincide, so there is no definite boundary between different states [137]. This fuzziness generally exists in objective things. The study of the interrelationship between fuzzy things is called fuzzy theory [137]. Hence, fuzzy theory

is a kind of intelligent technique with a complete fuzzy inference system, by introducing linguistic variables and approximation reasoning as fuzzy logic based on classical set theory, in order to achieve fuzzification of classical set theory.

In the study of fuzzy theory, the concept of membership function is introduced. This function is used describe a function from a fully membership status to a completely non-membership state, in which the degree of membership is employed to evaluate the degree of similarity of fuzzy information. The introduction of membership function can help fuzzy theory better solve the fuzziness of man's natural language, thus the membership function is one of the most core concepts of fuzzy theory. The characteristics of fuzzy theory lie in the positive recognition of the existence of subjective issues, thus the fuzzy set theory can be applied to deal with these issues that are not easy to be quantified in the real world, so as to deal with man's subjective evaluation issues in an appropriate and reliable manner. Fuzzy theory has been widely applied in comprehensive evaluation of things, and this evaluation method is called fuzzy comprehensive evaluation method. Its basic principle is demonstrated as follows:

- First, determine the evaluation factors and its evaluation criteria and weights, so as to establish the factor set of evaluation object. In addition, it is essential to construct the evaluation grade, for example, the operation state of power transformer can be divided into four grades, including normal state, attention state, abnormal state and serious state.
- Then, determine the fuzzy membership function that is used to conduct pre-processing of the original data of gases dissolved in transformer oil. Concretely, select the appropriate membership function to accurately establish the complicated fuzzy relationship between the transformer fault and fault phenomenon. A suitable membership function is crucial to the entire fault diagnosis of the transformer. In [138], Zhang et al. selects the fuzzy results of three ratios in the three-ratio method as the model input of the SVM, and they are x_1 = C_2H_2/C_2H_4, x_2 = CH_4/H_2, and x_3 = C_2H_4/C_2H_6. The corresponding membership functions $f_1(x_1)$, $f_2(x_2)$ and $f_3(x_3)$ can be seen in [139]. The outputs of the three membership functions represent the input matrix of the SVM model, which are used to train or test the SVM model.
- Next, adopt the degree of membership to describe the fuzzy boundaries of the factors according to the principle of fuzzy set transformation, so as to construct a fuzzy evaluation matrix.
- Lastly, determine the final grade of the evaluation object through repeated calculations.

In a power transformer, there is a lot of uncertainty and fuzziness in its fault phenomena, fault causes, and fault mechanisms. The traditional precise mathematical theory can hardly describe the relationship between them, so it is difficult to diagnose the true faults of the transformer and their causes. As stated above, the fuzzy theory can be used to make a quantitative analysis of human fuzzy thinking and fuzzy language, and find out the fuzzy judgment that is suitable for the computer to imitate the human brain. In the DGA data-based transformer fault diagnosis, there are more serious uncertainties and fuzziness among the fault phenomena, fault causes, fault mechanisms and fault classifications. To address it, the fuzzy theory is gradually employed by researchers in order to solve these issues which have fuzziness and uncertainty since the fuzzy theory was proposed. Hence, fuzzy theory has provided an effective approach to solving the issues with fuzziness and uncertainty in transformer fault diagnosis based on DGA.

Concretely speaking, the fuzzy theory applied in the DGA-based transformer fault diagnosis can be described as follows [140,141]:

(a) First, it is necessary to establish a DGA-based transformer fault database as the basic database, which is employed for the establishment of fuzzy rules.
(b) Then, the DGA data of the transformer is treated as the inputs, on which fuzzification, fuzzy processing and defuzzification are conducted to determine the results of fuzzy diagnosis.
(c) When the difference between the fuzzy diagnosis result and the actual result exceeds the pre-set threshold, it is essential to optimize the fuzzy rules based on the optimization

algorithms, and then circulate the whole process in turn until the optimal result of fault diagnosis is determined.

4.2. Fuzzy Theory in DGA-Based Transformer Fault Diagnosis: A Survey

Fuzzy theory, as a mathematical tool for accurately describing uncertainty relations, has unique advantages in the field of transformer fault diagnosis. At present, the research results in this area are rich. The current fuzzy diagnosis method is mainly focused on the following two research directions:

The first one is to introduce the functions of self-organizing and self-learning in simple fuzzy technology. For example, in view of the problem that traditional three-ratio and four-ratio methods have some defects in coding interval, Ma et al. [142] employed the fuzzy correlation matrix to determine the relationship between DGA and fault types, by implementing a fuzzification of the coding. In addition, the system identification method is used to optimize the parameters of the fuzzy correlation matrix, thus achieving good diagnostic effect.

The second one is to integrate the fuzzy diagnosis technique with other intelligent techniques to form hybrid fault diagnosis techniques, such as evolutionary fuzzy logic [52], grey relational fuzzy diagnosis algorithm [141], fuzzy Petri Nets knowledge representation algorithm [143], integrated neural fuzzy algorithm [55–57], FWNN [58], rough set based fuzzy diagnosis [58,144], fuzzy clustering algorithm [145–147], fuzzy C-means algorithm [148,149], and probabilistic fuzzy diagnosis algorithm [150–152]. For this research direction, a couple of examples are given as follows:

For the evolutionary fuzzy logic, Huang et al. [52] proposed an evolutionary programming-based fuzzy system development technique to identify the incipient faults of power transformers. They first built a preliminary framework of the fuzzy diagnosis system, and then employed the proposed evolutionary programming-based development technique to automatically modify the fuzzy if-then rules and simultaneously adjust the corresponding membership functions. In comparison to results of the conventional DGA and the ANN classification methods, the proposed method shows superior performance both in developing the diagnosis system and in identifying the practical transformer fault cases. Islam et al. [53] adopted a novel fuzzy logic approach to develop a computer based intelligent interpretation of transformer faults using VB and C/sup ++/programming. This proposed fuzzy logic based software is tested and tuned using over 800 DGA case histories. It is also utilized in detection and verification of 20 transformer faults and the results show that this proposed diagnostic tool is very useful to both expert and novice engineers in DGA result interpretation. In addition, Aghaei et al. [153] used three fuzzy methods for specifying the internal faults of transformer through the ratio method of oil-immersed gases. The results show that the proposed methods are effective enough in the diagnosis of transformers internal faults.

For the grey relational fuzzy diagnosis, Li et al. [141] adopted fuzzy clustering analysis method to acquire c kinds of cluster centres, in order to make up a standard chart for transformer fault diagnosis. On this basis, the grey incidence analysis theory was used to compute the incidence order of diagnosing pattern with the standard pattern. This method is proposed based on the combination of grey incidence analysis and fuzzy cluster. The tests show that its diagnosis accuracy is higher than other traditional methods. Besides, a concentration prediction model of dissolved gases in transformer oil based on grey relational analysis (GRA) and fuzzy SVM is proposed in [154]. In this method, the GRA is first used to extract key factors that have great influence on characteristic gases concentration. Then the fuzzy membership function is introduced to combine fuzzy mathematics with SVM. Here, each input sample is assigned to different weights according to its sampling time, which reflects the later data had a greater impact on the following prediction results than the earlier data. The result of an actual case proves that the proposed model can improve prediction precision and overcome drawbacks of traditional SVM and the shortcoming of considering only one or all characteristic gases method.

For the fuzzy Petri nets (FPN) knowledge representation in transformer fault diagnosis, Wang and Ji [143] proposed a method of FPN knowledge representation and its rigorous inference

algorithm. In this model, FPN is applied in transformer for the first time and it represents relations between fault symptoms and faults. This FPN is very simple and clear because it only uses simple matrix calculation based on Petri nets theory, thus fast and accurate results can be obtained. The case indicates the model is correct and can provide a new tool for fast fault diagnosis of the power transformer.

The integrated neural fuzzy algorithms has been widely adopted by researchers and engineers in theoretical research and practical application. Fan et al. [55] proposed a hybrid method which combines the relevance vector machine and the adaptive neural fuzzy inference system to address the misdiagnosis of conventional methods that is caused by ambiguous characteristic of some of the record data for the analysis. This algorithm can achieve an accuracy rate as high as 95% and exceeds single adaptive neural fuzzy inference system, SVM, and ANN in distinguishing multiple faults and samples with ambiguous characteristic. Analogously, a transformer fault diagnosis method based on neural network and fuzzy theory has been proposed in [56]. In [57], Naresh et al. presented a new and efficient integrated neural fuzzy approach for transformer fault diagnosis using DGA. This proposed approach first formulates the modelling problem of higher dimensions into lower dimensions and then uses the designed fuzzy rule base for the identification of fault. The approach has been tested on standard and practical data and it shows superior performance in identifying the transformer fault type. Besides, a transformer DGA diagnosis EPS based on neural network and fuzzy theory was developed in [98], which is called blackboard EPS. This system can use fuzzy theory to solve the problems of complexity, empiricism and fuzziness in transformer fault diagnosis, as well as can use the good pattern classification ability and self-learning ability of neural network to improve the accuracy of fault diagnosis of the whole system. The blackboard model structure of this system in [98] is shown in Figure 3.
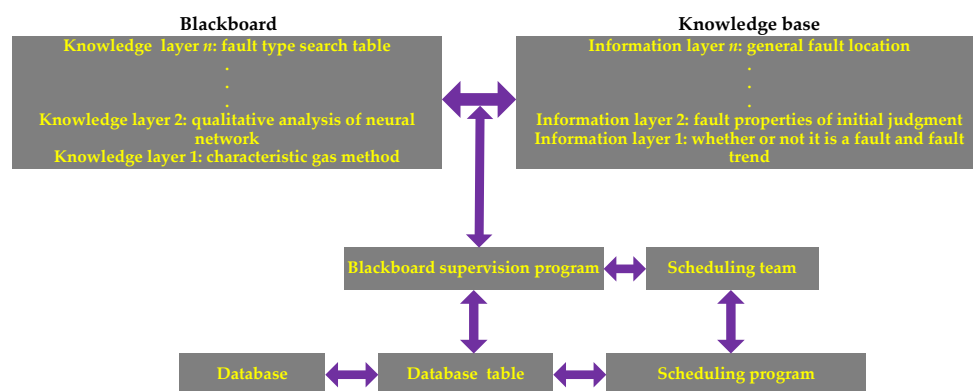
Figure 3. The blackboard model structure for transformer insulation fault diagnosis.

For the FWNN, Dong et al. [124] integrated a rough set and FWNN with a least squares weighted fusion algorithm-based fault diagnosis for power transformers using DGA. The rough set is used as a front end of the FWNN, which is integrated with least square weighted fusion algorithm to simplify the input of FWNN and mine the rules whose confidence and support satisfy some pre-set criteria. In the model, the diagnosis accuracy cannot be limited by the neural network hidden layer number and correlated training parameter. By using the FWNN, this mechanism has good classified diagnosis ability.

For the fuzzy clustering algorithm, an integrated grey clustering and fuzzy clustering fault diagnosis method is proposed in [145], based on which, a weighted fuzzy clustering algorithm has been applied in fault diagnosis of power transformers in [146]. In [146], the method of normalization and promotion compression has been proposed for the components and the component ratios of various characteristic gases. Besides, the attribute weights are utilized to express the relative degree of the importance of various data in fault partitioning, and the weighted fuzzy clustering algorithm is designed to accomplish fuzzy clustering and the calculation and optimization of clustering prototype and attribute weights. Moreover, in order to achieve an

accurate diagnosis by DGA without experienced experts, a novel diagnosis method using fuzzy clustering and a RBF neural network (RBFNN) is proposed in [147]. In this neural network, fuzzy clustering is effective for selecting the efficient training data and reducing learning process time. After conducting the fuzzy clustering, based on which, the RBFNN is used to analyze and diagnose the state of the transformer. Various experiments show that this proposed method has good performance and validity in transformer fault diagnosis based on DGA.

For the application of fuzzy C-means algorithm, Fu et al. [155] aimed at the collected 195 sets of fault samples, and used fuzzy clustering algorithm and fuzzy C-means algorithm for fault diagnosis respectively, with accuracies of 80% and 91.3%, respectively. This research shows that different diagnostic techniques have a great difference in the effect of diagnosis. Besides, an improved fuzzy C-means clustering algorithm for transformer fault has been proposed in [148], and a cross-correlation-aided fuzzy C-means for classification of dynamic faults in transformer winding during impulse testing is proposed in [149].

For the application of probabilistic fuzzy diagnosis algorithm, Duan et al. [150] developed a probabilistic neural network for fault diagnosis of transformer based on fuzzy input, and Yang et al. [151] applied the probability reasoning and fuzzy technique for identifying power transformer malfunction. Besides, in order to overcome the complexity of electric power transformer fault, Fu et al. [152] proposed an improved fault diagnosis model based on the research theories of electric power transformer fault diagnosis by predecessors. This model is developed based on fuzzy theory and probability reasoning, which not only can use the DGA and electric tests data, but also takes other observed information into account. The probability reasoning and parsimonious covering theory here are used to rebuild the relative probability function. The application of this model shows that it can identify the fault characteristic correctly even with some symptoms absent.

Although the fuzzy diagnosis technique can be employed to diagnose the DGA-based transformer faults by using fuzzy membership functions, fuzzy relation equations and fuzzy clustering analyses, etc., it still has some limitations due to the existence of ambiguous relationships between the transformer fault phenomena, fault causes, fault mechanisms and fault types. For example, the sample data is required to be complete in the fuzzy rule table, and the fuzzy membership function is difficult to be determined accurately. Hence, these factors have indirectly affected the comprehensiveness of the diagnosis results. In the future, for the fuzzy theory-based transformer fault diagnosis using DGA, more and more researchers will focus on the combination of fuzzy theory with other intelligent diagnosis tools, such as ANN, Petri nets, WNN, DL, RL, GST, fuzzy clustering algorithm, fuzzy C-means algorithm, SI algorithm, evolutionary algorithm, SVM, and probabilistic fuzzy diagnosis algorithm. A summary for the application of fuzzy theory in DGA based transformer fault diagnosis is presented in Table 7.

Table 7. A summary for the application of fuzzy theory in DGA based transformer fault diagnosis.

Advantages and Disadvantages	Working Process	Primary Means
▪ can well solve the issues with fuzziness and uncertainty ▪ hard to accurately determine the fuzzy membership function ▪ the relationship between the transformer faults phenomena, fault causes, fault mechanisms and fault types is ambiguous ▪ the sample data is required to be complete in the fuzzy rule table	▪ first establish a DGA-based transformer fault database to formulate fuzzy rules ▪ then input the DGA data to conduct fuzzification, fuzzy processing and defuzzification ▪ then optimize the fuzzy rules based on the optimization algorithms ▪ finally repeat calculation in turn until optimal result is determined	▪ employ fuzzy correlation matrix [142] ▪ combined with evolutionary fuzzy logic [52] ▪ combined with grey relational fuzzy diagnosis algorithm [141] ▪ combined with Petri Nets knowledge representation algorithm [143] ▪ combined with integrated neural fuzzy algorithm [55–57] ▪ combined with FWNN [58,124] ▪ combined with rough set [58,144] ▪ fuzzy clustering algorithm [145–147] ▪ fuzzy C-means algorithm [148,149,155] ▪ probabilistic fuzzy diagnosis algorithm [150–152] ▪ combined with expert system [98] ▪ combined with DL, RL, and other ML methods

5. Application of RST in DGA-Based Transformer Fault Diagnosis

5.1. Rough Sets Theory Description

The concept of rough sets has been used by more and more experts and scholars in transformer fault diagnosis. Moreover, the combination of RST with other intelligent means has been widely adopted in transformer fault diagnosis. The RST-based attribute reduction can ensure the selection of fewest characteristic sets with consistent diagnostic results of transformer faults, thus it provides a novel direction for fuzzy theory based transformer fault diagnosis [144]. The RST is an effective mathematical tool to deal with the fuzzy and uncertain knowledge because it does not need to provide any prior information beyond the data needed for the problem, thus it can be used for direct analysis and reasoning of data to find out the hidden knowledge and reveal the potential rules from the data. This is why RST has been widely used in transformer fault diagnosis, especially for the integrated intelligent approaches.

The rough sets can be defined as follows [144]. A four-element group $S = (U, A, V, f)$ is defined as an information system formally, among which $U = \{x_1, x_2, \ldots, x_n\}$; $A = \{a_1, a_2, \ldots, a_m\}$, represents non-empty finite set of attributes; $V = \bigcup\limits_{a \in A} Va$, representing value sets of attributes, here Va denotes the value domain of the attributes $a \in A$; f is the information function, and $f: U \times A \rightarrow V$, means it gives 1 information value for each attribute of each object, namely $\forall a \in A$, $x \in U$, $f(x, a) \in Va$. If $A = C \cup D$, $C \cap D = \emptyset$, here C denotes the condition attribute set, and D represents the decision attribute set, thus such type of information system is also called decision-making system. The relation between the attribute and value described above can form a two-dimensional condition-action table, called decision table.

Note that not all the condition attributes in the original decision table are necessary, and may some of them are unnecessary and can be removed without affecting the original decision-making results. Hence, for the knowledge representation using RST, the decision table after attribute reduction is an incomplete table which only contains necessary condition attributes used in decision-making, while these condition attributes possess all the knowledge of the original knowledge system. As illustrated in [144,156], the flow of fault diagnosis based on the RST is shown in Figure 4.
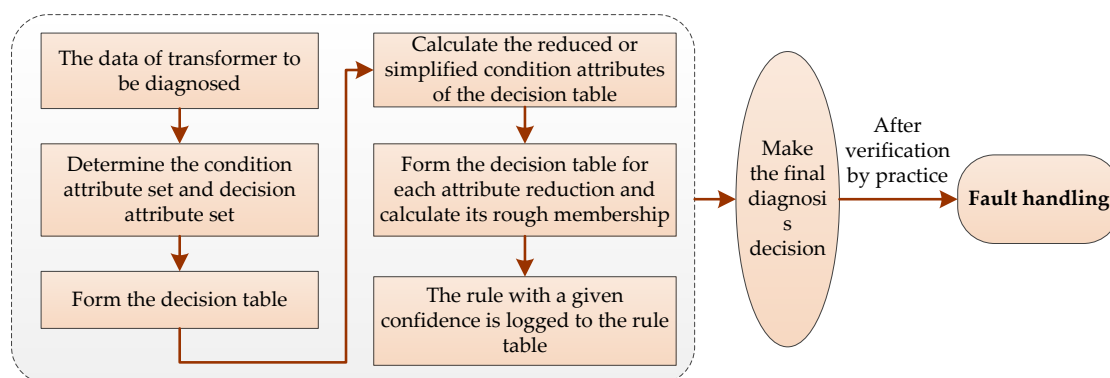
Figure 4. The flow of fault diagnosis based on the RST presented in [144,156].

5.2. Rough Sets Theory in DGA-Based Transformer Fault Diagnosis: A Survey

In current investigations, there are two major methodologies for transformer fault diagnosis using the RST [144]: the first one is fault diagnosis based on single RST, and the second is based on integration of the RST with other intelligent methods. The two categories of transformer fault diagnosis methods are summarized as follows.

(1) Single RST-based fault diagnosis

In this methodology, firstly, the fault symptoms are measured as the condition attribute of fault classification, and the actual existing faults are treated as decision attributes to establish a decision table. Then the attribute reduction ability of RST is used to simplify the original decision table to obtain multiple attribute reductions which are equivalent to the original decision table. Lastly, these attribute reductions are made further simplification operations in order to remove the unnecessary attributes, such that the fault diagnosis rules can be achieved. On this basis, Su et al. [157] developed a fault diagnosis model based on RST and information entropy, which increasingly accelerates the computation time of diagnosis. Yuan et al. [158] proposed a diagnosis model of transformer faults based on a new heuristic reduction algorithm using RST, in which the complexity is decreased obviously comparing with general attribute reduction approaches, and the computation time is shortened, such that the diagnostic efficiency is improved. In the case with high density data, this proposed model can still provide a faster computation speed with higher judgment accuracy, thus it highly enhances the computation of rough sets.

(2)   Integration of RST with other intelligent algorithms

The first direction is to integrate RST with EPS, which is generally focused on establishment of complete knowledge base in EPS-based transformer fault diagnosis system. Xiang [159] proposed a fault diagnosis EPS based on RST, which integrates RST with EPS. Based on the attribute reduction of the decision table formed by the historical fault data of the transformer, the knowledge base of node network rule set that meets the requirement of confidence level is established with different reductive levels, by calculating rough membership of the rule, thus it is able to achieve accurate diagnosis results with some fault-tolerant ability, even if the gas chromatograph analysis data is incomplete. In addition, Zuo [160] proposed a new intelligent fault diagnosis method based on RST and EPS, in order to improve diagnosis precision and decrease misinformation diagnosis, according to the intelligence complementary strategy. In this model, RST is employed to handle inexact and uncertain knowledge for pattern recognition with the target of removing redundant information and seeking for reduced decision tables, so as to obtain the minimum fault feature subset. Besides, EPS here with an independent knowledge base is used to make knowledge maintenance more convenient and have easy reasoning process to explain.

The second is integration with ANN, in which the abilities of ANN such as non-linear feature, parallel processing and self-organizing and self-learning can be perfectly employed. On this basis, Yu et al. [161] first used RST to conduct attribute reduction for the original sample sets to form reduced rule sets, hence rough set network is treated as front-end system and then the sample sets after attribute reduction by RST are conducted as input sample sets of the ANN to form a rough set and ANN-based transformer fault diagnosis system. Zhang et al. [162] proposed to firstly use DGA knowledge-based continuous attributes to discretize some attributes in the decision table and at the same time use natural algorithm and partition with same frequency to discretize some other attributes. After that, RST is used to reduce the attributed of the discretized data. Lastly, the obtained minimum decision table is used to train the error BF algorithm-based neural network. Besides, Li et al. [163] proposed a new power transformer fault diagnosis method based on RST and an improved artificial immune network classification algorithm, which can achieve the minimal diagnostic rules via simplifying expert knowledge and reducing fault symptoms, learning the features of fault samples, and obtaining the memory antibody cells pool with capability of representing the fault samples better than those without class information. This proposed model has better capability to classify single-fault and multiple-fault samples as well as higher diagnosis precision, by comparing with the IEC three-ratio method and BP neural network.

The third is integration with fuzzy sets theory. To this end, Xiong et al. [164] presented a new diagnosis measure with the gas ratios method for transformer incipient faults. In the diagnosis process, an information decision system has been built in which a data-mining algorithm is developed to extract fuzzy rough rules and thus determine the topology of multi-table decision base according to the attributes set. This proposed diagnosis system using the actual dissolved gases in transformer oil confirms that the extracted rules allow diagnosis results to be satisfied with a satisfactory accuracy for diagnosis ratio. However, the single RST for transformer fault diagnosis

needs a high requirement on precision of experiment data samples, besides the conventional RST cannot be employed to address continuous attributes, such that it generally needs to discretize the data samples. Hence, in actual application, more and more researchers choose to integrate the RST with other intelligent algorithms when using it for transformer fault diagnosis. Besides, Wang [165] also proposed a fault diagnosis method for power transformers based on rough set and fuzzy rules, which can realize effective fuzzy reasoning and obtain accurate diagnosis results.

The fourth is integration with BN, namely a Bayesian Network. In this direction, the BN and RST can be both employed to process the incomplete data. However, the direct utilization of the two cannot be satisfied with the actual demands of fault diagnosis due to the lower judgment accuracy when the key attributes are missing. To address this, Wang et al. [166] integrated the BN classifier with RST organically and applied it to fault diagnosis of a transformer, by developing a comprehensive transformer fault diagnosis model combining dissolved gas in transformer oil and other electrical testing data samples. The basic idea of the model is to use the attribute reduction technique in RST to achieve reduction of the expert knowledge and diminution of the fault features, such that the minimal diagnosis rule and inputting it to the BN to reduce the complexity of the network structure as well as the difficulty of acquiring fault features. Moreover, Wang et al. [167] proposed a new transformer fault diagnosis based on RST and BN, in which the expert knowledge can be simplified as well as fault symptoms can be reduced through the reduction approach of RST information table, and the diagnostic rules can be mined. Besides, the BN can realize probability reasoning to describe changes of fault symptoms and analyse fault reasons of the transformer. This proposed method shows correctness and effectiveness in some practical fault diagnosis examples. Furthermore, Xie et al. [168] combined the BN classifier and rough set reduction theory together in order to establish a BN classification model based on expert knowledge and statistical data, in which the DGA data and electrical tests are integrated as the input set of diagnosis, and the probabilistic reasoning and sequencing of potential fault types are actualized, such that improving the reliability of the diagnosis. This proposed method is capable of dealing with missing information and shows a better fault-tolerant feature and can achieve high accuracy.

The fifth is combination with SVM, in which the SVM can be employed to better address the issues of small sample learning and has been research highlights in the field of ML internationally. The combination of the two can fully take advantages of the SVM in aspect of accurate binary class classification as well as the RST in aspect of dealing with small complete information and rapid diagnosis. On this basis, Jiang and Ni [169] proposed a transformer fault diagnosis method based on the combination of rough sets and SVM, in which the rough sets are employed to establish the decision table and the rough set theory is applied to simplify the expert knowledge to obtain the diagnosis rules with attribute reduction and implement rough diagnosis for the transformer, and then the SVM is adopted to conduct accurate fault diagnosis with the function of accurate binary class classification. Wu et al. [170] employed the rough sets and SVM to build a model for the location of the transformer fault diagnosis. In this model, the results of the oil data and the electrical experimental data are first combined and reduced based on rough set theory, in order to establish the mapping of the faults and the information. Then, this mapping is classified by the SVM classifier, thus the rough faulty point of the transformer can be diagnosed. This proposed model shows a satisfactory accuracy of obtaining rough faulty point of the transformer.

The last direction is combination with a Petri network. In this application, the RST is generally employed to obtain the minimal diagnosis rule based on its stronger data analysis ability, compression capability and fault-tolerant, in order to establish the optimal Petri network model whose parallel reasoning ability is used for more effective transformer fault diagnosis. Wang et al. [171] developed a model to improve the efficiency of intelligent approaches based on prior knowledge, in which the RST is employed to reduce the many redundant features in the transformer fault diagnosis rules and the optimal Petri nets are built to realize fast and parallel reasoning. This developed model shows an invariable fault classification after reduction and that the main features are close to actual experiences. Besides, Wang et al. [172] integrated the RST and fuzzy Petri nets for synthetic fault diagnosis of oil-immersed power transformers, based on

complementary strategy. According to the minimal rules which are mined through reduction approach of RST information table, the complexity of fuzzy Petri nets structure and difficulties of fault symptom acquisition are largely lessened. Meanwhile, the fuzzy Petri nets are employed to describe changes of fault symptoms and analyse operating status of the transformer based on its parallel and fuzzy reasoning capability. This proposed method shows correctness and effectiveness in the practical fault examples.

Besides the directions summarized above, some researchers have also combined the RST with other intelligent tools for transformer fault diagnosis, for example, Zhou et al. [173] integrated the RST and evidence theory for transformer fault diagnosis, in which the rough set is induced to calculate the importance degree of condition attribute to decision attribute and act as basic probability assignment of recognition framework. In the same recognition framework, different evidence is combined to obtain information on the fault types of decision classification information. This proposed method can effectively improve the single fault diagnosis accuracy and also give information about compound fault analysis. In addition, Shu et al. [174] brought Extenics and RST into fault diagnosis of the transformer, in which the attribute predigesting method in RST is employed to classify the attribute term which needed by each fault diagnosis. In this method, the DGA testing datum is used to be attribute set and the standard fault model of the transformer is used to be the decision set for diagnosis. Besides, the association function from Extenics is utilized to count each fault degree. This method has been applied to diagnose 76 DGA testing data and it shows better diagnosis results than the IEC method. It is indicated that the RST can be combined with grey theory for fault prediction of power transformer, based on which, Fei and Sun [175] proposed a new method for transformer fault prediction, in which the improved three-ratio attribute decision table is constructed and simplified by the knowledge acquisition method based on rough sets, and the ratios of feature gases can be predicted by GM and their future state feature can be obtained. According to the minimal rules, the incipient fault can be predicted, and its probability can be acquired by combination rules' credibility with the number of the fault acquired from predicted feature of gases' ratios. The testing results show that this method is effective and correct in fault prediction examples. In addition, Song et al. [176] established an immune model for transformer fault diagnosis by combining the strong ability of recognition and learning in the artificial immune system with the attribute's objectively reduction of the RST together. Results show that this developed model has high diagnosis accuracy, strong robustness and good learning ability.

In the future, for the RST-based transformer fault diagnosis using DGA, more and more researchers will focus on the combination of it with other intelligent diagnosis tools, such as ANN, Petri nets, WNN, DL, GST, fuzzy clustering algorithm, fuzzy C-means algorithm, SI algorithm, SVM, and probabilistic fuzzy diagnosis algorithm. Especially for the combination of RST and ML algorithms and this may be aimed at the following aspects: the analysis of the cause of fault, the characteristic gases generation mechanism based fault diagnosis, and the exploration of new diagnosis approaches and strategies. This will be a new breakthrough in fault diagnosis techniques of the oil-immersed power transformer based on DGA. A summary of the applications of RST in DGA-based transformer fault diagnosis is presented in Table 8.

Table 8. A summary for the application of RST in DGA-based transformer fault diagnosis.

Advantages and Disadvantages	Categories	Primary Means
▪ effective in dealing with fuzzy and uncertain knowledge ▪ does not need to provide any prior information beyond the data needed for the problem ▪ direct analysis and reasoning of the data samples ▪ can effectively find the hidden knowledge and reveal potential rules from the data	▪ single RST based fault diagnosis ▪ integration of RST with other intelligent algorithms	▪ combined with information entropy [157] ▪ combined with new heuristic reduction algorithm [158] ▪ integrated with expert system [159,160] ▪ integrated with ANN [161–163] ▪ integrated with fuzzy set theory [164,165] ▪ integrated with Bayesian network [166–168] ▪ integrated with SVM [169,170] ▪ integrated with Petri network [171,172] ▪ integrated with evidence theory [173] ▪ integrated with attribute predigesting method [174] ▪ combined with improved three-ratio attribute decision [175]

▪ combined with artificial immune system [176]
▪ combined with DL, fuzzy C-means algorithm, SI algorithms, probabilistic fuzzy theory

6. Application of GST in DGA-Based Transformer Fault Diagnosis

6.1. Grey System Description

Grey system theory (GST) was proposed by Deng in 1982 [177,178] and has been developed rapidly [179–183] since that time. GST is a method to study the issues with features of less data, poor information and uncertainty. This method is a theoretical result developed on the basis of the practice of fuzzy mathematics. This theory after years of research and development has formed the analysis system relied on grey relational space, the method system based on grey sequence generation, the model system with GM as the core, and the technical system with the system analysis, evaluation, modelling, prediction and decision-making as the principal parts [184].

In GST, the small samples with some known information and some unknown information, as well as the uncertain systems with poor information are treated as research objectives, and their valuable information is extracted mainly through the generation and development of the known information part of the research object, such that realizing correct description and effective control of the operation behaviour and evolution rule of the system [185–187]. In the field of engineering, the depth of colour is generally adopted to describe the clarity degree of information. For example, the black box is used to describe a system or object whose internal information is completely unknown. Hence, in GST, black is used to express the meaning of the information completely unknown, white to express the information completely known, and grey to express that part of the information is clear and part of the information is unknown. Correspondingly, the system with unknown information is called black system, the system with completely known information is called white system, and the system with partial known information and partial unknown information is called grey system [177–179].

For the research objective in this paper, namely the oil-immersed power transformer, its fault diagnosis system can be seen as a typical grey system, due to the fact the relationships between some fault causes and fault results in the transformer fault diagnosis system are not well-defined, as well as it cannot clearly determine which kinds of gases dissolved in oil cause even when a fault occurs [187]. Consequently, the GST model as an effective tool is with the characters of less data, high precision and without prior information, which has been widely used in transformer fault diagnosis based on DGA. As defined in [177,178], the system that only masters or can only obtain part of the control information is called a grey control system, or grey system for short. Accordingly, the matrix with some known mathematical properties as well as some known elements is called grey matrix, and the parameters that have some known mathematical properties while its concrete values are unknown are called grey parameters. Hence, as first defined in [178], the grey matrix A is given as:

$$A \begin{cases} A = \begin{bmatrix} & \vdots & \\ \cdots & a_{ij} & \cdots \\ & \vdots & \\ \cdots & \otimes_{ij} & \cdots \\ & \vdots & \end{bmatrix} \begin{pmatrix} i = 1,2,\cdots,n \\ j = 1,2,\cdots,m \end{pmatrix}, n,m \in W \bigcap R^1 \\ S_G = \left\{ \otimes_{ij} \right\} \\ S_W = \left\{ a_{ij} \right\} \\ S = S_G \bigcup S_W \\ S_G = S \setminus S_W \end{cases} \tag{5}$$

where A is the grey matrix, denoted by $A = \left(\otimes_{ij} \right)$, here the general grey parameters are denoted by $\otimes_{ij}$, namely $\otimes_{ij} =$ unknown, $\otimes_{ij} \in R^1$. R^1 is the set of all real numbers. The zero operation of grey

parameters is defined as $\otimes \cdot 0 = \otimes$, $\otimes \pm 0 = \otimes$. Apart from zero operation, the results of the four fundamental operations of arithmetic between grey parameters, as well as between grey parameters band white parameters are still grey parameters. G generally represents the grey area, system, concept, matrix, number, control law, etc. Accordingly, W is the general symbol for the white. S is the element set of the matrix A. S_G and S_W are the grey parameter set and white parameter set in A, respectively. Based on (5), the following system is called grey linear system in [178], denoted by G_L as:

$$\begin{cases} \dot{X} = AX + BU, \quad A = \left(\otimes_{ij} \right), B = \left(\otimes_{pq} \right) \\ Y = CX, \, C = \left(\otimes_{kl} \right) \\ \left(a_{ij}, b_{pq}, c_{kl} \right) \in W \end{cases} \tag{6}$$

Based on the definition given in [178], the GST model mainly includes grey analysis model, grey clustering evaluation model, GM series model, grey decision model, grey combination model, grey system prediction model and GRA model. Among them, the most important theory in the grey decision model is the weighted grey target theory. Recently, the weighted grey target theory and GRA model have been extensively adopted in fault diagnosis of the transformer [59–66,185–187]. The two are briefly introduced as follows:

(1) Weighted grey theory. The substance of weighted grey theory is to set a grey target under the condition of no standard mode, and then find the bull's eye in the grey target through the grey target theory. Next, the models of indexes are compared with the standard model, and finally the models of these indexes are implemented grade division to determine the evaluation grade [59]. As the author already studied in [59], the approaching degree (i.e., the grey-correlation degree) $\gamma(x_0, x_j)$ and the weight value q_i can be calculated as:

$$\gamma(x_0, x_j) = \frac{1}{n} \sum_{k=1}^{n} \gamma(x_0(k), x_j(k)) \tag{7}$$

$$q_i = \begin{cases} \dfrac{1}{n}(1 + \dfrac{|\gamma_{\text{mea}} - \gamma(i)|}{\gamma_{\text{mea}}}), \, \text{if } \gamma(i) > \gamma_{\text{mea}} \\ \dfrac{1}{n}(1 - \dfrac{|\gamma_{\text{mea}} - \gamma(i)|}{\gamma_{\text{mea}}}), \, \text{if } \gamma(i) < \gamma_{\text{mea}} \end{cases}, \quad \text{where } \sum_{i=1}^{n} q_i = 1 \tag{8}$$

where $\gamma(x_0, x_j)$ represents the degree of the bull's-eye of a mode close to that of the standard state mode, called the approaching degree. $k = 1, 2, \cdots, n$, and n is the sum total of the index modes and the indexes. i is the number of the index. γ_{mea} is the average value of the contribution degree of all the indexes, thus the weight that corresponds to γ_{mea} should be $1/n$.

Based on (7) and (8), finally, the weighted approaching coefficient $\gamma(\omega_0(k), \omega_i(k))$ and the weighted approaching degree of ω_i, namely $\gamma(\omega_0, \omega_i)$ can be obtained as [59]:

$$\gamma\left(\omega_0(k), \omega_i(k)\right) = \frac{\min\limits_{i} \min\limits_{k} \Delta_{0i}(k) + \rho \max\limits_{i} \max\limits_{k} \Delta_{0i}(k)}{\Delta_{0i}(k) + \rho \max\limits_{i} \max\limits_{k} \Delta_{0i}(k)} \tag{9}$$

$$\gamma(\omega_0, \omega_i) = \sum_{k=1}^{n} \gamma\left(\omega_0(k), \omega_i(k)\right) q_k \tag{10}$$

where $\rho \in [0,1]$. $\Delta_{0i}(k)$ represents the grey-correlation difference information between the sequence to be evaluated ω_i and the bull's-eye ω_0. $\gamma(\omega_0, \omega_i)$ means the degree of each mode close to the standard state mode.

According to (7)–(10), the weighted grey theory can used to carry out pattern recognition, pattern clarification and pattern optimal selection. The evaluation flow of weighted grey theory is shown in Figure 5. In Figure 5, the data of DGA is made as the state evaluation parameters to

conduct evaluation on the internal oil-paper insulation system of the transformer, so as to achieve the grade of operation status of the transformer [59], as shown in Figure 6.
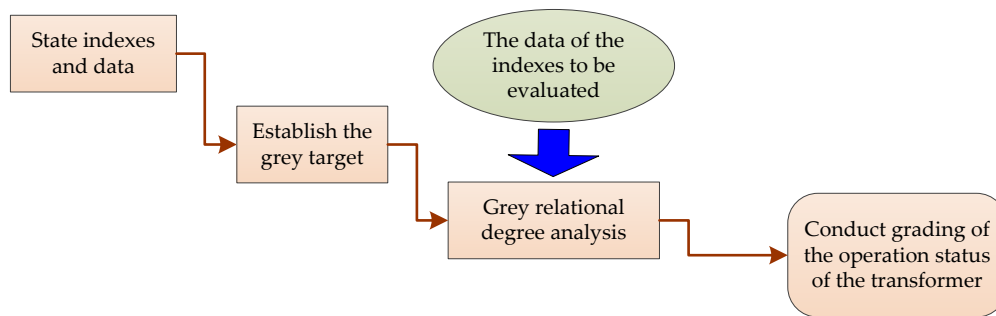
Figure 5. The evaluation flow of weighted grey theory.
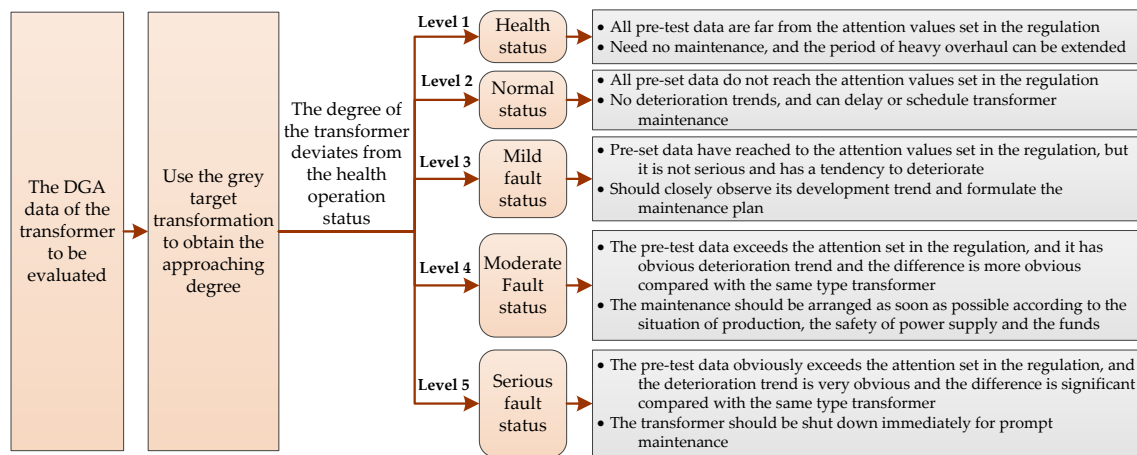
Figure 6. The grading of operation status of the transformer.

(2) GRA model. GRA is a kind of analysis method which is based on the GST. The basic idea of GRA is to determine the degree of correlation between the factors according to the similarity degree of the geometric shape of their variation curves. Through quantitative analysis of the development trend of dynamic processes, this method can achieve the comparison of geometric relations of statistical data related to time series, so as to find out the grey relational degree among all factors [154,188]. As elaborated in [154], the grey relational degree is introduced to measure the affinity among the factors, in order to obtain the main factors affecting the concentration of each kind of characteristic gas. This is because no definite qualitative and quantitative description for the relation between the content of gas dissolved in transformer oil, oil temperature and load can be found, and uncertainty exists in the mutual restriction relation between the gases. Hence, according to [154], assume that the reference array is $X_0 = \{X_0(k)\,|\,k = 1, 2, \ldots, n\}$, and the comparative array is $X_i = \{X_i(k)\,|\,k = 1, 2, \ldots, n\}$, where $i = 1, 2, \ldots, n$. Firstly, the original data are made being dimensionless as:

$$x_i(k) = \frac{X_i(k)}{X_i(1)} \tag{11}$$

where $k = 1, 2, \ldots, n$ and $i = 1, 2, \ldots, n$.

Then, the grey relational coefficient [154] can be calculated as:

$$\xi_i(k) = \frac{\min\limits_{i}\min\limits_{k}|x_0(k) - x_i(k)| + \rho\max\limits_{i}\max\limits_{k}|x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho\max\limits_{i}\max\limits_{k}|x_0(k) - x_i(k)|} \tag{12}$$

where $\xi_i(k)$ is the grey relational coefficient between $x_0(k)$ and $x_i(k)$, which reflects the tightness of the two sequence at a certain time. The constant ρ is the discrimination coefficient, with a value

range of (0, 1), and generally the smaller the ρ, the larger the discrimination. In order to improve the difference between the relational coefficients, the ρ is taken 0.5 in [154]. Based on the grey relational coefficient of each point, the grey relational degree between X_i and X_0 is obtained [154] as:

$$\gamma_i = \frac{1}{n}\sum_{k=1}^{n}\xi_i(k) \tag{13}$$

where $k = 1, 2, …, n$ and $i = 1, 2, …, m$.

The expression of the grey relational degree γ_i reflects the degree of correlation between X_i and X_0, and it shows that the larger the γ_i, the higher the degree of correlation between the two, thus the closer the relationship between them and the closer the development trend and speed of them.

6.2. Grey System Theory in DGA-Based Transformer Fault Diagnosis: A Survey

As previously stated, the fault diagnosis system of the transformer can be considered as a typical grey system. Here, the GRA of fault of the transformer is performed by employing the grey theory to identify and classify the symptom pattern of the faults as well as the fault modes. Hence, according to [59,154], the procedures of GRA can be described as follows:

(a) First, construct a comparative sequence based on the inputs of the data of DGA.
(b) Next, use the GRA method to calculate the grey correlation between the comparative sequence and the reference sequence.
(c) Lastly, according to the calculated grey correlation, the principle to be followed is that the larger the grey correlation, the closer the actual fault mode to the reference fault mode is.

Based on the procedures above, the application of GRA in transformer fault diagnosis has presented a lot of research achievements in recent years. Li et al. [185] used GST to analyse transformer insulation fault, in which a grey cluster model and the relevant model are developed for insulation fault diagnosis. This proposed method has been successfully applied in some fault examples using oil-chromatogram data of five transformers, which shows that it is valid to analyse fault pattern and locate the fault position with a good prospect of wide application. On this basis, the transformer fault diagnosis method based on grey correlation entropy is proposed in [189], which has been verified feasible and effective by an example. Compared with the traditional three-ratio method, this proposed method is better in fault diagnosis under the same conditions. However, the diagnosis result of the method in [189] is susceptible to external disturbances. To address this issue, Li and Zhao [187] proposed a transformer fault diagnosis method based on entropy weight optimization and weighted grey correlation degree, in which five kinds of gases dissolved in oil are made as characteristic parameters to verify that this proposed model is valid and good in fault diagnosis, thus the problem of external interference is solved well. In addition, Li [190] proposed to use the weighted grey target theory to evaluate the operation status of the transformer. In this work, seven groups of fault identification sequence are obtained through statistical analysis of 300 sets of transformer fault data samples. In the 100 sets of normal operation data of the power transformer, the accuracy rate of fault judgment is reached 98%. In the 100 sets of fault data, the accuracy rate is reached 96%.

Besides the research work introduced above, the author in [59] proposed a method which realizes dynamic modelling for reliability assessment of transformer oil-paper insulation systems using hot spot temperature (HST) and grey target theory. This developed model contains a HST-based static ageing failure model and a grey target theory based dynamic correction model, thus it corresponds to two stages: transformer ageing process description and winding HST calculation stage, and life expectancy dynamic modification stage. The combination of the two models can dynamically modify the life expectancy of the transformer using actual data of DGA. The entire dynamic correction process can be seen in Figure 7.
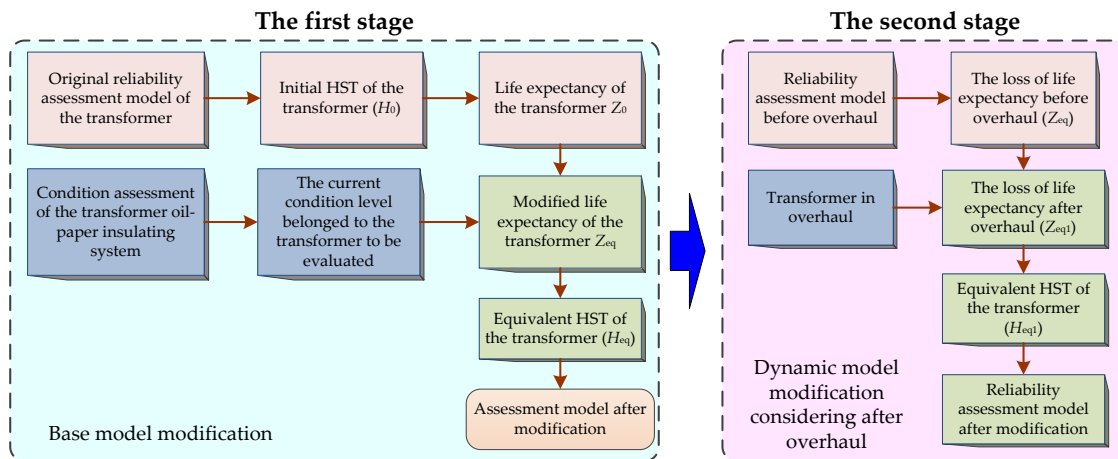
Figure 7. The dynamic modification process of the model developed by the author in [59].

In addition, a concentration prediction model of dissolved gases in transformer oil based on GRA and fuzzy SVM is proposed in [154], which considers the influence of oil temperature and loads on oil-dissolved gases. In this model, the GRA is employed to extract key factors that have great influence on characteristic gases concentration as attributes of input samples of the SVM regression modelling. Besides, the fuzzy membership function is employed to combine fuzzy mathematics and SVM. The result of an actual case shows that this proposed model is effective, and can improve prediction precision and overcome drawbacks of traditional SVM and the shortcoming of considering only one or all characteristic gases method. Dong et al. [60] presented an approach of fault diagnosis based on model-diagnosis for power transformers after analyzing in depth the relationship between the reason and symptom of the fault. In the method, the action and function of the transformer, symptom set and fault set can be established based on the known knowledge, experiences, and collected fault examples. The grey correlation in the model is employed to assist to describe the similarity between the faults and symptoms, such that the diagnosis results in more detail can be achieved. The examples of diagnosis show that the approach is quite efficient, flexible and fault-tolerant. Song et al. [61] presented a new method based on grey relation entropy to address the issue of code deficiency exists in the IEC/IEEE standard (such as ratio code nonentity) and complexity of fault diagnosis for the transformer. This method integrates grey relation analysis and information entropy, which can overcome defects of original grey relation analysis, such as partial relation and information losing. Analogously, Chang et al. [62] proposed a fault diagnosis method for transformer based on the DGA and grey relational theory, which is available for the transformer fault diagnosis and has fault classified ability. Lin et al. [63] proposed a method for dissolved-gases prediction and fault diagnosis in oil-immersed transformers using grey prediction-clustering analysis. In this model, DGA is employed to detect and monitor abnormal conditions in transformer, the grey prediction GM(1, 2) model is used to forecast the further trends of both combustible and non-combustible gases by using the variant information of hydrogen, and the grey clustering analysis is applied for internal faults diagnosis. Tests with field gas records show the model is effective in dissolved gases forecast and fault diagnosis. Song et al. [64] employed the GRA method to diagnose the fault patterns of power transformers, in which a group of reference sequences are selected from fault data and they are analyzed and compared with other methods. The results show that GRA is a useful tool for evaluating the faults of power transformers and the diagnosis method is effective. Aimed at the all gas features of a traction transformer when a fault occurs, Zhao and Li [191] proposed a method based on the improved grey correlation analysis model for fault diagnosis of traction transformers. This method can fully utilize the overall DGA information and can make use of the advantages of grey correlation analysis in dealing with less samples and lean grey information, such that it can avoid the partial correlation and information loss. Examples show that this model can determine fault types of the traction transformer effectively with higher diagnosis accuracy than ever. In addition, in order to solve the problem of

randomness and fuzziness in transformer fault diagnosis, Xu et al. [192] proposed a new fault diagnosis method based on feedback cloud entropy model, in which the collected fault examples of chromatographic data for transformer oil after statistical analysis are put into Bayesian feedback backward cloud generator as cloud drop, and then the parameter values of fault characteristic gases cloud model is employed to build the transformer fault diagnosis standard normal cloud model. This built model has integrated cloud correlation coefficient and information entropy theory, which can reduce the dependence on the single standard normal cloud model and dig more information of the dissolved-gases in oil, such that improving the accuracy of transformer fault diagnosis. Results of example show that the model has well theoretical value and application prospects with a higher accuracy of transformer fault diagnosis. Besides, a unified GRA on transformer DGA fault diagnosis is conducted in [193]; Liu et al. [194] carried out GRA for insulation condition assessment of power transformers based on conventional dielectric response measurement, which can provide reliable and effective insulation diagnosis; Zhou et al. [195] proposed to use GRA and integrated weight determination for timely fault identification, in which the weight of each indicator is determined by integrating analytic hierarchy process and entropy methods. This model can effectively improve the accuracy of fault diagnosis.

Based on the above research summary, the GRA method has been widely applied for DGA-based transformer fault diagnosis and fault identification, which has good accuracy for some faults that are more difficult to be judged, such as dampness. However, GRA for DGA data under normal circumstances sometimes suffers from misjudgment phenomena, and some researchers have pointed out that this may be caused by the diagnostic system input [187], but the specific reasons are not very clear currently. This is also one of the reasons that limits the wide application of GRA in transformer fault diagnosis based on DGA. Hence, as previously stated, many scholars deem that the chromatographic data should be compared with the warning value by the conventional method before utilization of the GRA. If the data shows a fault, then the GRA can be applied for fault judgment and diagnosis. In the future, the development direction of GRA should be focused on its combination with other intelligent diagnosis tools, such as improved SVM, fuzzy theory, cloud entropy model, BN, ML and data mining techniques. A summary for the application of GST in DGA-based transformer fault diagnosis is presented in Table 9.

Table 9. A summary for the application of GST in DGA-based transformer fault diagnosis.

Advantages and Disadvantages	Procedures of GRA	Primary Means
▪ needs less data for fault diagnosis ▪ high precision ▪ without prior information ▪ good at dealing with the small samples with some information known and some information unknown, as well as the uncertain systems with poor information	▪ first construct a comparative sequence ▪ then use GRA to calculate grey correlation between comparative sequence and reference sequence ▪ lastly determine the actual fault mode according to the calculated grey correlation	▪ weighted grey target theory [59–66,185–187,190] ▪ GRA model [154,188] ▪ combined with hot spot temperature [59] ▪ combined with fuzzy SVM [154] ▪ grey relation entropy [61] ▪ grey prediction-clustering analysis [63] ▪ improved grey correlation analysis model [191] ▪ combined with feedback cloud entropy model [192] ▪ combined with improved SVM, cloud entropy model, BN, ML and data mining techniques

7. Application of Other Intelligent Algorithms in DGA-Based Transformer Fault Diagnosis

The DGA-based transformer fault diagnosis system is a complex system in which various uncertain factors and unknown information are remained under cover, causing fuzziness and randomness in addressing these uncertain issues. In addition to the five main categories of research approaches and techniques summarized above, there are some other intelligent algorithms, such as artificial immune algorithm (AIA) [72,163,176], GA [67,68,196], improved artificial fish swarm

optimizer (IAFSO) [197–199], PSO [69,77,80], dynamic clustering (DC) [79,81], WA [83,124–127], SVM [5,68,72,77,80,85,154,169,170,188,199], BN [87,166–168], information fusion technology [200–202], extreme learning machine (ELM) [203–205], DL [70,71,105,206,207], optimized neural network [208,209], and evidential reasoning approach [45,75,151,210–217], that have been consecutively employed by more and more scholars in fault diagnosis and decision making of the transformer based on DGA in recent years. In this section, these intelligent algorithms and techniques are divided into SI algorithms, data mining techniques, ML approaches and other intelligent tools, which are systematically summarized as follows.

7.1. Swarm Intelligence Algorithms

7.1.1. Swarm Intelligence Algorithms Introduction

With the study of biologically inspired computation, the self-organization behaviour of some social animals has aroused the widespread interest of scientists, who have found that the individuals of some social animal species in Nature tend to possess no intelligence and simple behaviour while a swarm of them exhibits strong intelligence with complex behaviour characteristics when they work together, such as birds foraging, fish fleeing, etc. Based on this phenomenon, the SI algorithm was proposed and developed by scholars, which performs excellently in solving complex problems in the aspects of searching and optimization [218]. The basic idea of an SI algorithm is reflected in imitating the population behaviour of the biological species in Nature to construct a stochastic optimization algorithm in which the optimization and search process is simulated as an individual's foraging or evolution process in a population. In this simulated process, the point in the search space is used to imitate the individual of a population in nature and meanwhile the objective function of the issue to be solved is measured as the adaptive ability of the individual to the environment in the population, such that the process of positive natural selection or foraging process is compared to the optimization iteration process of replacing poor feasible solution with better feasible solution in the search process. Hence, a SI algorithm as a type of iterative optimization algorithm represents the collective behaviour of decentralized and self-organized systems, regardless of natural or artificial, with features of generation and test [219]. SI algorithm includes GA, AIA, ant colony optimizer (ACO), PSO, bacterial foraging optimization (BFO), artificial fish swarm optimizer (AFSO), artificial bee colony (ABC), firefly optimization algorithm (FOA), bat optimization algorithm (BOA), etc. These optimization algorithms as a new type of evolutionary algorithm have been successfully applied to the fields of function optimization due to the characters of distribution, self-organization, and strong robustness [219]. Several typical SI algorithms mentioned here are briefly introduced as follows, as well as their possible applications in the DGA-based transformer fault diagnosis and decision making.

7.1.2. Application of SI Algorithms in Transformer Fault Diagnosis

(1) *GA:* it is a randomized search method evolved from imitating of the evolutionary laws of the biosphere [220], which is initially proposed by Holland. The main principle of GA is based on Darwin's concept of biological evolution and Mendel's theory of genetic variability, with the aim of achieving random global search and optimization by imitating the mechanism of biological evolution in nature [221]. The main features of GA are reflected in the following aspects: conduct direct operation to structural objects; have better global optimization ability and a search space that can automatically obtain and guide optimization; the search direction can be adjusted adaptively; there is no need for certain rules. The mathematical model of standard GA (SGA) can be described as: $SGA = (C, E, P_0, N, \Phi, \Gamma, \Psi, T)$, where $C, E, P_0, N, \Phi, \Gamma, \Psi$, and T represent the individual coding method, individual fitness evaluation function, initial population, size of population, selection operator, crossover operator, mutation operator and iterative termination condition of GA, respectively. Based on this, the flow chart of SGA is illustrated in Figure 8.
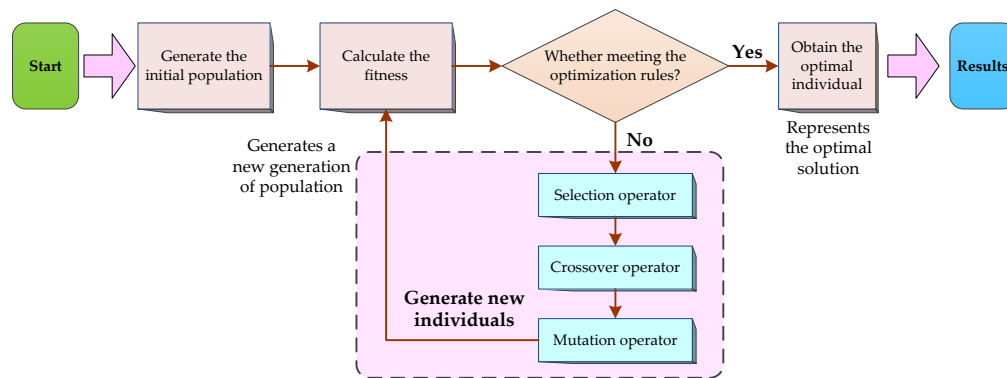
Figure 8. The flow chart of SGA.

According to the principle of GA shown in Figure 8, Pan et al. [67] presented a fault diagnostic method based on a real-encoded hybrid GA evolving a wavelet neural network (WNN), which can be employed to optimize the structure and the parameters of WNN instead of humans in the same training process. This method overcomes some defects of a BP algorithm of WNN, the optimal procedure is easily stacked into the local minima and cases strictly demand initial value, for example, and can achieve a satisfactory compromise among network complexity, convergence and generalization ability. A number of examples are carried out in this model, which show that it has good classification capability for the single- and multiple-fault samples of power transformers as well as high fault diagnostic accuracy. In order to select appropriate SVM parameters, Fei and Zhang [68] proposed a SVM with genetic algorithm-based model for fault diagnosis of a power transformer, in which the GA is used to optimize the parameters of the SVM. This model is employed to test the experimental data from several electric power companies in China and the results indicate that this developed model can achieve higher diagnostic accuracy than IEC three-ratio methods, normal SVM classifier and ANN. Besides, aimed at the inherent disadvantages of BPNN, such as local optimization, over-fitting and difficulties in convergence, Zhang et al. [196] integrated a combination ratio of taking advantages of IEC and Doernenburg into GA and fuzzy C-means clustering algorithm optimized BP, based on which, a novel model has been built successfully and it shows a better diagnosis accuracy rate and generalization ability than other models. In this model, fuzzy C-means clustering algorithm and GA can significantly overcome the disadvantages of data training and BP, thus it offers the potential of implementation for real-time diagnosis systems. Analogously, in order to avoid getting easily trapped into the minimal value locally and strict requirements on the initial value which would make fault diagnosis difficult to some extent, Chen and Yun [222] employed the evolutionary rule of the survival of the fittest to carry out a global optimization search for the transformer fault results which may contain the possible solutions. Finally, the optimal solution is found. The example shows that GA applied in this developed model can effectively prevent the diagnosis results from falling into local optimum, and the convergence performance is better than the traditional least square method. In addition, Mahvi and Behjat [223] also used the GA to estimate the detailed model of the damaged winding by the fault from the measured low-frequency response data up to 10 kHz. The experiments made on a test transformer show that the newly developed method is sufficiently able and sensitive to detect and localize faults of only few shorted turns on the transformer windings.

(2) *AIA:* it is a new evolutionary theory which is inspired by the biological immune system, which introduces the immune mechanism based on the original theoretical framework of evolutionary algorithm, and imitates the function of the natural immune system [224]. In AIA, the affinity of antibodies and antigens is treated as the matching degree between the feasible solution and the objective function, such that the affinity between antibodies ensures the diversity of feasible solutions, the heredity and variation of the superior antibody is promoted by calculating the expected survival rate of antibodies, and the feasible solutions after selection and optimization stored by the memory cell unit are employed to restrain the continuous generation of similar feasible solutions and to accelerate the search to the global optimal solution. At the same time,

when the similar problems appear again, the better solution and even the optimal solution of these problems can be generated rapidly. Hence, as stated here, the basic procedures of the AIA are successively presented as [224]: problem identification; antibody group generating; calculate the fitness value of antibodies; produce the immune memory cells; selection of the antibodies, including promotion and restraint; evolution of the antibodies; updating of the antibody population; and termination. Based on this, the basic flow chart of AIA [224] is illustrated in Figure 9.
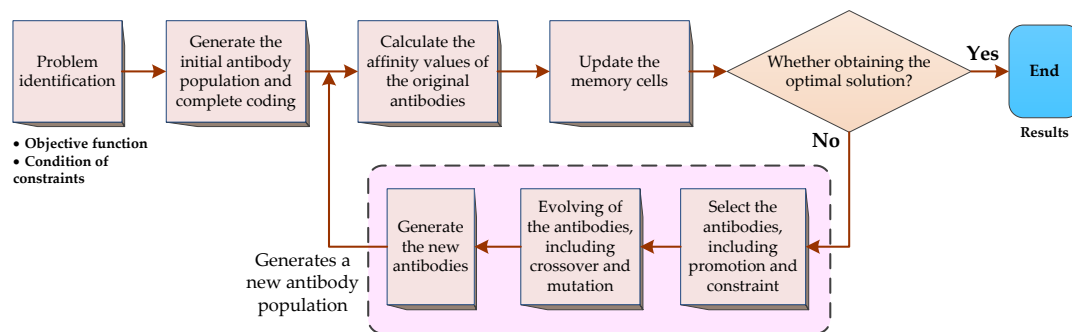
Figure 9. The basic flow chart of AIA.

Based on the principle of AIA introduced above, Yuan et al. [225] employed AIA in DGA-based transformer fault diagnosis, in which the optimal fault results are screened by the immune mechanism, and the accuracy of diagnosis is improved compared with the traditional three-ratio method. Moreover, based on genetic SVM and grey AIA, Zheng et al. [72] proposed a two-classifier cascade power transformer fault diagnosis algorithm, which is employed to solve the problem of both single and multiple power transformer fault diagnosis. Here, the first classifier is presented as SVM classified fault or normal state of power transformer and the high-frequency variation based on dynamic vaccine mechanism generates a new antibody. The experiments indicate that this proposed method combines genetic SVM with grey AIA and dynamic vaccine mechanism can effectively classify single and multi-fault of the power transformer and raise the fault diagnosis accuracy and diagnosis speed. In this diagnosis method [72], the flow charts of transformer fault diagnosis and grey AIA based dynamic vaccine mechanism are shown in Figures 10 and 11, respectively.

Besides the research work in [72], AIA-improved can be combined with Li et al. [163] proposed a new power transformer fault diagnosis method based on RST (RST) and improved artificial immune network classification algorithm, according to the complementary strategy. In this method, the minimal diagnostic rules are obtained through reduction approach of RST information table. In the first step, both the antigens and memory antibodies with class information are added into the artificial immune network and then are trained to learn the features of fault samples. In the second step, the *k*-nearest neighbor method is employed to classify the fault samples. Tests show that this proposed method has better capability to classify single- and multiple-fault samples with higher diagnosis accuracy than the IEC three-ratio method and BPNN. Similarly, Song et al. [176] also developed a transformer fault diagnosis model based on the rough set and AIA, which has high diagnosis accuracy, strong robustness and good learning ability. Yuan et al. [225] also used AIA for fault diagnosis of the power transformer, which can obtain higher accuracy of diagnosis compared with the results of IEC three-ratio methods.

Currently, many immune mechanisms and theories have been investigated, and some of them have been applied in the artificial immune system [224], including B-cells, T-cells, antibody, antigen, immune learning, immune memory, immune network theory, immune risk theory, clonal selection theory, affinity maturation, negative selection, affinity, gene pool, diversity, distribution, the innate immune system, the adaptive immune system, immune response, immune tolerance, immune system hierarchy, etc. The DGA-based transformer fault diagnosis can obtain inspiration from these mechanisms and theories, which will be a new development direction in the future. Besides, the AIA can be combined with the computational immunology to make a new breakthrough in the

future, which will be able to establish some new artificial immune system theories, such as the optimization oriented immune computation theory, data mining oriented immune mining theory, and control oriented immune control theory.
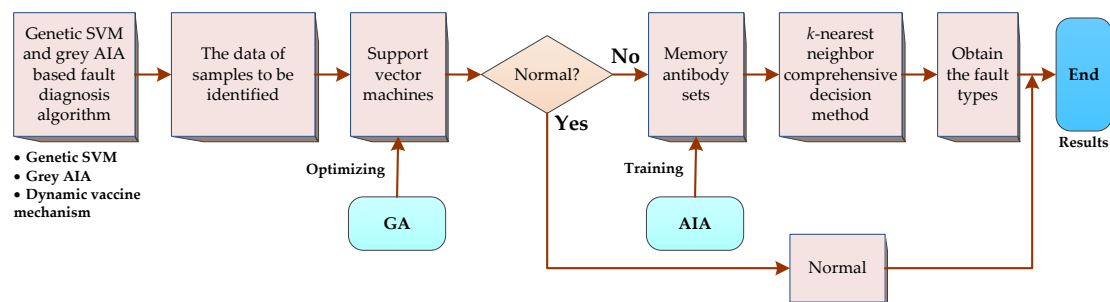
Figure 10. The genetic SVM and grey AIA-based transformer fault diagnosis flow chart presented in [72].
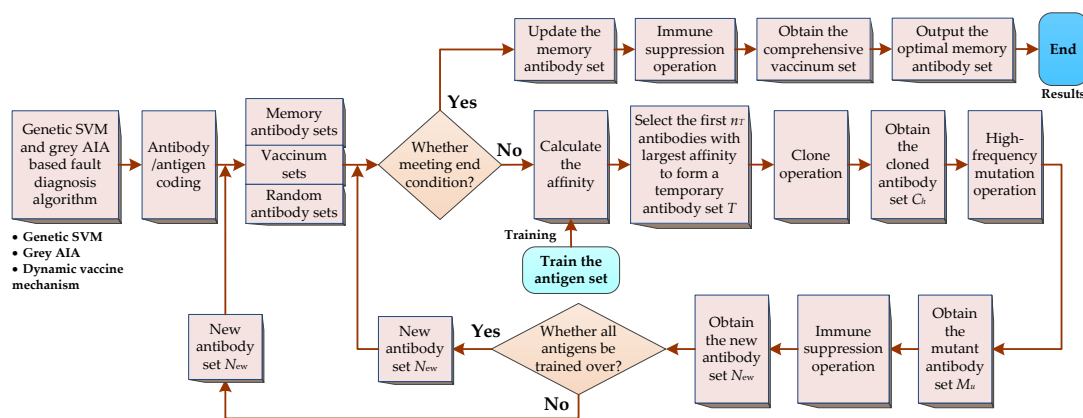
Figure 11. The grey AIA based dynamic vaccine mechanism flow chart presented in [72].

(3) *ACO:* it is also called ant algorithm, is a probabilistic technique employed to find the best solution. ACO was proposed by Marco Dorigo [226,227]. In ACO, a single ant in the population can leaves a substance called pheromone on the path it passes during the process of foraging, and it can perceive the intensity of the pheromone in this process. At the same time, the ants move towards the high intensity pheromone. Hence, the collective foraging of ant population is shown as a positive feedback phenomenon to the pheromone, based on which, the optimal path is gradually approximated and finally it can be found. Here, the optimal path is found through the positive feedback and distributed collaboration, which can be seen as the main feature of ACO. Aiming at parameters are confirmed by the cross-validation, Mo [228] proposed the ACO, SVM and IEC method-based model, called ACSVM-IEC, for the transformer fault diagnosis, which can find out the optimum accurately in a wide range, and is robust and practical for transformer fault diagnosis. Based on the accurate assessment of DGA on the insulation condition of power transformer, Liu et al. [229] proposed an approach of ACO-SVM to recognize histograms of characteristic dissolved-gas in transformer oil. Here, the DGA data of transformer oil with normal operation and four types of typical power transformer faults are selected, and the characteristic dissolved-gases are employed to establish histograms. This approach to DGA-based transformer fault diagnosis integrates the IEC three-ratio method, SVM and ACO-SVM, which shows effectiveness in improving the accuracy of the recognition for characteristic dissolved-gas histograms compared with the other two methods. Analogously, Niu et al. [230] also proposed a fault diagnosis method for the power transformer based on ACO-SVM classifier, which is effective to detect failure of transformer. Moreover, Tian et al. [231] proposed an improved ACO based on the transformer diagnosis data reduction, which is proved to have higher diagnosis accuracy rate in data reduction and have fast diagnosis speed compared to either traditional algorithm. Li et al. [232] combined the RBF network with ACO and

Fisher ratio algorithm to develop a novel model for the fault diagnosis of oil-immersed transformer, in which the combined ACO and Fisher ratio algorithm is employed to optimize the RBF structure. This method shows more effective than the ACO alone in fault diagnosis of oil-immersed transformer. Wei and Cui [233] investigated the power transformer fault test sequence optimization based on ACO. The power transformer fault test path is optimized with global optimization and heuristic optimal of ACO. The simulative results show the ACO based fault diagnosis method is feasible and effective. Besides, Qiao [234] employed the ACO to develop a transformer intelligent breakdown fault diagnosis system, which can overcome the defects of network to fall into partial minimum, and is a practical and effective method, with higher diagnosis accuracy.

(4) *PSO:* it was proposed by Kennedy et al. [235] as a kind of evolutionary computing technique. In the PSO, each feasible solution can be treated as a particle that has two properties: location x and speed v. The fitness function of each particle is calculated in each round of iteration. Then two optimal particles will be tracked continuously and they are: the optimal location experienced by the current particle, called *pBest*, and the particle which has a global optimal location, called *gBest*. Based on the two particles, the speed and location of any particle are updated [218] as follows:

$$v = \omega v + c_1 r_1 (pBest - x) + c_2 r_2 (gBest - x), \quad x \in [x_{\min}, x_{\max}], v \in [v_{\min}, v_{\max}] \tag{14}$$

$$x = x + v, \quad x \in [x_{\min}, x_{\max}], v \in [v_{\min}, v_{\max}] \tag{15}$$

where ω represents inertia weight, which controls the impact of past speed on current speed. c_1 and c_2 are normal numbers, representing acceleration factors. r_1 and r_2 are two random numbers distributed uniformly in [0, 1]. Hence, the algorithm flow of PSO is introduced as follows:

Step 1: initialization, namely the speed v and location x of each particle are set randomly.

Step 2: calculate the fitness function of each particle.

Step 3: for each particle, its fitness is compared with the *pBest*, if the current particle is better, then *pBest* = x.

Step 4: for each particle, its fitness is compared with the *gBest*, if the current particle is better, then *gBest* = x.

Step 5: update the speed and location of each particle according to (14) and (15).

Step 6: if the end conditions are not met, then go back to step 2; otherwise, output the speed v and location x of the optimal particle.

On this basis, PSO and the combination of it with other intelligent approaches have been gradually employed by researchers in the field of DGA-based transformer fault diagnosis. To this end, Tang et al. [69] developed a Parzen windows-based classifier for transformer fault diagnosis, which is able to interpret transformer DGA with a probabilistic scheme. The PSO as a global optimizer is employed to optimize the parameters of Parzen windows to improve the fault classification accuracy. This method improves both the diagnosis accuracy and computational efficiency when it is compared with a number of fault classification techniques. Fei et al. [77] proposed a PSO-SVM model to forecast dissolved gases content in power transformer oil, among which the PSO is employed to determine the free parameters of the SVM. The tests show that this model can achieve greater forecasting accuracy than GM and ANN under the circumstances of small samples. In addition, Liao and Zheng [80] developed a forecasting model on dissolved gases in oil-filled power transformers based on PSO-least squares support vector regression. In order to improve the performances of SVM classifier for the purpose of incipient faults syndrome diagnosis of power transformers, Lee et al. [236] proposed a PSO-based encoding technique to improve the accuracy of classification. Experiments on real operational data demonstrate that this proposed approach is effective, has high efficiency, and can make operation faster and also increase the accuracy of the classification. Moreover, in view of non-linear characteristics between fault symptoms and fault types of transformers, Cheng et al. [237] combined the WNN with an improved PSO in order to design a novel model for transformer fault diagnosis using the data of DGA. This model is constructed by three-layer WNNs, and is trained by an improved PSO. The PSO applied in

this model can accelerate the training speed of WNN and improve the accuracy of training. The simulative experiments demonstrate that this improved PSO and WNN can be effectively applied to transformer fault diagnosis using DGA and provides a new way for transformer fault diagnosis.

(5) *BFO:* it is a new optimization method proposed based on the basic laws of the growth and evolution of bacterial colonies. The biological basis of BFO is the intelligent expression of Escherichia coli in the process of foraging in human intestinal tract. The location of bacteria is updated according to three iterative procedures: chemotaxis, reproduction and dispelling, so that bacteria can tend to a nutrient rich place. Based on this main principle of BFO, Geethanjali et al. [238] proposed an entirely new approach for detection and discrimination of different operating and fault conditions of power transformers. In the proposed scheme, ANN techniques have been applied to transformer protection to distinguish internal faults from normal operation, magnetizing inrush currents and external faults. In addition, Gopila and Gnanambal [239] developed a Hyperbolic S-Transform BFO technique to solve an optimization problem that is formulated by the task of detecting inrush and internal fault in power. In this technique, the BFO has been demonstrated the capability of identifying the maximum number of faults covered with minimum test cases and therefore improving the fault detection efficiency in a wise manner.

(6) *AFSO:* it was proposed by Li in [240,241] as a SI algorithm to simulate the foraging activities of fishes. The main principle of AFSO is to adopt the idea of a bottom-up approach to complete optimization, by simulating the three basic behaviours of fishes: foraging, gathering and pursuing. Hence, AFSO has some advantages such as fast optimization speed, global optimization ability and strong parallel processing capability. Although the AFSO has a lower requirement for the initial values and can be realized easily, it also has some defects, such as the optimization precision is not high and the convergence rate is slower in later period. To address it, some improved AFSO algorithms have been proposed by scholars and have been successfully applied in the transformer fault diagnosis using DGA data. For some examples, Yu et al. [197] proposed an IAFSO to optimize the weight and threshold of the BP. The global searching ability of the IAFSO approach is utilized to find the global optimization solution, so that it can overcome the slower convergence velocity and easily getting into local extremum of the BP neural network. Experimental results indicate that the proposed IAFSO can improve both convergence velocity and veracity to some extent. Geng et al. [198] proposed a hybrid AFSO and frog leaping algorithm to identify J-A parameters, which combined the advantages of fast convergence from AFSO and high local search accuracy from a frog leaping algorithm. The results show that the hysteresis curve generated from the proposed hybrid algorithm has a great consistency with the measured curve. In addition, Yu et al. [199] also developed a model based on the IAFSO and SVM and it has been applied in transformer fault diagnosis. In this model, the IAFSO is utilized to find the optimization solution of the SVM parameters. Experimental results show that the proposed algorithm can find out the optimum accurately in a wide range.

(7) *ABC:* it was proposed by Karaboga in 2005, as a kind of biomimetic algorithm employed to simulate the intelligent search behaviour of a bee colony [242]. ABC is developed to simulate the process of honey collection by bees in Nature, and in this process, various stages of the tasks will be completed by bees in the population in the process of honey collecting, according to their different functions, and then the optimal solution of the problem is found through the collection and sharing of information and food sources. Hence, ABC has the characteristics of less control parameters, simple calculation and easy implementation. To this end, Yilmaz [243] proposed a multi-objective ABC algorithm to estimate transformer equivalent circuit parameters. Besides, ABC is rarely employed in transformer fault diagnosis based on DGA data. In the future, the improved ABC and the combination of it with other intelligent methods may be a new breakthrough in DGA-based transformer fault diagnosis due to excellent characteristics of ABC.

(8) *FOA:* it was proposed by Krishnanand [244]. The main principle of FOA is introduced as follows: a firefly attracts other companions through the fluorescence emitted by the individual's fluorescein during the simulation of firefly movement, and then fireflies move to the firefly that has the brightest fluorescent and better location in the area, so that achieving the best location of its own.

On this basis, Huang et al. [245,246] proposed a transformer fault diagnostic method based on grey fuzzy FOA. In this method, characteristic gas coding sequences are used as inputs of training samples, and transformer fault types corresponding to the inputs are used as outputs to build an FOA-LM network, and the weight value and the threshold value of the LM network are optimized through an FOA. The pre-treated data of the characteristic gases of the transformer are used to train the network, with the aim of obtaining an optimal nerve net weight value. This method can solve the problems of data source shortage of transformer fault gases and low result accuracy in a conventional analysis method.

(9) *BOA:* it was proposed by Yang in 2010 as a type of metaheuristics optimization algorithm [247]. BOA, as a SI algorithm, is employed to imitate bats in Nature that use sonar to avoid obstacles to detect prey. BOA is a new intelligent algorithm, which offers obvious improvements in validity and accuracy, and meanwhile it has the characteristics of a simple model, strong searching ability and fast convergence speed. Hence, Gong et al. [248] employed it for power transformer fault diagnosis based on improved BPNN. During the diagnosis, the bat algorithm of BPNN weights and threshold parameters optimization can improve the speed of convergence. The BPNN model is constructed according to the obtained parameter values, and the data are trained and tested. An example analysis shows that the optimization of BPNN for the fault diagnosis of the transformer is practical and effective.

(10) *Hybrid SI algorithms:* they have been adopted by more and more researchers in power transformer fault diagnosis using DGA. This is because multiple limitations of a single SI algorithm are becoming more and more prominent, especially when dealing with a very complicated optimization problem. Therefore, many scholars have proposed to improve the single SI algorithm by drawing lessons from the features of other intelligent algorithms, as well as a hybrid SI algorithm to optimize the SVM parameters. Li et al. [219] found that the experimental results are easy to fall into local optimum, such as GA, ABC and flower pollination algorithm (FPA), after they compared the performance of each SI algorithm in SVM parameters optimization; At the same time, the convergence is slow in later stage, such as the PSO; and the optimization speed is too slow, although some algorithms have fast optimization speed, there is a problem of low optimization precision; AFSO demonstrates the best performance in the SVM parameter optimization methods when it is compared with several other SI algorithms; the main problem of ACO is how to effectively solve the problem of continuous optimization; BOA and AFSO both have the problem of low optimization accuracy. Hence, Li et al. gave a detailed comparison of advantages and disadvantages of optimizing the SVM parameters of various SI algorithms in [219], as shown in Table 10.

Table 10. A detailed comparison of advantages and disadvantages of optimizing the SVM parameters of various SI algorithms in [219].

SI Algorithm	Advantages	Existing Problems
GA	▪ Possess global optimization ability ▪ Perform distributed computing based on intrinsic parallelizability	▪ Results easily fall into local optimum ▪ Have longer training time ▪ Have lower classification precision
ACO	✓ Stronger global optimization ability ✓ Seen as a kind of positive feedback algorithm ✓ Stronger robustness	✓ Failed to deal with continuous optimization problems
PSO	▪ Have fast speed of convergence in early stage ▪ High optimization precision	▪ Have slow convergence speed in later stage ▪ Easy to fall into local optimum
AFSO	✓ Stronger parallel processing ability ✓ Have fast speed of optimization ✓ Possess global optimization ability	✓ Precision of optimization is not high ✓ Have slower convergence rate in the later stage
ABC	▪ Less control parameters ▪ Easy to calculate	▪ Slow convergence rate ▪ Easy to fall into the local optimum
FOA	✓ Have fast speed of optimization	✓ Slow convergence rate in the early stage ✓ Unstable convergence rate in the later

		stage
BOA	▪ Have fast speed of convergence ▪ Have strong ability of searching	▪ Optimization precision is not high
SFLA [1]	✓ Less parameters and easy to implement ✓ Have superior global optimization ability	✓ Easy to fall into the local optimum
FFOA [2]	▪ Have simple structure ▪ Easy to implement	▪ Easy to fall into the local optimum ▪ Optimization precision is not high
BFO	✓ Parallel search ability ✓ Easily jump out of local minima	✓ Convergence rate is not fast ✓ Optimization precision is not high
HIAs [3]	▪ Possess stronger global search ability ▪ Avoid premature convergence of algorithms ▪ Convergence rate is improved in later stage	/
Hybrid SI algorithm	✓ Avoid prematurely falling into local optimum ✓ Optimization of algorithms are accelerated ✓ Optimization precision of algorithms are improved	/

[1] SFLA = shuffled frog leaping algorithm; [2] FFOA = fruit fly optimization algorithm; [3] HIAs = hybrid intelligence algorithms, which represents the combination of SI algorithms with other algorithms

Based on Table 10, it can be concluded that the mixed use of SI algorithms can find the optimal parameters and greatly improve the prediction and classification accuracy of SVM. In addition, the improved algorithms not only possess the advantages of one SI algorithm, but also possess the advantages of another one, thus this improvement is feasible and effective. On this basis, some improved SI algorithms have been developed and successfully applied in the transformer fault diagnosis using DGA data. A brief summary is presented as follows.

In view of timely and accurate grasp of the health status of the power transformer, and carrying out predictive analysis of the incipient faults, Zhang et al. [249] developed a diagnosis model for transformer fault based on chemical reaction optimization BPNN and fusion DGA method, which combines both advantages of AI and DGA method. In this proposed model, CRO and fusion DGA method are employed to overcome the defects of BPNN and traditional DGA method. The results reveal that the accuracy, iterations and training time of the model are 87.88%, 1991 and 1927 ms, respectively. This demonstrates that the model has distinct advantages when compared with those of other models. Zang et al. [250] proposed a hybrid intelligent method for power transformer fault diagnosis, based on the integration of evolutionary programming, fuzzy theory, ANN and case-based reasoning. Huang and Zhao [251] proposed to combine the rough sets with multi-population GA for transformer fault diagnosis, in which the immigration operator and partial competitive rules are employed to maintain the diversity of population so as to avoid the results falling into local optimum. Yao [252] proposed the transformer fault diagnosis model based on the integration of improved AFSO and RBFNN using the data of DGA. In this model, the parameters and behaviours of AFSO are improved by introducing appropriate strategies, including adaptive strategy, fragmentation strategy, jumping behaviour, and stepping behaviour. This designed AFSO-RBF network shows strong superiority in diagnosis performance, and has good practical value in transformer fault diagnosis. Illias et al. [89] proposed a hybrid modified evolutionary PSO-time varying acceleration coefficient-ANN for power transformer fault diagnosis. Meanwhile, Illias and Zhao [253] employed hybrid SVM-modified evolutionary PSO to identify transformer faults based on DGA. Wang et al. [254] developed a new hybrid evolutionary algorithm combining PSO and BP algorithm, called HPSO-BP algorithm, to select optimal value of probabilistic neural network parameters for power transformer fault diagnosis. In this model, PSO is employed to perform a global search to give a good direction to the global optimal region, and then BP algorithm is used as a fine tuning to determine the optimal solution at the final. The experimental results show that the proposed approach has a better ability in terms of diagnosis accuracy and computational efficiency.

To sum up, SI algorithms are generally employed to optimize the parameters, such as SVM parameters, in the process of DGA-based transformer fault diagnosis, and a lot of research work has

been carried out on this. However, some defects still exist in SI algorithms. Taking GA and ACO, although the improved GA can avoid falling into a local optimum under certain conditions, it still has the possibility of falling into local optima, thus it is essential to further investigate and solve this problem in judging whether the algorithm is falling into and jumping out of the local optimum; although ACO is able to obtain a better classification performance of optimal parameter combinations than other SI algorithms, it still has some issues to be addressed, such as the longer operation time and the fact the time complexity of the algorithm is higher when the number of samples is increased to a certain extent, so that the convergence performance of ACO still needs to be further studied. In the transformer fault diagnosis using DGA data, the selection of a SVM parameter model is still a typical problem. A number of SI algorithms have been proposed to address this through optimizing the SVM parameters, based on the advantages of SI algorithms, such as the characters of strong distribution, self-organization and robustness. Some SI algorithms have been investigated, improved and applied continuously, especially for the application of optimizing SVM parameters, in transformer fault diagnosis using DGA data in recent years, and some achievements have been made in transformer fault diagnosis, which demonstrates that SI algorithms possess stronger parallel processing capability, fast optimization speed, and can effectively avoid falling into local optimum and carry out global optimization, with a high prediction and classification precision. However, SI algorithms cannot ensure a strong capability of optimization in each condition, and it also cannot guarantee that the obtained optimal parameters have good classification and prediction abilities for each kind of parameter model.

Hence, there is still much room for the theoretical investigation and application of SI algorithms in DGA-based transformer fault diagnosis, and it is essential to combine the convergence of the algorithm with the prediction of optimization together and consider it at the same time. The research direction of SI algorithms in transformer fault diagnosis using DGA should be focused on the following aspects: improving the performance of SI algorithms to avoid falling into local optima; investigating the setting of algorithm parameters to improve the algorithm performance; conducting the proof of algorithm convergence; carrying out the combination of SI algorithms with other intelligent algorithms to develop more effective and efficient fault diagnosis algorithms.

7.2. Data Mining Technology

7.2.1. Data Mining Technology Introduction

The application of data mining technology in DGA-based transformer fault diagnosis is reflected in the use of a computer to automatically process a large amount of data samples to find hidden relationships or rules. In this process, the data mining technology can be employed to model the fault symptoms of the power transformer to find the law of describing the interrelationship between the operating status of the transformer and its external performance. The process of data mining contains four procedures: data preparation, define the topics, establish the corresponding model, and understand the built model. For example, in the last step, after the model is established, the program begins to perform the task of understanding the built model, which is implemented mainly through the analysis of frequency index and influence value index. The former index explains the percentage of fault records in each group of data accounts for the current records. The latter index indicates the importance of a record value to the prediction, and its value is taken from 0 to 100.

7.2.2. Application of Data Mining Technology in Transformer Fault Diagnosis

Data mining technology can be applied to perform data mining of data sources and establish relevant models, which is beneficial to obtain the intuitive relationship between the operation state of transformer and the result of fault analysis, thus the prediction accuracy is high. Moreover, when the data source is more complex and the amount of data is larger, the more adaptable the built model is, thus the more practical the predictive result is. However, the process of setting up the

data source will be more complex. To address it, the clustering technology has been introduced in the DGA-based transformer fault diagnosis. Fu et al. [155] introduced the weighted fuzzy kernel clustering method to the power transformer fault diagnosis based on DGA, which can effectively solve the problem that fuzzy C-means algorithm is susceptible to the influence of sample distribution and initial parameters. The examples demonstrate that this proposed method can quickly and effectively cluster the sample data, thus it can meet the requirements of transformer fault diagnosis. In addition, Hao et al. [79] used the dynamic clustering algorithm to diagnose transformer faults, in which the artificial immune network is firstly employed to carry out immune memory and learning of the fault samples, so that the useful characteristics that effectively represent the fault samples can be extracted and used as the initial clustering centres of kernel-based probabilistic clustering algorithm. In the second step, GA is used to dynamically optimize and select the number and centres of clustering to achieve the classification of the fault samples. The diagnostic results indicate that the fault samples are effectively classified using this proposed method and the fault diagnosis precision is improved, by comparing it with the results obtained by BPNN. Lin et al. [63] combined the grey prediction with clustering analysis, and developed a relevant model to enhance oil-immersed transformer fault diagnosis using dissolved gases forecasting. Aiming at the problem that power transformer fault reasons are very complicated owing to the fuzziness and uncertainty between the failure phenomenon and failure mechanisms, Zheng et al. [81] proposed an iterative self-organizing data analysis technique algorithm, called ISODATA, based on DGA, which can largely overcome the dependence on initial cluster centre and can be easily applied to oil-immersed transformer fault diagnosis. In addition, Sima and Shu [255] established a SVM based multilevel binary tree transformer fault diagnosis model, in which the adaptive k-means clustering algorithm is put forward to resolve multi-class problem.

Apart from the clustering analysis method as a kind of data mining technology applied in the DGA-based transformer fault diagnosis, some scholars have employed information fusion technology in transformer fault diagnosis. Hu et al. [200] put forward a new fault diagnosis model which has combined the on-line monitoring of five characteristic gases dissolved in transformer oil with the feature extraction of three-dimensional temperature field of power transformer. This model can realize fusion of the multivariate fault information of power transformer and greatly improve the fault diagnosis accuracy of the power transformer based on the information fusion technology. Based on the model, an on-line state monitoring and fault diagnosis system has been developed, which can change the existing maintenance and repair pattern for power transformer as well as realize accurate diagnosis or forecast faults of transformer intelligently. Li et al. [201] employed the multi-sensor information fusion technology in the power transformer fault diagnosis. Besides, Gong and Zhang [202] also developed the fault diagnosis model of transformer based on the technology of information fusion, which can achieve a perfect goal, namely the precision of fault diagnosis result has been enhanced definitely.

7.3. Machine Learning

7.3.1. Machine Learning (ML) Description

With the continuous improvement of computing power and computing theory innovation, great progress has been made in ML in the past 30 years, and it is being considered by more and more researchers. ML has been widely employed in various areas, such as biology, medicine, energy, transportation and environment. Based on a conventional ML framework, new ML methods and theoretical framework have been continuously proposed by scholars [256,257]. The essence of a ML algorithm is to find a target function (f), with the aim of making it the best mapping between the input variable (X) and the out variable (Y), that is, $Y = f(X)$. The most common type of ML is to find such $Y = f(X)$ and use which to forecast value of Y corresponding to the new X. Such a process is called predictive modelling or predictive analysis whose goal is to derive the most accurate predictive result as much as possible. Aiming at the relationship between data acquisition and action selection, the mathematical models can be established to describe the theoretical

framework of some common ML algorithms presented in [257]. Assume that a set of data are obtained to make up a collection $X = \{x_i\}$, $i = 1, 2, …, I$. if the research objective is a complex system, then these data are usually observed system states or outputs. For these data, a series of actions a_k can be taken to make up a collection $A = \{a_k(X')\}$, $k = 1, 2, …, J$. $X' \subseteq X$ represents a subset of data set X. Each action can generate a reward $R(a_j)$, and the data acquisition and taking action can be separated in time. The target is to maximize the long-term reward via ML [257], namely:

$$\max_{a_j, 1 \leq j \leq J} \sum_{k=1}^{k=J} R(a_k) \tag{16}$$

If each action a_k causes a loss $L(a_k)$, then according to (16), the objective function can be transformed into minimizing the long-term loss as:

$$\max_{a_k, 1 \leq k \leq J} \sum_{k=1}^{k=J} L(a_k) \tag{17}$$

For the common supervised learning, the model presented in (17) can be further simplified as follows: when all data are known and have been classified correctly, then take an action to establish a function mapping (is usually the classification function) to minimize the classification error. Generally speaking, the pre-set data are subject to independent and identically distributed (i.e., i.i.d), then the objective function can be further expressed [257] as:

$$\min_{a} L(a(X)) = \min_{a} \sum_{i=1}^{i=I} L(x_i, a) \tag{18}$$

Compared with supervised learning, online ML [258] emphasizes that data is gradually acquired and every time a new data is acquired, the system can take an action based on all the acquired data. For a special case of online ML, called sequential learning, every time only one data is acquired as x_i, and one predictive action $f(x_i)$ is generated according to the mapping function $f(\cdot)$. Next, the real labeled data $y(x_i)$ of x_i is acquired, and then the loss generated by it can be calculated as $L[f(x_i), y(x_i)]$. Finally, the objective function is obtained by selecting the appropriate $f(\cdot)$ to minimize the long-term regret value [257]:

$$\min_{f} L(a(X)) = \min_{f} \sum_{i=1}^{i=I} L\left[y(x_i), f(x_i) \right] \tag{19}$$

where multiple actions exist, and with the increase of the number of data obtained, the actions taken will be optimized continuously.

Similar to online ML, RL allows the action a_t at time t to influence the data x_{t+1} that is obtained at time $(t + 1)$, thus a specific state transfer function exists as $T(\cdot, \cdot): X \times A \rightarrow X$, namely:

$$T\left(x_t, a(x_t), x_{t+1} \right) = \Pr\left(x_{t+1} \mid x_t, a(x_t) \right) \tag{20}$$

Then, the objective function is as:

$$\max_{a} \sum_{t=1}^{\infty} T(x_t, a(x_t), x_{t+1})[R_a(x_t, a(x_t)) + \chi \cdot V(x_{t+1})] \tag{21}$$

where $R_a(x_t, a(x_t))$ is the immediate reward obtained by the system in state x_t taking action $a(x_t)$ at time t. $V(x_{t+1})$ is the long-term mean reward of the system in state x_{t+1}. χ is the discount factor. It can be seen from (19) and (20) that RL is one kind of active learning [259], thus specific action can be taken to give consideration to both optimization objective function and exploration of input data set X. This is a significant improvement against online ML. Therefore, the basis of RL is to maximize the cumulative value of the reward obtained by agent from the environment, in order to learn the optimal strategy to accomplish the goal. This indicates that RL is more focused on learning strategy for solving problems. Deep learning (DL) is originated from investigation of ANN, thus multi-layer preceptor with multiple hidden layers is a common used DL model. DL has been widely used and a

series of breakthroughs have been achieved using DL in image recognition, speech recognition and natural language processing, etc. DL has been concerned by more and more researchers due to its strong representational ability and generalization ability. The perceived ability of DL can be combined with RL that has ability of decision learning to form a new ML, called deep RL. According to [260,261], the illustrations of RL and deep RL are demonstrated in Figures 12a,b, respectively.
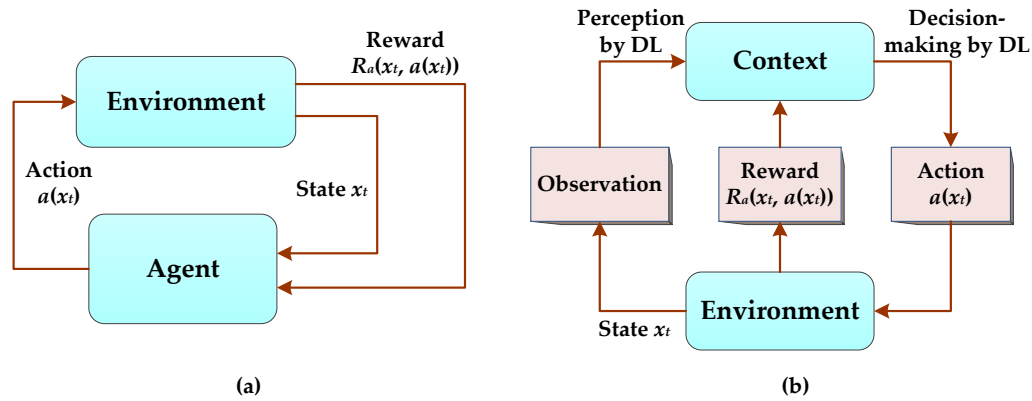
Figure 12. Principle framework diagram of RL and deep RL. (**a**) is the principle framework of RL; (**b**) is the principle framework of deep RL (DRL).

7.3.2. ML-Based Transformer Fault Diagnosis

Actually, the process of ML-based transformer fault diagnosis using DGA can be divided into two procedures. Firstly, establish the mapping model of state vector space of transformer to transformer fault type space, and then use this built model to identify the fault types of the transformer according to the obtained unknown data samples. According to the principle explanation of ML above, a brief summary is conducted in this section on the application of ML methods such as SVM, ELM, and DL in transformer fault diagnosis using DGA. Besides, the difficulty and future development direction of these ML methods in transformer fault diagnosis are presented as well.

(1) *SVM:* in view of fault diagnosis of the transformer using ML, SVM has been widely used since it was proposed by Vapnik et al. in 1995. SVM is a kind of ML method based on statistical learning theory [262], which has great advantages in solving small sample size and non-linear problems. SVM is a two-classifier which distinguish transformer faults by constructing multi-branch classification SVM. In recent years, SVM is generally employed in combination with AI algorithms to develop new algorithms or improved AI algorithms, which are then used to optimize and design misjudgment penalty factors, thus enhancing the role of empty rotation sample in the construction of the classification hyperplane, and suppressing deviation of the hyperplane. On this basis, Bacha et al. [5] proposed an intelligent fault classification approach to power transformer DGA. Here, SVM as a powerful tool is employed to deal with the problem with small sampling (i.e., small amounts of training data), nonlinearity and high dimension (i.e., large amounts of input data), hence SVM is applied to establish the power transformer faults classification and to choose the most appropriate gas signature. In this method, the experimental data from the Tunisian Company of Electricity and Gas are tested and the test results indicate that the extension method and SVM approach can significantly improve the diagnosis accuracy for power transformer fault classification. In order to effectively select appropriate SVM parameters, Fei and Zhang [68] combined SVM with GA to develop a novel algorithm for fault diagnosis of a power transformer, called SVM-GA algorithm. In this algorithm, GA is employed to select appropriate free parameters of SVM. Experimental results indicate that this proposed method can achieve higher diagnosis accuracy than IEC three-ratio methods, normal SVM classifier and ANN. Zheng et al. [72] proposed a two-classifier cascade power transformer fault diagnosis algorithm to

solve the problem of both single and multiple power transformer fault diagnosis. In this algorithm, SVM classified fault or normal state of power transformer is made as the first classifier, and GA is used to optimize the kernel function parameter of SVM. Genetic SVM combined with grey artificial immune and dynamic vaccine mechanism can effectively classify single- and multi-fault of power transformer. In addition, Fei et al. [77] made full use of strong global search capability of PSO and then developed a PSO-SVM model to forecast dissolved gases content in power transformer oil, among which PSO is employed to determine free parameters of SVM. This PSO-SVM method can achieve greater forecasting accuracy than GM, and ANN under the circumstances of small sample. Analogously, Liao and Zheng proposed a PSO-least squares support vector regression-based forecasting model on dissolved gases in oil-filled power transformer, in which the least squares-SVM regression model with RBF kernel is established to facilitate the forecasting model, and the PSO is employed to optimize the hyper-parameters needed in least squares-SVM regression. Shah and Bhalja [85] implemented discrimination between internal faults and other disturbances in transformer using the SVM-based protection scheme. Sima et al. [154] combined the GRA with fuzzy SVMs to form a novel concentration prediction model of dissolved gases in transformer oil, which has considered the influence of oil temperature and loads on oil-dissolved gases. This developed model has made full use of advantages of GRA to extract the key factors that have great influence on characteristic gases concentration and that acted as attributes of input samples of the SVM regression modelling using fuzzy membership function. Likewise, the SVM-based fault diagnosis algorithm of transformer using DGA has been developed both in [169,170]. In the two investigations, the rough set theory has been integrated into SVM to obtain the rough faulty point of the power transformer with a satisfactory accuracy. In addition, Zhang et al. [188] developed a new SVM model for fault diagnosis of oil-immersed transformers based on an improved imperialist competitive algorithm (IICA), in which SVM is introduced as an effective fault diagnosis technique based on DGA for transformers with maximum generalization ability, and the IICA is employed to optimize the SVM parameters appropriately. Three classification benchmark sets are investigated in [188] based on PSO-SVM and IICA-SVM with four multiple classification schemes to select the best scheme for transformer fault diagnosis. Meanwhile, Chao et al. [199] thoroughly investigated the combined improved artificial fish swarm and SVM applied in transformer fault diagnosis; and Wang et al. [263] made use of the distinctive strength of SVM algorithm in solving small sample size problems and applied the SVM in DGA-based transformer fault diagnosis, by employing the cross-validation based grid search method to determine the parameters of SVM, so as to construct the power transformer fault diagnosis model, which is better used in practice.

(2) *Extreme learning machine (ELM):* ELM has been introduced into DGA-based transformer fault diagnosis in recent years. ELM is an emerging learning algorithm which is proposed for the single-hidden-layer feedback neural network by Professor Huang in 2004 [264]. ELM is characterized by the fact that the weight matrix and biasing of the input layer and the hidden layer are generated randomly only at one time, without the need of iterative optimization. The only solution to the parameter is the weight matrix of hidden layer and output layer, which is obtained by the generalized inverse matrix method, so that the solving process is more quickly [265,266]. ELM has achieved good application in speech recognition, fault diagnosis, and image classification, especially the ELM has been applied to the fault diagnosis of the power transformer currently due to the characters of fast learning speed and good generalization of it. According to [267], the network structure of ELM is illustrated in Figure 13a, where m, L, and n are the number of nodes of the input layer, hidden layer and output layer, respectively, y is the output of the ELM network, and x_1, x_2, …, x_p are the training samples.

Based on Figure 13a, the output of ELM network is presented [267] as:

$$y = \sum_{i=1}^{L} \beta_i g(\boldsymbol{\omega}_i^{\mathrm{T}} \boldsymbol{x}_i + b_i) \qquad (22)$$

where x_1, x_2, …, x_p are the training sample sets of the ELM network, and their corresponding labels are t_1, t_2, …, t_p, respectively. $g\left(\boldsymbol{\omega}_i^{\mathrm{T}}\boldsymbol{x}_i + b_i\right)$ is the activation function of the hidden layer. w is the

weight matrix sized $m \times L$. $\boldsymbol{\omega}_i$ is the weight vector between the ith node of the hidden layer and the input layer. b_i is the bias parameter the ith node of the hidden layer. β is the weight matrix between the hidden layer and the output layer, sized $L \times n$. β_i is the weight vector between the ith node of the hidden layer and the output layer. $\boldsymbol{\omega}_i$ and b_i are generated randomly, making the ELM can directly generate the global optimum, which is finally transformed into the minimum norm least squares solution, with a high speed in solving. Assume that

$$H = \begin{bmatrix} g(\boldsymbol{\omega}_1^T \boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{\omega}_L^T \boldsymbol{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\boldsymbol{\omega}_1^T \boldsymbol{x}_m + b_1) & \cdots & g(\boldsymbol{\omega}_L^T \boldsymbol{x}_m + b_L) \end{bmatrix}$$, then the optimization objective of ELM is demonstrated [267]

as $\min_{\beta} \|H\beta - T\|^2 + \dfrac{C}{2}\|\beta\|^2$, where C is a penalty factor. The solution of this optimization objective of ELM can be solved as $\beta = (H^T H + CI)^\dagger H^T T$, among which $A^\dagger$ refers to the Moore-Penrose generalized inverse of the matrix A. This indicates that the parameters can be obtained via direct calculation by ELM according to $\beta = (H^T H + CI)^\dagger H^T T$, thus the complexity is significantly better than the Max-margin domain transforms [267]. Based on this description of ELM and according to [268], the flow chart of ELM-based transformed fault diagnosis is demonstrated in Figure 13b.
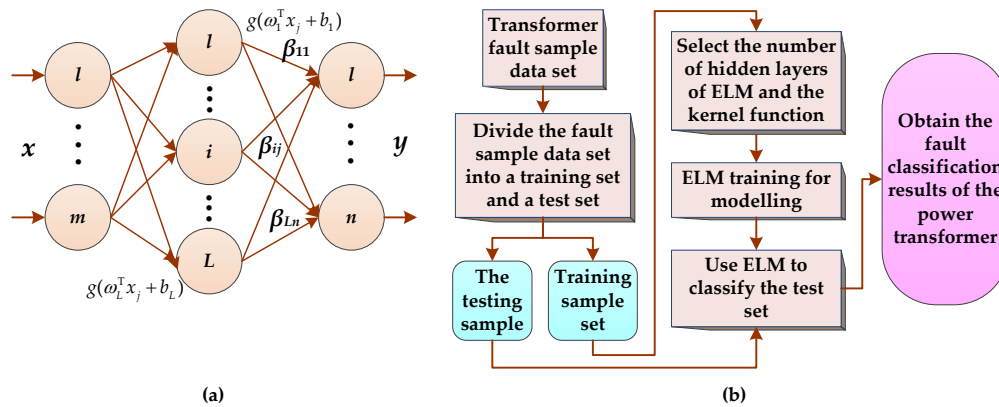
Figure 13. Illustration of ELM, among which (**a**) is the ELM network structure; and (**b**) shows the ELM-based transformer fault diagnosis flow chart.

Therefore, in view of the ELM-based transformer fault diagnosis using DGA, some research achievements have been achieved. In the process of fault diagnosis of the transformer, Malik and Mishra [203] applied principle component analysis using RapidMiner software to IEC TC10 and related databases to identify most relevant input variables for incipient fault classification of the power transformer. Thereafter, ELM is implemented to classify the incipient faults of power transformer and its performance is compared with fuzzy-logic and ANN. The compared results show that ELM can provide better diagnosis results with proposed input variables. Wang et al. [204] presented the optimization algorithm of integrated ensemble of online sequential ELM, which has been applied to transformer fault diagnosis using a limited number of sample data. The experimental results show that this developed model has better performance in response to online monitoring and real-time data processing. In addition, Yuan et al. [269] proposed an integrated PSO and ELM method for fault diagnosis of the power transformer based on DGA. Moreover, in order to overcome the deficiency of three-ratio method that fault diagnosis cannot be made due to missing ratio coding, Du et al. [205] combined ELM with the three-ratio method to diagnose the fault of the power transformer through making use of the good generalization performance of ELM. This ELM fault diagnosis model includes the ratio of coded information by taking the component content of the characteristic gases in the sample and the corresponding ratio of codes as the input of ELM. The diagnosis results show that the model is feasible and effective in transformer fault diagnosis.

(3) *Deep learning (DL):* it is derived from the study of neural networks, which can be known as a a deep layer neural network. DL, by establishing a hierarchical model structure similar to human brain, can conduct feature extraction of the data that needs trained from bottom layer to top layer to depict the intrinsic information rich in data, so that it can improve the accuracy of classification or prediction. Deep Auto-Encoder Network (DAEN) is a kind of DL method, in which the training samples are implemented feature transformation layer by layer through constructing a ML model with multiple hidden layers, and then the characteristic representation of samples in the original space is transformed to a new feature space, so that making the classification easier and finally improving the accuracy of the classification. DAEN is more capable of depicting the rich intrinsic information of the data, thus it is a hot spot in ML internationally [105]. On this basis, DL has been employed by more and more researchers in DGA-based transformer fault diagnosis. Mlakić et al. [70] employed the DL method and infrared imaging as a tool for transformer faults detection. Cui et al. [71] investigated a DL-DBN (deep belief network) and two BP artificial neural networks based on Matlab programming by using directly DGA and characteristic gas method in transformer oil chromatographic analysis. Shi et al. [105] firstly constructed a classified DAEN model, and employed the typical classified data set to analyze and verify the classification performance of this model. Then, combined with the on-line monitored data of DGA for power transformed, they proposed a method of transformed fault diagnosis based on the classified DAEN model, which can optimize the parameters of classified DAEN model by the pre-training with massive unlabeled samples and adjust them with a few labeled samples. The results of case analysis demonstrate that the proposed method has higher diagnosis accuracy than those based on the BPNN and the SVM. Besides, Shi and Zhu [206] employed the DL neural network to the fault diagnosis of the power transformers, which indicates that the proposed DL neural network can effectively utilize massive oil characteristic online monitoring unlabeled data and a small number of fault DGA data to accomplish data training, and then the fault diagnosis results are presented in probabilistic forms with a better fault judgment performance. Moreover, aiming at current transformer saturation classification issues, Ali et al. [207] employed a DL approach to develop an accurate current transformer saturation classification method based on unsupervised feature extraction and supervised fine-turning strategy. In this method, auto-encoders and deep neural networks are used to extract features automatically without prior knowledge of optimal features. Simulation results show that the method can classify the different levels of current transformer saturation with a remarkable accuracy and has unique feature extraction capabilities.

To sum up the survey in this section, ML methods have provided new ideas for the fault diagnosis of power transformers using DGA, and lots of substantive results have been achieved currently. However, there are still some problems to be solved. Firstly, although ML demonstrates a better performance in transformer fault diagnosis based on DGA, these achieved research results are mostly focused on fault classification while rarely involving fault location. ML can be combined with other intelligent algorithms for accurate location of the faults, but the location precision of them is not so perfect. Hence, it is one of research directions to accurately diagnose the fault of the transformer and locate it with a high precision in the future. Secondly, it may become possible to realize omnidirectional monitoring of the operation status of the power transformer in the future along with the high-speed development of computer information technology and AI. Hence, it is of great significance to investigate the intelligent ML methods by using the obtained power transformer operation state data samples, which can realize identification of new faults in the power transformer for high accuracy of diagnosis. This will be another research direction in future.

7.4. Other Intelligent Diagnosis Tools

In addition to the methods summarized in the previous sections in this chapter, some other intelligent methods as powerful diagnosis tools have been developed and applied to power transformer fault diagnosis using DGA. Among them, mathematical statistics method [270], WA [83,124–127], optimized neural network [208,209], BN [87,166–168], and evidential reasoning

approach [45,75,151,210–217] have already appeared and have been preliminarily applied in the DGA-based fault diagnosis of the power transformers, and they are briefly reviewed as follows:

(1) *Mathematical statistics methods:* they are a branch of mathematics that uses statistical methods to analyze data and derive its conceptual regularity (i.e., statistical laws), based on the theory of probability. The mainstream analysis methods include regression analysis, variance analysis, covariance analysis, clustering analysis, discriminant analysis, principal component analysis, etc. On this basis, Zou [270] developed a fuzzy fault diagnosis model for power transformer based on a new coding membership function, in which the probability theory and mathematical statistics are employed to analyze the distribution of true gas-in-oil volume fraction and its ratio. Based on the true ratio distribution of gas-in-oil volume fraction, a method combined with the three-ratio coding boundary is proposed to solve the coding membership function. In addition, the calculation methods of code-combination fuzzy set, fault fuzzy set and cut set are investigated. Two sets of historical chromatogram data are tested to verify this fuzzy diagnosis and the test results demonstrate that the diagnosis robustness and accuracy are both improved.

(2) *WA:* it is an emerging time-frequency analysis method. WA is seen as a breakthrough of Fourier transformation, by which the localization idea of window Fourier transformation is developed. The window width of WA decreases with the increase of frequency, thus it meets the high resolution requirement of the high-frequency signal. WA has good time-frequency localization characteristic and the ability of adaptive and multi-scale analysis to the signal, thus it is suitable for detecting transient anomalies occurring in normal signals and can demonstrate their components. At present, WA has been successfully applied to the fault signal analysis of electrical equipment and the vibration signal analysis of the mechanical equipment. In the fault diagnosis of transformers using DGA, WA has been adopted and some notable achievements have been made. Babu et al. [83] investigated the application of WA technique to transformer fault diagnosis using ANN. In the process of fault diagnosis, wavelets provide an efficient means of decomposing voltage and current signals to a detectable and discriminate features as it convolutes into different frequency components. In addition, Dong et al. [124] carried out fault diagnosis research for power transformers, by integrating the rough set and FWNN with the least square weighted fusion algorithm. Mao and Aggarwal [125] proposed a novel approach to the classification of the transient phenomenon in power transformers using combined wavelet transform and neural network. In this method, the wavelet transform is wavelet transform is employed to decompose the differential current signals of the power transformer into a series of detailed wavelet components whose spectral energies are calculated and used to train a neural network to discriminate an internal fault form the magnetizing inrush current. Similarly, the adaptive wavelet neural network (WNN) is adopted in [126,127] to distinguish between inrush and internal fault of the transformer. Considering the good time-frequency characteristics of WA, Li et al. [271] combined WA with neural network to form a new approach to fault diagnosis of power transformer, which can improve the efficiency and accuracy of fault diagnosis of the neural network system, and have achieved good results.

(3) *Optimized neural network.* A neural network can be optimized by some intelligent algorithms, such as PSO. To this end, the neural network is evolved by a modified PSO algorithm to form a new approach to power transformer fault diagnosis in [208]. This approach can overcome the problem of premature convergence observed in many applications of error BP algorithm and enhance the fault diagnostic ability of conventional DGA in power transformer. In addition, in order to improve the correct judgment rate in power transformer fault diagnosis, Jia et al. [209] investigated a DGA method of transformer via neural network based on PSO with neighborhood operator. In this method, some typical gases in transformer oil are selected as the input of neural network for training according to correlation analysis and data pre-treatment. After that, the neural network is trained and optimized so as to accomplish the fault diagnosis. The experimental results indicate that this developed method gains good classification result and can identify faults under the difficult situation where transformer overheating and partial discharge coexist. Moreover, this method shows a higher correct judgement rate.

(4) *BN:* it is a directed acyclic graph, which is briefly introduced as follows. Suppose that the nodes X_1, X_2, X_3 and X_4 in a BN refer to the random variables, and the directed arc between nodes represents the causal relationship between variables. Here, X_1, X_2 and X_3 are root nodes, and X_4 is a sub-node. Each node has the corresponding conditional probability table, that is, the root nodes X_1, X_2 and X_3 correspond to their respective marginal distribution $P(X_1)$, $P(X_2)$ and $P(X_3)$. The sub-node X_4 corresponds to its conditional probability distribution $P(X_4|X_1,X_2,X_3)$. Hence, by using the unique conditional independence of BN, the joint probability distribution can be simplified to

$$P(X_1,X_2,X_3,X_4)=\prod_{i=1}^{4}P\left(X_i|P_a(X_i)\right)=P(X_1)P(X_2)P(X_3)P(X_4|X_1,X_2,X_3) \quad [272], \quad \text{where} \quad P_a(X_i)$$

represents the parent node of X_i with i = 1, 2, 3, and 4. Based on the basic principle of BN, Wu et al. [87] proposed a novel method for transformer fault integrated diagnosis based on BN classifier. Wang et al. in [166] and [167] developed a new transformer fault diagnosis model based on rough set theory and BN, according to complementary strategy. In this model, the complexity of BN structure and difficulties of fault symptom acquisition are largely decreased based on the minimal rules. Meanwhile, probability reasoning can be realized by BN, which can be employed to describe changes of fault symptoms and analyze fault reasons of transformer. Analogously, Xie et al. [168] developed a transformer fault diagnosis model based on BN and rough set reduction theory, which is capable of dealing with missing information, embodies fault-tolerant feature and can achieve high accuracy. In addition, Bai et al. [272] established a three-layer BN by analyzing the causal relation of undesirable service conditions, fault modes and abnormal symptoms. In this network model, a BN reasoning method is employed to obtain the most probable explanation of the network, including the concurrent fault that he transformer possibly confronts, and the condition of abnormal symptoms that are not detected. This BN reasoning model provides an important basis for following diagnostic tests. Moreover, Zhu and Wu [273] conducted synthesized diagnosis on transformer faults based on BN, naive Bayesian classifier model, tree augmented naive Bayesian classifier model and BN augmented naive Bayesian. This approach uses the results of DGA attributes to classify power transformer's fault types. The computing tests are implemented on actual samples of transformer faults and the results show that the diagnosis performance of this proposed hybrid approach prevails that of separated BN based classifiers and the rough set based approach.

(5) *Evidential reasoning approaches:* they have been widely used in fault diagnosis of transformers using DGA. Concretely speaking, Qian et al. [75] proposed a case-based reasoning approach to power transformer fault diagnosis using DGA data, which has higher reliability and is more practical for the transformer incipient fault diagnosis. Moreover, Yang et al. [151] combined probability reasoning with fuzzy technique to identify power transformer malfunction. Ming et al. [210] developed an evidential reasoning approach to transformer fault diagnosis, which is effective in insulation diagnosis of transformer. In [211], the evidential reasoning approach, oil testing and DGA are used to implement transformer condition assessment. Based on a fuzzy reasoning method, Shi et al. [212] designed a transformer fault diagnosis EPS. In addition, based on the reasoning integration of rough set, fuzzy set and Bayesian optimal classifier, Su and Dong [213] developed a model for transformer fault diagnosis. Similarly, Qian et al. [214] developed a fault diagnosis method of power transformer, by integrating the case-based reasoning with fuzzy theory and neural network, through which satisfactory accuracy and well practically could be achieved. Liao et al. [215] developed an integrated decision-making model for condition assessment of power transformers using fuzzy approach and evidential reasoning method. Irungu et al. [216] developed an integrated fuzzy-evidential reasoning approach in fault identification of power transformers using DGA. The results show that the assessing model is capable of offering an overall evaluation of the observed transformer. In addition, Xie et al. [217] put forward a new diagnosis method based on fuzzy normal partition and logic reasoning for insulation fault of power transformer. In this method, fuzzy processing of the insulation diagnosis parameters are realized, and then the insulation diagnosis knowledge is acquired and the reasoning rules are built, and finally the reasoning results are obtained by applying reasoning in fault diagnosis. The method may improve

the reasoning efficiency. Moreover, it can increase the accuracy of fault diagnosis and maneuver ability by actual computation.

Some conclusions can be drawn in this section. First, SI algorithms, data mining technology, ML and other intelligent methods have been gradually used in transformer fault diagnosis using DGA by more and more researchers, and these intelligent methods have achieved good diagnostic results and produced great economic benefits; especially for the SI algorithms, WA, ML, etc., which shows the potential of application in the field of transformer fault diagnosis. Second, these intelligent methods are not isolated in application, and they can be combined with each other to achieve better diagnostic results in many cases. Third, although some intelligent methods have been successfully applied in the field of transformer fault diagnosis, the theory of these methods is not very mature, especially some applications are still in exploration and experimental stage, and there is still a certain distance from the actual engineering application. Therefore, it needs to be perfected in both theoretical research and engineering application. A summary of the application of other intelligent algorithms in DGA-based transformer fault diagnosis is presented in Table 11.

Table 11. A summary for the application of other intelligent algorithms in DGA based transformer fault diagnosis.

Other Intelligent Algorithms	Advantages and Disadvantages
Swarm intelligence algorithms - GA [67,68,196,222,223] - AIA [72,163,176] - ACO [228–234] - PSO [69,77,80] - BFO [238,239] - AFSO [197–199] - ABC [243] - FOA [245,246] - BOA [248] - Hybrid SI algorithms [219]	- GA: better global optimization ability - AIA: accelerated global optimal searching - ACO: higher diagnosis accuracy rate - PSO: a good global optimizer - BFO: identify maximum faults with minimum test cases - AFSO: fast optimization speed, global optimization ability and strong parallel processing capability - ABC: less control parameters, simple calculation and easy implementation - FOA: solve the problems of data source shortage of fault gases with high accuracy - BOA: obvious improvement in validity and accuracy, strong searching ability and fast convergence speed - Hybrid SI algorithms: high-speed searching, strong global optimization, and avoid falling into local optimality
Data mining technology - weighted fuzzy kernel clustering method [155] - dynamic clustering algorithm [79,81] - combined grey prediction and clustering analysis [63] - iterative self-organizing data analysis technique [81] - SVM-based multilevel binary tree [255] - information fusion technology [202]	- model the fault symptoms of transformer to find the law of describing interrelationship between operating status and external performance - beneficial to obtain intuitive relationship between operation state and result of fault analysis - high prediction accuracy - strong adaptability
ML methods - SVM [5,68,72,77,80,85,154,169,170,188,199] - ELM [203–205] - DL, DRL [70,71,105,206,207]	- SVM: enhance the role of empty rotation sample in the construction of classification hyperplane, and suppress deviation of hyperplane; powerful tool to deal with issues with small sampling, nonlinearity and high dimension - ELM: an emerging learning algorithm; without need of iterative optimization; fast learning speed and good generalization - DL/DRL: high accuracy of classification or prediction; unique feature extraction capability with a remarkable accuracy
Other intelligent diagnosis tools - mathematical statistics [270] - WA [83,124–127] - optimized neural network [208,209] - BN [87,166–168,272,273] - evidential reasoning approach [45,75,151,210–217]	- mathematical statistics methods: can improve the diagnosis robustness and accuracy - WA: good time-frequency localization characteristic and ability of adaptive and multi-scale analysis to signals; suitable for detecting transient anomalies - optimized neural network: can overcome the issue of premature convergence and enhance the fault diagnosis ability - BN: difficulty of fault symptom acquisition is largely decreased based on minimal rules

▪ Evidential reasoning approach: widely used; higher reliability and more practical for transformer incipient fault diagnosis

8. Discussion and Prospects

8.1. Discussion

The power transformer DGA is not affected by external electric fields and magnetic fields, which easily happen in an electrified state and online, thus it has become an effective method for fault diagnosis of oil-immersed power transformers. On this basis, the traditional methods such as the characteristic gas method, the three-ratio method, and the Rogers method have been developed. Combined with these, some AI methods such as EPS, ANN, fuzzy theory, GST, rough sets, SAIs, DL, SVM, ELM, and WA have been applied in transformer fault diagnosis. However, the traditional DGA methods have the main defects such as lack of coding, and over absolute coding limitation. Currently, the main issues in the field of power transformer fault diagnosis are as follows:

- More serious uncertainties and fuzziness among the fault phenomena, fault causes, fault mechanisms and fault classifications in the DGA data-based transformer fault diagnosis.
- The accuracy of fault diagnosis by DGA without experienced experts is not high.
- The complexity of electric power transformer fault is hard to overcome.
- Randomness and fuzziness in transformer fault diagnosis usually exist.
- Some intelligent fault diagnosis approaches are easily get stacked into the minimal value locally and strict requirement on the initial value which would make fault diagnosis difficult to some extent.
- The deficiency of three-ratio method that fault diagnosis cannot be made due to missing ratio coding is hard to overcome.
- The correct judgment rate in power transformer fault diagnosis is not high.
- Insulation condition assessment is usually performed by experts with special knowledge and experience due to the complexity of the transformer insulation structure and various degradation mechanisms under multiple stresses.
- Different orders of magnitude of the input variables in the network have an impact on the network convergence performance.
- The relationship between some fault causes and fault results in the transformer fault diagnosis system is not well-defined, as well as it cannot clearly determine which kinds of gases dissolved in oil cause even when a fault occurs.
- Relevant data samples of transformer fault diagnosis are hard to be obtained accurately.

Hence, due to their stronger fault diagnosis ability, the intelligent algorithms have achieved great success in the fault diagnosis of transformer based on DGA. The AI technique based transformer DGA methods have become increasingly mature and practiced in transformer fault diagnosis. However, it should also be seen that a single intelligent state detection method can only reflect the status of the transformer from one aspect, thus it has a different degree of malpractice. For several examples, the knowledge base acquisition and validation of EPS is rather difficult, and it has poor abilities of learning and fault tolerance, which largely limits the development of fault diagnosis EPS; ANN has the defects of slow convergence speed, easy oscillation and easy to fall into local optimum; SVM is essentially a two-classifier algorithm, thus it has the disadvantages of constructing the learner and the low efficiency of classification in solving multi-class issues, and the kernel function selection and parameter determination are rather difficult; ELM has a fast training speed, but it has worse stability. In brief, the main advantages, existing problems and development trend of the main intelligent techniques and methods summarized in this paper in the DGA data based transformer fault diagnosis are presented in Table 12.

Table 12. A brief summary of the main advantages, existing problems and development trend of the main intelligent techniques and methods summarized in this paper in transformer fault diagnosis using DGA data.

Main Intelligent Techniques and Methods	Advantages	Existing Problems	Development Trend
EPS [1]	▪ Comprehensive analysis and judgement for transformer operating status ▪ Early diagnosis of the internal potential faults in the operating transformer ▪ Effectively simulate the process of fault diagnosis by the fault diagnosis experts	▪ Hard to obtain the complete knowledge base ▪ Hard to verify the completeness of the knowledge ▪ Generally does not have learning ability ▪ Has a diagnostic error or no result sometimes ▪ Has a poor fault tolerance ▪ Troublesome maintenance	▪ Combined with self-learning and self-association ANN ▪ Combined with fuzzy theory based reasoning method under uncertainty ▪ Combined with Petri nets ▪ Combined with RST
ANN [2]	✓ Strong learning ability ✓ Self-organization of knowledge ✓ Strong fault-tolerant ability ✓ Easy to parallel processing ✓ A faster speed of execution	✓ A slow convergence speed ✓ Network structure and parameter are determined by experiences ✓ Hard to obtain optimal network structure for a specific problem	✓ Improved BPNN ✓ Combined with other AI technologies ✓ Improved FWNN ✓ Improved GRNN ✓ Improved EDA-ANN
Fuzzy theory	▪ Adapt to deal with uncertainty ▪ Fuzzy knowledge base is closer to man's expression habits ▪ Can deal with critical value coding defects in IEC/IEEE three-ratio	▪ Hard to determine appropriate membership function for the input and output variables ▪ Fuzzy rule requires the sample data to have completeness	▪ Improved fuzzy reasoning models ▪ Combined with other intelligent algorithms, such as SI algorithms ▪ Combined with GRA
RST [3]	✓ Effectively deal with incomplete, imprecise and inconsistent information ✓ Does not need any prior knowledge ✓ An EPS knowledge base with attributes reduction ✓ Better fault tolerance ability ✓ Simplicity, practicality	✓ A single RST can hardly show good performance in fault data processing if it is not combined with other intelligent techniques and algorithms	✓ Combined with EPS ✓ Combined with fuzzy theory ✓ Combined with ML ✓ Combined with data driven ✓ Combined with ANN ✓ Integrated with BN and Petri nets
GST [4]	▪ Can effectively extract valuable information ▪ Has better ability to judge accurately for some difficult faults, such as dampness	▪ Misjudgment occurs with normal DGA data sometimes	▪ Combined with clustering technique ▪ Combined with entropy theory ▪ Combined with some intelligent optimization approaches
SI algorithms	✓ Strong ability of searching ✓ Global optimization ability ✓ The cooperative individuals are distributed in a population ✓ No centralized control, has robustness, expandability ✓ Individual is simple and easy to implement	✓ Easy to fall into the local optimum ✓ The convergence rate is unstable ✓ The optimization precision is not high	✓ Improvement of algorithm performance ✓ Setting of algorithm parameters ✓ Combined with other algorithms
DMT [5]	▪ Effective in fault symptom modelling ▪ Beneficial to obtain the direct connection between transformer operation state and fault analysis results ▪ Predictive accuracy is high ▪ Strong adaptability	▪ The process of setting up of data sources is complex	▪ Combined with ML, especially DL, ELM, and SVM ▪ Integrated with other intelligent diagnosis tools, such as evidential reasoning, fuzzy theory, RST, and

			ANN
ML [6]	✓ Fast learning ability, and good generalization performance, such as ELM ✓ Can improve the precision of classification and predictive, such as DL	✓ Less involving the location of faults ✓ Hard to achieve a strict analysis	✓ Integrated with some advanced computer information technique ✓ Combined with data mining technology ✓ Combined with other intelligent fault diagnosis tools and methods

[1] EPS = expert system; [2] ANN = artificial neural network; [3] RST = rough sets theory; [4] GST = grey system theory; [5] DMT = data mining technology; [6] ML = machine learning.

Therefore, combined with Table 12 above, the shortcomings of transformer fault diagnosis using intelligent algorithms is mainly reflected in the following three aspects:

- Most of the existing intelligent diagnosis methods only diagnose the fault types of the transformer separately, without consideration of some of the inherent connections between various faults. In addition, some of them are not very mature and still in the stage of exploration and experiment, which will inevitably affect the results of fault diagnosis of the transformer using DGA.
- Due to the cumulative effect of dissolved gas in oil and the effect of its error on sampling, the current intelligent diagnosis method of transformer fault based on DGA data indicates a larger error of diagnosis when the gas content is less, and it needs people to judge the existence of the fault in advance, which is no doubt harmful to the diagnosis of the potential fault.
- In the actual operation of the transformer, there are a lot of incomplete or imperfect data of dissolved gas in oil, thus it is difficult to implement intelligent diagnosis according to these data.

According to the issues above, in view of the shortage of single intelligent fault diagnosis method, on the one hand, it can be improved from the aspect of algorithm, that is, multiple intelligent algorithms are integrated to form a compounded network in which the algorithms are complementary. For example, ANN, GA and EPS can be combined closely to develop an intelligent fault diagnosis system with comprehensive diagnosis ability, meeting the requirements of improving transformer safety and economic operation level, thus it can be seen as an important direction for the development of fault detection and diagnosis technique of the power transformers. In addition, the improved PSO is integrated with fuzzy neural network in [274] to form fault diagnosis strategy for transformer oil chromatography monitoring, which is beneficial to balance the relationship between the local searching and global searching of the BP neural network, thus avoiding it falling into a local optimum.

On the other hand, it can be improved from the angle of transformer detection means. When the fault occurs in transformer or the transformer has a potential fault, in addition to the change of dissolved gas in oil, the mechanical vibration and electrical properties of the transformer will also be changed, thus it is necessary to extract the feature data with reasonable detection methods, and then combine these characteristic data with the DGA data in a reasonable manner, in order to find the best transformer fault diagnosis method.

8.2. Prospects

For the existing problems in transformer fault diagnosis discussed in Section 8.1, how to conduct high-precision and high-accuracy fault diagnosis for the transformer based on imperfect DGA data requires immediate solutions. To address this, new ML algorithms provide a new idea for transformer fault diagnosis using DGA data. A lot of substantial results have been achieved in this aspect at present. ML needs to be built based on a good knowledge representation system, which has achieved a long-term development in the past 30 years. The defects of traditional ML theoretical framework have been found and determined gradually, and new ML theoretical

framework has been proposed continuously [256,257]. In recent years, ML has developed rapidly, especially when AlphaGo was developed in 2016. Since then, the multi-layer ANN based DL as perception together with Markov decision process based RL as decision form a pair of golden components. In [257], Li et al. put forward a novel theoretical framework of ML, called parallel ML, based on the parallel system, which can employ the parallel virtual system to generate massive virtual samples for ML. This provides a significant research direction in transformer fault diagnosis using DGA data. It needs to note that these new ML theories are still in theoretical stage, thus they may have some defects that have not been found. Therefore, it is essential to combine them with actual engineering applications and improve them continuously, which will be of great significance to improve the intelligent levels of fault diagnosis of the power transformer using DGA.

Besides, the generative adversarial net (GAN) [275] can also be considered, which is able to automatically produce massive simulation model data via constructing a Max-min adversarial game system. This can solve the small sample size problems in the real environment to a large extent. From AlphaGo [276], AlphaGoZero [277], AlphaZero [278], and parallel system [257] to GAN [276], scientists have been looking for ways to solve the data sample issues of the ML. The obstacles to the improvement of ML intelligence have been gradually removed. ML has been developed from the known training sample set (limited small data) to the era of acquiring massive imaginary training samples (infinite large data) via self-exploration. This is a watershed in AI that transcends human intelligence. Hence, it will be a promising future for the DGA-based transformer fault diagnosis through applying the emerging ML methods and GAN technique.

9. Conclusions

This paper presents a detailed overview on the application status of intelligent methods in fault diagnosis of the oil-immersed power transformers based on DGA, including EPS, ANN, fuzzy theory, RST, GST, SI algorithms, data mining technology, ML, and other intelligent methods. These intelligent methods provide an idea for high-precision transformer fault diagnosis. The main contributions can be summarized as follows:

(1) The application of these intelligent methods compensates for the shortcomings of the traditional DGA method, and improves the fault diagnosis ability and diagnostic accuracy of the system. Through the analysis of the principle, characteristics, effectiveness and feasibility of these intelligent diagnosis methods, the merits and defects of them are demonstrated, as well as their improvement schemes. This provides a reference for the researchers to choose the optimal approach to fault diagnosis of the oil-immersed power transformer. It is considered that the application of AI technology to power transformer fault diagnosis is determined by the characteristics of AI and the importance of power system fault diagnosis. It is the inevitable choice for the development of power system. Finally, the intelligent diagnosis method of transformer fault based on DGA is prospected, and the future development direction is analysed.

(2) Years of operation practice have proved that the online monitoring technology of dissolved gases in transformer oil can diagnose, predict and track the development trends of faults, but it has some major defects such as coding deficiencies, excessive coding boundaries and critical value criterion defects. A single intelligent algorithm can meet the requirements of fault diagnosis under certain conditions, but inevitably will have some limitations. To address this, on the one hand this can be improved from the aspect of the algorithm, that is, by combining the traditional DGA methods with multiple AI algorithms to constitute a compound network in which the algorithms are complementary, and further to develop a novel composite intelligent algorithm, which will be the main direction of the future development of transformer fault diagnosis technology, and will have potential practical value and broad application prospect. On the other hand, it can be improved from the angle of transformer detection means. Concretely, when a fault occurs in transformer or the transformer has a potential fault, the mechanical vibration and electrical properties of the transformer will change, in addition to the change of dissolved gas in oil, thus it is necessary to extract the

feature data with reasonable detection methods. These data then are combined with the DGA data in a rational manner, in order to find the best fault diagnosis method for the transformer.

(3) In the future, it will be very promising for developing new intelligent comprehensive fault diagnosis systems through introducing new ML theories and frameworks, the new DL based on multi-layer ANN, and the GAN to fault diagnosis of the transformer based on DGA. Such systems can automatically identify and delete bad data in some cases, with better real-time capability and self-adaptation. Besides, they should have the function of self-organization, self-learning, associations and memories, and continuous innovation in the operation. This system will have a very good prospect of application and it is of great significance to the realization of high-precision transformer fault diagnosis and fault location.

(4) Combined with the survey made in this paper, and the status of transformer fault diagnosis in practice, several suggestions are given as follows: (a) we should collect a large number of existing examples of power transformer fault diagnosis in practice to build up an abundant and perfect knowledge base and case database through sorting and analyzing; (b) combine multiple intelligent algorithms with existing diagnosis methods to make full use of detection and experimental data for comprehensive diagnosis, so as to improve the comprehensive diagnosis capability of the system, and make the diagnostic conclusion of the system more instructive to the maintenance of the transformer; (c) enhance the reliability and openness of the diagnostic system, thus the knowledge and experience gained by the maintenance personnel in practice can conveniently extend and modify the knowledge base of the system so as to improve the diagnosis accuracy of the system; (d) speed up the development of online detection technology to achieve diagnosis online by the diagnostic system, so as to improve the level of automation of the diagnostic system; and (e) fully understand the merits and defects of various intelligent methods in power system fault diagnosis, and then integrate them with conventional IEC/IEEE three-ratios to develop an intelligent comprehensive diagnosis system, in which the comprehensive complementarity between the advantages of these intelligent methods are continuously realized to improve the security and economy of the transformer.

(5) This paper presents a detailed and systematic survey on various intelligent methods applied in faults diagnosing and decisions making of the oil-immersed power transformers, by thoroughly investigating their merits and demerits. Moreover, their improvement schemes and future development trends are demonstrated. The research summary, empirical generalization and analysis of predicament in this paper can provide thoughts and suggestions for the research of complex power grid in the new environment, as well as references and guidance for researchers to choose the optimal approach to fault diagnosis and decision making of the large oil-immersed power transformers using DGA in preventive electrical tests.

Acknowledgments: The authors gratefully acknowledge the support of the National Natural Science Foundation of China (Grant. 51477055 & 51777078), and the Key Science and Technology Projects of China Southern Power Grid (CSGTRC-KY2014-2-0018).

Author Contributions: Lefeng Cheng and Tao Yu conducted the survey, that is, the investigation of the application status of EPS, ANN, fuzzy theory, RST, GST, and other intelligent algorithms in fault diagnosis of the power transformer using DGA. Lefeng Cheng wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Nomenclature

DGA	dissolved gas analysis	BN	Bayesian network
ANN	artificial neural network	GRA	grey relational analysis
EPS	expert system	HST	hot spot temperature
RST	rough sets theory	AIA	artificial immune algorithm
GST	grey system theory	DC	dynamic clustering
BPA	basic probability assignment	WA	wavelet analysis

HAE	hydrogen-acetylene-ethylene	ELM	extreme learning machine
TD	thermal-discharge	DL	deep learning
AI	artificial intelligence	SI	swarm intelligence
SVM	support vector machine	ACO	ant colony optimizer
MLP	multi-layer perceptron	BFO	bacterial foraging optimization
PSO	particle swarm optimizer	AFSO	artificial fish swarm optimizer
T2-FLS	type-2 fuzzy logic system	ABC	artificial bee colony
RBF	radial basis function	FOA	firefly optimization algorithm
BP	back propagation	BOA	bat optimization algorithm
BPNN	back propagation neural network	SGA	standard genetic algorithm
GRNN	generalized regression neural network	WNN	wavelet neural network
GA	genetic algorithm	FPA	flower pollination algorithm
GM	grey model	ML	machine learning
FWNN	fuzzy wavelet neural network	RL	reinforcement learning
LM	Levenberg-Marguardt algorithm	IICA	improved imperialist competitive algorithm
EDA	estimation of distribution algorithm	DAEN	DeepAuto-Encoder network
FPN	fuzzy petri nets	DL-DBN	deep learning-deep belief network
RBFNN	radial basis function neural network	GAN	generative adversarial net

References

1. Wang, Y.Y.; Gong, S.L.; Grzybowski, S. Reliability evaluation method for oil-paper insulation in power transformers. *Energies* **2011**, *4*, 1362–1375.

2. Sang, Z.X.; Mao, C.X.; Lu, J.M.; Wang, D. Analysis and simulation of fault characteristics of power switch failures in distribution electronic power transformers. *Energies* **2013**, *6*, 4246–4268.

3. Sun, C.X.; Chen, W.G.; Li, J.; Liao, R.J. *Online Monitoring and Fault Diagnosis of Dissolved Gas in Electrical Equipment Oil*; Science Press: Beijing, China, 2003; pp. 5–10.

4. Tenbohlen, S.; Coenen, S.; Djamali, M.; Müller, A.; Samimi, M.H.; Siegel, M. Diagnostic measurements for power transformers. *Energies* **2016**, *9*, 347.

5. Bacha, K.; Souahlia, S.; Gossa, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electr. Power Syst. Res.* **2012**, *83*, 73–79.

6. Chen, W.G.; Chen, X.; Peng, S.Y.; Li, J. Canonical correlation between partial discharges and gas formation in transformer oil paper insulation. *Energies* **2012**, *5*, 1081–1097.

7. Xiang, C.M.; Zhou, Q.; Li, J.; Huang, Q.D.; Song, H.Y.; Zhang, Z.T. Comparison of dissolved gases in mineral and vegetable insulating oils under typical electrical and thermal faults. *Energies* **2016**, *9*, 312.

8. Haroldo, D.F., Jr.; João, G.S.C; Jose, L.M.O. A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. *Renew. Sustain. Energy Rev.* **2015**, *46*, 201–209.

9. Faiz, J.; Soleimani, M. Dissolved gas analysis evaluation in electric power transformers using conventional methods a review. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 1239–1248.

10. Ministry of Electric Power, People's Republic of China. *DL/T722-2000 Guides for Analysis and Judgement of Dissolved Gases in Transformer Oil*; China Electric Power Press: Beijing, China, 2000; pp. 1–10.

11. Liu, X.L.; Mo, Y.L. Analysis and treatment of abnormality of characteristic gas contents caused by bad breath in electric locomotive main transformer. *Transformer* **2013**, *50*, 64–67.

12. Chen, W.G.; Zhou, Q.; Gao, T.Y.; Su, X.P.; Wan, F. Pd-Doped SnO_2-based sensor detecting characteristic fault hydrocarbon gases in transformer oil. *J. Nanomater.* **2013**, *1*, 2527–2531.

13. Wan, F.; Chen, W.G.; Peng, X.J.; Shi, J. Study on the gas pressure characteristics of photoacoustic spectroscopy detection for dissolved gases in transformer oil. In Proceedings of the 2012 IEEE International Conference on High Voltage Engineering and Application, Shanghai, China, 17–20 September 2012; pp. 286–289.

14. Chen, X.; Chen, W.G.; Gan, D.G. Properties and gas production law of surface discharge in transformer oil-paper insulation. In Proceedings of the IEEE 2010 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, West Lafayette, IN, USA, 17–20 October 2010; pp. 1–4.

15. Liang, Y.L.; Li, K.J.; Zhao, J.G.; Niu, L.; Ren, J.G. Research on the dynamic monitoring cycle adjustment strategy of transformer chromatography on-line monitoring devices. *Proc. CSEE* **2014**, *34*, 1446–1453.

16. Zeng, W.L.; Yang, Y.F.; Gan, C.Y.; Liu, G. Study on intelligent development of power transformer on-line monitoring based on the data of DGA. In Proceedings of the IEEE Computer Society Asia-Pacific Power and Energy Engineering Conference, Wuhan, China, 25–28 March 2011; pp. 1–4.

17. Nogami, T.; Yokoi, Y.; Ichiba, H.; Atsumi, Y. Gas discrimination method for detecting transformer faults by neural network. *Electr. Eng. Jpn.* **1995**, *115*, 93–103.

18. Jiang, X.Q.; Gong, Y.; Han, S.; Zhou, K. Application of the improved three-ratio method in chromatographic analysis of locomotive transformer oil. *Adv. Mater. Res.* **2014**, *1030–1032*, 29–33.

19. Yang, T.F.; Liu, P.; Lu, L.J.; Yi, H. New fault diagnosis method of power transformer by combination of FCM and IEC three-ratio method. *High Volt. Eng.* **2007** *33*, 66–70.

20. Dhote, N.K.; Helonde, J.B. Diagnosis of power transformer faults based on five fuzzy ratio method. *Wseas Trans. Power Syst.* **2012**, *7*, 114-125.

21. Xu, W.J.; Ruan, J.G.; Song, B. Application of grey correlation analysis method in transformer fault diagnosis with missing code in three-ratio method. *Bull. Sci. Technol.* **2017**, *33*, 129–164.

22. Liu, Z.X.; Song, B.; Li, E.W.; Mao, Y.; Wang, G.L. Study of code absence in the IEC three-ratio method of dissolved gas analysis. *IEEE Electr. Insul. Mag.* **2015**, *31*, 6–12.

23. Zhang, W.H.; Yuan, J.S.; Wang, S.; Zhang, K. A calculation method for transformer fault basic probability assignment based on improved three-ratio method. *Power Syst. Prot. Control* **2015**, *43*, 115–121.

24. Zhang, W.H.; Yuan, J.S.; Zhang T.F.; Zhang, K. An improved three-ratio method for transformer fault diagnosis using B-spline theory. *Proc. CSEE* **2014**, *34*, 4129–4136.

25. Xu, K.J. The problem when using three-ratio method to judge the fault in dissolved gas analysis of transformer. *Transformer* **2010**, *47*, 75–76.

26. Cao, D.K. *Gas Analysis Diagnosis and Fault Inspection in Transformer Oil*; China Electric Power Press: Beijing, China, 2005; pp. 10–25.

27. Electric Power Industry Standard of the People's Republic of China. *Guide for Analysis and Judgment of Gas in Transformer Oil*; China Electric Power Press: Beijing, China, 2003; pp. 2–6.

28. BOI Standard. *IEC 60599-1999-Mineral Oil-Impregnated Electrical Equipment in Service—Guide to the Interpretation of Dissolved and Free Gases Analysis*; International Electrotechnical Commission: Geneva, Switzerland, 2006.

29. Transformers Committee of the IEEE Power Engineering Society. *IEEE C57.104-1991-IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*; IEEE: New York, NY, USA, 1993.

30. Singh, S.; Bandyopadhyay, M.N. Dissolved gas analysis technique for incipient fault diagnosis in power transformers: A bibliographic survey. *IEEE Electr. Insul. Mag.* **2010**, *26*, 41–46.

31. Liu, N; Tan, K.X.; Gao, W.S. Fault diagnosis method for power transformers based on patterns of dissolved gases in the oil. *J. Tsinghua Univ.* **2003**, *43*, 301–303.

32. Xie, R.B.; Xue, J.; Zhang, L.; Shen J.; Li, Q.H.; Lei, Y.; Zhao L.H. Transformer fault diagnosis and application based on oil chromatography analysis. *Guangdong Electr. Power* **2017**, *8*, 117–121.

33. Zhao, X.T.; Shen, Q.; Xu, J.X.; Yang, C.; Liu, L. Fault diagnosis for power transformer based on IEC three-ratio and extension association with combined weights. *Power Syst. Clean Energy* **2013**, *29*, 18-22.

34. Waghmare, H.V.; Kulkarmi, H.H. Modeling of transformer DGA using IEC & fuzzy based three gas ratio method. *Int. J. Eng. Res. Technol.* **2014**, *3*, 1149–1152.

35. Cai, J.D.; Wang, S.F. An improved three-ratio fault diagnosis based on rough set theory for power transformer. *Adv. Technol. Electr. Eng. Energy* **2004**, *23*, 8–12.

36. Cai, J.D.; Wang, S.F. Application of decision rules for IEC-60599 three-ratio fault diagnosis based on rough set theory. *Proc. CSEE* **2015**, *25*, 134–139.

37. Zhang, Y.; Ding, X.; Liu, Y.L.; Griffin, P.J. An artificial neural network approach to transformer fault diagnosis. *IEEE Trans. Power Deliv.* **1996**, *11*, 1836–1841.

38. Colorado, D.; Hernández, J.A.; Rivera, W.; Martinez, H.; Juárez, D. Optimal operation conditions for a single-stage heat transformer by means of an artificial neural network inverse. *Appl. Energy* **2011**, *88*, 1281–1290.

39. Bhalla, D.; Bansal, R.K.; Gupta, H.O. Function analysis based rule extraction from artificial neural networks for transformer incipient fault diagnosis. *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 1196–1203.

40. Lin, J.; Sheng, G.H.; Yan, Y.J.; Dai, J.J.; Jiang, X.C. Prediction of dissolved gas concentrations in transformer oil based on the KPCA-FFOA-GRNN model. *Energies* **2018**, *11*, 225, doi:10.3390/en11010225.

41. Yi, J.H.; Wang, J.; Wang, G.G. Improved probabilistic neural networks with self-adaptive strategies for transformer fault diagnosis problem. *Adv. Mech. Eng.* **2016**, *8*, 1–13.

42. Meng, K.; Dong, Z.Y.; Wang, D.H.; Wong, K.P. A self-adaptive RBF neural network classifier for transformer fault analysis. *IEEE Trans. Power Syst.* **2010**, *25*, 1350–1360.

43. Castro, A.R.G.; Miranda, V. Knowledge discovery in neural networks with application to transformer failure diagnosis. *IEEE Trans. Power Syst.* **2005**, *20*, 717–724.

44. Miranda, V.; Castro, A.R.G. Improving the IEC table for transformer failure diagnosis with knowledge extraction from neural networks. *IEEE Trans. Power Deliv.* **2005**, *20*, 2509–2516.

45. Peng, H.; Wang, J.; Pérez-Jiménezc, M.J.; Wang, H.; Shao, J.; Wang, T. Fuzzy reasoning spiking neural P system for fault diagnosis. *Inf. Sci.* **2013**, *235*, 106–116.

46. 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. *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 1346–1353.

47. Lin, C.E.; Ling, J.M.; Huang, C.L. An expert system for transformer fault diagnosis using dissolved gas analysis. *IEEE Trans. Power Deliv.* **1993**, *8*, 231–238.

48. Wang, Z.; Liu, Y.; Griffin, P.J. A combined ANN and expert system tool for transformer fault diagnosis. *IEEE Trans. Power Deliv.* **2000**, *13*, 1224–1229.

49. Saha, T.K.; Purkait, P. Investigation of an expert system for the condition assessment of transformer insulation based on dielectric response measurements. *IEEE Trans. Power Deliv.* **2004**, *19*, 1127–1134.

50. Mani, G.; Jerome, J. Intuitionistic fuzzy expert system based fault diagnosis using dissolved gas analysis for power transformer. *J. Electr. Eng. Technol.* **2014**, *9*, 2058–2064.

51. Li, J.P.; Chen, X.J.; Wu, C.M. Application of comprehensive relational grade theory in expert system of transformer fault diagnosis. In Proceedings of the IEEE International Workshop on Intelligent Systems and Applications, Wuhan, China, 23–24 May 2009; pp. 1–4.

52. Huang, Y.C.; Yang, H.T.; Huang, C.L. Developing a new transformer fault diagnosis system through evolutionary fuzzy logic. *IEEE Trans. Power Deliv.* **1997**, *12*, 761–767.

53. Islam, S.M.; Wu, T.; Ledwich, G. A novel fuzzy logic approach to transformer fault diagnosis. *IEEE Trans. Dielectr. Electr. Insul.* **2000**, *7*, 177–186.

54. Xu, W.; Wang, D.; Zhou, Z.; Chen, H. Fault diagnosis of power transformers: Application of fuzzy set theory, expert systems and artificial neural networks. *IEE Proc.* **1997**, *144*, 39–44.

55. Fan, J.M.; Wang, F.; Sun, Q.Q.; Bin, F.; Liang, F.W.; Xiao, X.Y. Hybrid RVM-ANFIS algorithm for transformer fault diagnosis. *IET Gener. Transm. Distrib.* **2017**, *11*, 3637–3643.

56. Yu, J.; Zhou, R. Transformer fault diagnosis based on neural network and fuzzy theory. *J. Cent. South Univ.* **2013**, *44*, 243–247.

57. Naresh, R.; Sharma, V.; Vashisth, M. An integrated neural fuzzy approach for fault diagnosis of transformers. *IEEE Trans. Power Deliv.* **2008**, *23*, 2017–2024.

58. Youssef, O.A.S. Applications of fuzzy-logic-wavelet-based techniques for transformers inrush currents identification and power systems faults classification. In Proceedings of the IEEE PES Power Systems Conference and Exposition, 10–13 October 2004; Volume 1, pp. 553–559.

59. Cheng, L.; Yu, T.; Wang, G.; Yang, B.; Zhou, L. Hot spot temperature and grey target theory-based dynamic modelling for reliability assessment of transformer oil-paper insulation systems: A practical case study. *Energies* **2018**, *11*, 249.

60. Dong, M.; Yan, Z.; Taniguchi, Y. Fault diagnosis of power transformer based on model-diagnosis with grey relation. In Proceedings of the 7th IEEE International Conference on Properties and Applications of Dielectric Materials, Nagoya, Japan, 1–5 June 2003; Volume 3, pp. 1158–1161.

61. Song, B.; Yu, P.; Luo, Y.B.; Wen, X.S. Fault diagnosis for power transformer based on grey relation entropy. *Autom. Electr. Power Syst.* **2005**, *29*, 76–79.

62. Chang, W.P.; Zhao, B.; Li, J.H. Application of grey relational theory in transformer fault diagnosis. In Proceedings of the 2011 IEEE Asia-Pacific Power and Energy Engineering Conference, Wuhan, China, 25–28 March 2011; pp. 1–3.

63. Lin, C.H.; Chen, J.L.; Huang, P.Z. Dissolved gases forecast to enhance oil-immersed transformer fault diagnosis with grey prediction-clustering analysis. *Expert Syst.* **2011**, *28*, 123–137.

64. Song B, Ping Y, Luo Y, et al. Study on the fault diagnosis of transformer based on the grey relational analysis. In Proceedings of the IEEE International Conference on Power System Technology, Kunming, China, 13–17 October 2002; Volume 4, pp. 2231–2234.

65. Zeng, F.; Cheng, X.; Guo, J.C.; Tao, L.; Chen, Z.X. Hybridising human judgment, AHP, grey theory, and fuzzy expert systems for candidate well selection in fractured reservoirs. *Energies* **2017**, *10*, 447.

66. Dinmohammadi, A.; Shafiee, M. Determination of the most suitable technology transfer strategy for wind turbines using an integrated AHP-TOPSIS decision model. *Energies* **2017**, *10*, 642.

67. Pan, C.; Chen, W.; Yun, Y. Fault diagnostic method of power transformers based on hybrid genetic algorithm evolving wavelet neural network. *IET Electr. Power Appl.* **2008**, *2*, 71–76.

68. Fei, S.W.; Zhang, X.B. Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Syst. Appl.* **2009**, *36*, 11352–11357.

69. 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 Trans. Power Deliv.* **2008**, *23*, 751–759

70. Mlakić, D.; Nikolovski, S.; Majdandžić, L. Deep learning method and infrared imaging as a tool for transformer faults detection. *J. Electr. Eng.* **2018**, *6*, 98–106.

71. Cui, Q.M.; Liang, K.; Gao, H.Y.; Chen, G.D.; Wang, N.Y.; Jin, S.Y.; Zhong, C.Q.; Cui, S.T.; Sun, D.J.; Fang, J.; et al. Research of the transformer fault diagnosis expert system based on ESTA and deep learning neural network programmed in Matlab. In Proceedings of the International Conference on Civil, Transportation and Environment, Hsinchu, Taiwan, 27–29 May 2016; pp. 772–778.

72. Zheng, R.R.; Zhao, J.Y.; Zhao, T.T.; Lim M. Power transformer fault diagnosis based on genetic support vector machine and gray artificial immune algorithm. *Proc. CSEE* **2011**, *31*, 56–63.

73. Lee, B.E.; Park, J.W.; Crossley, P.A.; Kang, Y.C. Induced voltages ratio-based algorithm for fault detection, and faulted phase and winding identification of a three-winding power transformer. *Energies* **2014**, *7*, 6031–6049.

74. Yang, Q.; Su, P.Y.; Chen, Y. Comparison of impulse wave and sweep frequency response analysis methods for diagnosis of transformer winding faults. *Energies* **2017**, 10, 431.

75. Qian, Z.; Gao, W.S.; Wang, F.; Yan, Z. A case-based reasoning approach to power transformer fault diagnosis using dissolved gas analysis data. *Int. Trans. Electr. Energy Syst.* **2009**, *19*, 518–530.

76. Xia, F.R.; Zhu, Y.L.; Gao, Y. Shape-space based negative selection algorithm and its application on power transformer fault diagnosis. In the Proceedings of the IEEE International Conference on Robotics and Biomimetics, Sanya, China, 15–18 December 2017; pp. 2149–2154.

77. Fei, S.W.; Wang, M.J.; Miao, Y.B.; Tu, J.; Liu, C.L. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. *Energy Convers. Manag.* **2009**, *50*, 1604–1609.

78. Ghoneim, S.S.M.; Taha, I.B.M. A new approach of DGA interpretation technique for transformer fault diagnosis. *Int. J. Electr. Power Energy Syst.* **2016**, *81*, 265–274.

79. Xiong, H. Zhang, X.X.; Liao, R.J.; Chang, T.; Sun, C.X. Fault diagnosis of power transformer using dynamic clustering algorithm. *Chin. J. Sci. Instrum.* **2007**, *28*, 710–714.

80. Liao, R.; Zheng, H. Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electr. Power Syst. Res.* **2011**, *81*, 2074–2080.

81. Zheng, X.L.; Wang, S.Q.; Cao, Y.Z.; Transformer fault diagnosis based on improved fuzzy ISODATA algorithm. In Proceedings of the 2014 IEEE International Conference on Power System Technology, Chengdu, China, 20–22 October 2014; pp. 1279–1286.

82. Rigatos, G.; Siano, P. Power transformers' condition monitoring using neural modelling and the local statistical approach to fault diagnosis. *Int. J. Electr. Power Energy Syst.* **2016**, *80*, 150–159.

83. Prasanth Babu, B.; Surya Kalavathi, M.; Singh, B.P. Use of wavelet and neural network (BPFN) for transformer fault diagnosis. In Proceedings of the 2006 IEEE Conference on Electrical Insulation and Dielectric Phenomena, Kansas City, MO, USA, 15–18 October 2006; pp. 93–96.

84. Shah, A.M.; Bhalja, B.R. Fault discrimination scheme for power transformer using random forest technique. *IET Gener. Transm. Distrib.* **2015**, *10*, 1431–1439.

85. Shah, A.M.; Bhalja, B.R. Discrimination between internal faults and other disturbances in transformer using the support vector machine-based protection scheme. *IEEE Trans. Power Deliv.* **2013**, *28*, 1508–1515.

86. Hong, K.; Huang, H.; Zhou, J. Winding condition assessment of power transformers based on vibration correlation. *IEEE Trans. Power Deliv.* **2015**, *30*, 1735–1742.

87. Wu, L.Z.; Zhu, Y.L.; Yuan, J.S. Novel method for transformer faults integrated diagnosis based on Bayesian network classifier. *Trans. China Electrotech. Soc.* **2005**, *20*, 45–51.

88. Abu-Siada, A.; Hmood, S. A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis. *Int. J. Electr. Power Energy Syst.* **2015**, *67*, 401–408.

89. Illias, H.A.; Xin, R.C.; Bakar, A.H.A. Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. *Measurement* **2016**, *90*, 94–102.

90. Pandya, A.A.; Parekh, B.R. Interpretation of sweep frequency response analysis (SFRA) traces for the open circuit and short circuit winding fault damages of the power transformer. *Int. J. Electr. Power Energy Syst.* **2014**, *62*, 890–896.

91. Zhang, Y.Y.; Liu, J.F.; Zheng, H.B.; Wei, H.; Liao, R.J. Study on quantitative correlations between the ageing condition of transformer cellulose insulation and the large time constant obtained from the extended Debye model. *Energies* **2017**, *10*, 1842.

92. Jürgensen, J.H.; Nordström, L.; Hilber, P. Individual failure rates for transformers within a population based on diagnostic measures. *Electr. Power Syst. Res.* **2016**, *141*, 354–362.

93. Babita Jain, B.; Srinivas, M.B.; Jain, A. A novel web based expert system architecture for on-line and off-line fault diagnosis and control (FDC) of transformers. In Proceedings of the Joint International Conference on Power System Technology and IEEE Power India Conference, New Delhi, India, 12–15 October 2008; pp. 1–5.

94. Yang, Q.P.; Xue, D.; Fu, Y.; Lan, Z.D. Application of expert system in transformer fault diagnosis. *Transformer* **1996**, *2*, 35–38.

95. Yao, C.G.; Liao, R.J.; Cheng, Y.Y.; Sun, C.X. Object oriented knowledge base for transformer fault diagnosis. *Adv. Technol. Electr. Eng. Energy* **2001**, *20*, 61–66.

96. Ma, D.L.; Zhang, W.J.; Yao, W. Establish expert system of transformer fault diagnosis based on dissolved gas in oil. In Proceedings of the 2013 IEEE International Conference on Information Science and Cloud Computing Companion, Guangzhou, China, 7–8 December 2013; pp. 681–685.

97. Su, H.C.; Sun, X.F.; Si, D.J. A RS approach to founding and maintaining ES knowledge base for fault diagnosis of power transformer. *Proc. CSEE* **2002**, *22*, 32–35.

98. Li, J.C.; Zhou, N.; Lv, B. Transformer DGA diagnosis expert system based on neural network and fuzzy theory. *Power Syst. Technol.* **2006**, *30*, 125–128.

99. Du, J.G. Transformer Fault Diagnosis Expert System. Master's Thesis, North China Electric Power University, Beijing, China, 2003; pp. 15–26.

100. Flores, W.C.; Mombello, E.E.; Corvo, A.M. Expert system for the assessment of power transformer insulation condition based on type-2 fuzzy logic systems. *Expert Syst. Appl.* **2011**, *38*, 8119–8127.

101. Ranga, C.; Chandel, A.K.; Chandel, R. Expert system for condition monitoring of power transformer using fuzzy logic. *J. Renew. Sustain. Energy* **2017**, *9*, 3199–3208.

102. Žarković, M.; Stojković Z. Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics. *Electr. Power Syst. Res.* **2017**, *149*, 125–136.

103. Tu Y, Zheng Q. DGA based insulation diagnosis of power transformer via ANN. In Proceedings of the 6th International Conference on Properties and Applications of Dielectric Materials, Xi'an, China, 21–26 June 2000; Volume 1, pp. 133–136.

104. Segatto, E.C.; Coury, D.V. A power transformer protection with recurrent ANN saturation correction. In Proceedings of the IEEE Power Engineering Society General Meeting, San Francisco, CA, USA, 12–16 June 2005; Volume 2, pp. 1341–1346.

105. Shi, X.; Zhu, Y.L.; Ning, X.G.; Wang, L.W.; Sun, G.; Chen, G.Q. Transformer fault diagnosis based on deep auto-encoder network. *Electr. Power Autom. Equip.* **2016**, *36*, 122–126.

106. Sun, Y.J.; Zhang, S.; Miao, C.X.; Li, J.M. Improved BP neural network for transformer fault diagnosis. *Int. J. Min. Sci. Technol.* **2007**, *17*, 138–142.

107. Ding, S.; Chang, X.H.; Wu, Q.H.; Yang, Y.L. Study of transformer fault diagnosis based on GRNN and DGA. *Electr. Meas. Technol.* **2014**, *37*, 142–146.

108. Shi, D.Y.; Buse, J. Wu, Q.H.; Jiang, J.; Xue, Y.S. Fast identification of power transformer magnetizing inrush currents based on mathematical morphology and ANN. In Proceedings of the IEEE Power and Energy Society General Meeting, Detroit, MI, USA, 24–29 July, 2011; pp. 1–6.

109. Ghunem, R.A.; El-Hag, A.; Assaleh K. Prediction of furan content in transformer oil using artificial neural networks (ANN). In Proceedings of the 2010 IEEE International Symposium on Electrical Insulation, San Diego, CA, USA, 6–9 June, 2010; pp. 1–4.

110. Malik, H.; Tarkeshwar, M.K.; Yadav, A.K.; Kr Anil, B. Application of physical-chemical data in estimation of dissolved gases in insulating mineral oil for power transformer incipient fault diagnosis with ANN. *Int. J. Comput. Appl.* **2012**, *41*, 43–50.

111. Wang, H.X.; Yang, Q.P.; Zheng, Q.M. Artificial neural network for transformer insulation aging diagnosis. In Proceedings of the IEEE Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, Nanjing, China, 6–9 April 2008; pp. 2233–2238.

112. Zakaria, F.; Johari, D.; Musirin, I. The Taguchi-artificial neural network approach for the detection of incipient faults in oil-filled power transformer. In Proceedings of the 2013 IEEE 7th International Power Engineering and Optimization Conference, Langkawi, Malaysia, 3–4 June 2013; pp. 518–522.

113. Chen, M.J.; Zeng, X.; Li, G.H.; A new transformer protection based on the artificial neural network model. In Proceedings of the World Automation Congress, Hawaii, HI, USA, 28 September–2 October 2008; pp. 1–4.

114. Fu, Y. A neural network approach to power transformer fault diagnosis. *J. Shanghai Inst. Electr. Power* **1998**, *14*, 1–7.

115. Wang, X.M.; Li, W.S.; Yan, Z. Study on fault diagnosis for power transformer based on BP neural network. *High Volt. Technol.* **2005**, *31*, 12–14.

116. Liu, C.P. Expert System Research of Power Transformer Fault Diagnosis Based on DGA. Master's Thesis, Guangxi University, Nanning, China, 2007; pp. 29–42.

117. Zhao, J.Y.; Zheng, R.R.; Li, J.P. Transformer fault diagnosis based on homotopy BP algorithm. In Proceedings of the IEEE 9th International Conference on Electronic Measurement & Instruments, Beijing, China, 16–19 August 2009; Volume 4, pp. 622–626.

118. Sugumaran, V.; Sabareesh, G.R.; Ramachandran, K.I. Fault diagnosis of roller bearing using kernel based neighborhood score multi-class support vector machine. *Expert Syst. Appl.* **2008**, *34*, 3090–3098.

119. Tripathy, M.; Maheshwari, R.P.; Verma, H.K. Power transformer differential protection based on optimal probabilistic neural network. *IEEE Trans. Power Deliv.* **2010**, *25*, 102–112.

120. Wang, S.F.; Cai, J.D. Application of hybrid algorithm based on GA-BP in transformer diagnosis using GAS chromatographic method. *High Volt. Eng.* **2003**, *29*, 3–6.

121. De, A.; Chatterjee, N. Recognition of impulse fault patterns in transformers using Kohonen's self-organizing feature map. *IEEE Trans. Power Deliv.* **2002**, *17*, 489–494.

122. Dong, L.X.; Xiao, D.M.; Liu, Y.L. Rough set and radial basis function neural network based insulation data mining fault diagnosis for power transformer. *J. Harbin Inst. Technol.* **2007**, *14*, 263–268.

123. Zhang, J.; Pan, H.; Huang, H. Electric power transformer fault diagnosis using OLS based radial basis function neural network. In Proceedings of the IEEE International Conference on Industrial Technology, Chengdu, China, 21–24 April 2008; pp. 1–4.

124. Dong, L.X.; Xiao, D.M.; Liang, Y.S.; Liu, Y.L. Rough set and fuzzy wavelet neural network integrated with least square weighted fusion algorithm based fault diagnosis research for power transformers. *Electr. Power Syst. Res.* **2008**, *78*, 129–136.

125. Mao, P.L.; Aggarwal, R.K. A novel approach to the classification of the transient phenomena in power transformers using combined wavelet transform and neural network. *IEEE Trans. Power Deliv.* **2001**, *16*, 654–660.

126. Li, H.F.; Wang, G.; Li, X.H.; Hu, S.P. Distinguish between inrush and internal fault of transformer based on adaptive wavelet neural network. *Proc. CSEE* **2005**, *25*, 144–150.

127. Gong, M.; Zhang, X.; Gong, Z. Study on a new method to identify inrush current of transformer based on wavelet neural network. In Proceedings of the IEEE 2011 International Conference on Electrical and Control Engineering, Yichang, China, 16–18 September 2011; pp. 848–852.

128. Wu, H.Q.; Zhou, N.N.; Wang, C.Y. Research for transformer fault diagnosis and MATLAB simulation based on RBF neural networks. *Sci. Technol. Eng.* **2010**, *10*, 1249–1275.

129. Cardoso, G.; Rolim J.; Zurn H.H. Application of neural-network modules to electric power system fault section estimation. *IEEE Trans. Power Deliv.* **2004**, *19*, 1034–1041.

130. Ghanizadeh, A.J.; Gharehpetian, G.B. ANN and cross-correlation based features for discrimination between electrical and mechanical defects and their localization in transformer winding. *IEEE Trans. Dielectr. Electr. Insul.* **2014**, *21*, 2374–2382.

131. Zou, K.X.; Zhao, W.; Xu, W.D.; Yin, W.J. EDA-ANN based transformer fault recognition with dissolved gas. In Proceedings of the 2013 IEEE Conference on Innovative Smart Grid Technologies-Asia, Bangalore, India, 10–13 November 2013; pp. 1–5.

132. Patel, N.K.; Khubchandani, R.K. ANN based power transformer fault diagnosis. *J. Inst. Eng.* **2004**, *85*, 60–63.

133. Vanamadevi, N.; Arivamudhan, M.; Santhi, S. Detection and classification of impulse faults in transformer using wavelet transform and artificial neural network. In Proceedings of the IEEE International Conference on Sustainable Energy Technologies, Singapore, 24–27 November 2008; pp. 72–76.

134. Wu, Y.; Yu, Q.; Luo, R.C.; Chen, C.; Wang, F.F. FAHP and ANN in power transformer risk assessment. *J. Changsha Univ. Sci. Technol.* **2012**, *19*, 982–987.

135. Nashruladin, K.N. *Application of ANN and GA for Transformer Winding/Insulation Faults*; University Teknologi Petronas: Perak, Malaysia, 2007; pp. 5–12.

136. Zhang, J. Application of super SAB ANN model for transformer fault diagnosis. *Trans. China Electrotech. Soc.* **2004**, *19*, 49–54.

137. Leondes, C.T. *Fuzzy Theory Systems: Techniques and Applications*; Academic Press: Pittsburgh, PA, USA, 1999; pp. 1–30.

138. Zhang, J.J.; Liang, Y.S.; Yin, Y.J.; Guo, C.X. Transformer fault diagnosis based on fuzzy theory and support vector machine. *J. Electr. Power Sci. Technol.* **2011**, *26*, 61–66.

139. Zeng, J.Z.; Xu, Y.Z. On the method for transformer fault diagnosis based on fuzzy neural network. *J. Fujian Comput.* **2004**, *25*, 27–30.

140. Su, Q.; Mi, C.; Lai, L.L.; Austin, P. A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer. *IEEE Trans. Power Syst.* **2000**, *15*, 593–598.

141. Li, J.; Sun, C.X.; Liao, R.J.; Zhou, Q. Study on analysis method about fault diagnosis of transformer and degree of grey incidence based on fuzzy clustering. *Chin. J. Sci. Instrum.* **2004**, *25*, 587–589.

142. Ma, H.Z.; Li, Z.; Ju, P.; Han, J.D.; Zhang, L.M. Diagnosis of power transformer faults based on fuzzy three-ratio method. In Proceedings of the IEEE 7th International Power Engineering Conference, Singapore, 29 November–2 December 2005; pp. 162–186.

143. Wang, J.Y.; Ji, Y.C. Application of fuzzy Petri nets knowledge representation in electric power transformer fault diagnosis. *Proc. CSEE* **2003**, *23*, 121–125.

144. Sun, N. Review of fault diagnosis method for power transformer based on rough set theory. *Guangdong Electr. Power* **2010**, *23*, 14–17.

145. Li, J.; Sun, C.X.; Chen, W.G.; Zhou, Q.; Du, L. A method of synthesis based on the grey cluster and fuzzy cluster about internal fault diagnosis of transformer. *Proc. CSEE* **2003**, *23*, 112–115.

146. Chen, D.; Cui, D.W.; Li, X.; Wang, Z.R. A weighted fuzzy clustering algorithm and its application to fault diagnosis of power transformer. *J. Xian Univ. Technol.* **2008**, *24*, 182–186.

147. Lee, J.P.; Lee, D.J.; Kim, S.S.; Ji, P.S.; Lim, J.Y. Dissolved gas analysis of power transformer using fuzzy clustering and radial basis function neural network. *J. Electr. Eng. Technol.* **2007**, *2*, 157–164.

148. Tang, S.P.; Peng, G.; Zhong, Z.X. An improved fuzzy C-means clustering algorithm for transformer fault. In Proceedings of the 2016 China International Conference on Electricity Distribution, Xi'an, China, 10–13 August 2016.

149. Rajamani, P.; Dey, D.; Chakravorti, S. Cross-correlation-aided fuzzy C-means for classification of dynamic faults in transformer winding during impulse testing. *Electr. Mach. Power Syst.* **2010**, *38*, 1513–1530.

150. Duan, H.D.; Wang, Z.L.; Zhou, Z.X.; Liu, W.B. Probabilistic neural network for fault diagnosis of power transformer based application on fuzzy input. *Coal Min. Mach.* **2007**, *28*, 190–192.

151. Yang, L.; Shang, Y.; Zhou, Y.F.; Yan, Z. Probability reasoning and fuzzy technique applied for identifying power transformer malfunction. *Proc. CSEE* **2000**, *20*, 19–23.

152. Fu, Y.; Jiang, Y.R.; Cui, C.H.; Cao, J.L. Fuzzy theory and probability reasoning applied for identifying power transformer fault. *High Volt. Eng.* **2008**, *34*, 1040–1044.

153. Aghaei, J.; Gholami, A.; Shayanfar, H.A.; Dezhamkhooy, A. Dissolved gas analysis of transformers using fuzzy logic approach. *Int. Trans. Electr. Power Syst.* **2010**, *20*, 630–638.

154. Sima, L.P.; Shu, N.Q.; Zuo, J.; Wang, B.; Peng, H. Concentration prediction of dissolved gases in transformer oil based on grey relational analysis and fuzzy support vector machines. *Power Syst. Prot. Control* **2012**, *76*, 381–401.

155. Fu, Y.; Tian, Z.N.; Jiang, Y.R.; Cao, J.L. Power transformer fault diagnosis using weighted fuzzy Kernel clustering. *High Volt. Eng.* **2010**, *36*, 371–373.

156. Mo, J.; Wang, X.; Dong, M.; Yan, Z. Diagnostic model of insulation faults in power equipment based on rough set theory. *Proc. CSEE* **2004**, *24*, 162–167.

157. Su, P.; Yuan, J.S.; An, X.L.; Li, Z.; Shen, T. Diagnosis method of transformer faults based on rough set theory. *Electr. Power Sci. Eng.* **2008**, *24*, 56–59.

158. Yuan, J.S.; Su, P.; Li, Z.; Yao, M. Diagnosis model of transformer faults based on a new heuristic reduction algorithm. *J. North China Electr. Power Univ.* **2008**, *35*, 12–17.

159. Xiang, X.J. The study of rough set theory in fault diagnosis expert system of transformer. *Bull. Sci. Technol.* **2003**, *19*, 288–291.

160. Zuo, Y.G. Fault diagnosis model based on rough set theory and expert system. In Proceedings of the IEEE International Colloquium on Computing, Communication, Control, and Management, Sanya, China, 8–9 August 2009; pp. 498–500.

161. Yu, X.D.; Ma, F.Y.; Zang, H.Z. Rough sets theory and artificial neural networks applied in the transformer fault diagnosis. *Relay* **2006**, *34*, 10–14.

162. Zhang, J.M.; Xiao, Q.H.; Wang, S.S. Transformer fault diagnosis by combination of rough set and neural network. *High Volt. Eng.* **2007**, *33*, 122-125.

163. Li, W.W.; Huang, H.X.; Wang, C.H.; Tang, H.Z. synthetic fault diagnosis method of power transformer based on rough set theory and improved artificial immune network classification algorithm. In Proceedings of the IEEE Computer Society Fourth International Conference on Natural Computation, Jinan, China, 18–20 October 2008; pp. 676–681.

164. Xiong, H.; Li, W.G.; Chang, G.H.; Guo, H.M. Application of fuzzy rough set theory to power transformer faults diagnosis. *Proc. CSEE* **2008**, *28*, 141–147.

165. Wang, Z.Y. A fault diagnosis method for power transformer based on rough set and fuzzy rules. *Relay* **2006**, *34*, 37–40.

166. Wang, Y.Q.; Lv, F.C.; Li, H.M. Synthetic fault diagnosis method of power transformer based on rough set theory and Bayesian network. *Proc. CSEE* **2006**, *26*, 137–141.

167. Wang, Y.Q.; Lu, F.C.; Li, H.M. Synthetic fault diagnosis method of power transformer based on rough set theory and Bayesian network. In Proceedings of the International Symposium on Neural Networks: Advances in Neural Networks, Part II, Beijing, China, 24–28 September 2008; pp. 498–505.

168. Xie, Q.J.; Zeng, H.X.; Ruan, L.; Chen, X.M.; Zhang, H.L. Transformer fault diagnosis based on Bayesian network and rough set reduction theory. In Proceedings of the IEEE Tencon Spring Conference, Sydney, Australia, 17–19 April 2013; pp. 262–266.

169. Jiang, Y.J.; Ni, Y.P. Fault diagnosis method of transformer based on rough set and support vector machine. *High Volt. Eng.* **2008**, *34*, 1755–1760

170. Wu, Z.L.; Yang, J.; Zhu, Y.L.; Liu, Z. Power transformer fault diagnosis based on rough set theory and support vector machines. *Power Syst. Prot. Control* **2010**, *38*, 80–83.

171. Wang, N.; Lv, F.C.; Liu, Y.P.; Li, H.M. Study on application of Petri nets model of transformer fault diagnosis based on decision table reduction. *Trans. China Electr. Soc.* **2003**, *18*, 88–92.

172. Wang, N.; Lv, F.C.; Liu, Y.P.; Li, H.M. Synthetic fault diagnosis of oil-immersed power transformer based on rough set theory and fuzzy Petri nets. *Proc. CSEE* **2003**, *23*, 127–132.

173. Zhou, A.H.; Yao, Y.; Song, H.; Zeng, X.H. Power transformer fault diagnosis based on integrated of rough set theory and evidence theory. In Proceedings of the IEEE Computer Society 2013 Third International Conference on Intelligent System Design and Engineering Applications, Hong Kong, China, 16–18 January 2013; pp. 1049–1052.

174. Shu, H.C.; Hu, Z.J.; Sun, S.Y.; Yang, Q. The fault diagnosis algorithm for transformer based on Extenics and rough set theory. In Proceedings of the IEEE Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, Nanjing, China, 6–9 April 2008; pp. 1269–1272.

175. Fei, S.W.; Sun, Y. Fault prediction of power transformer by combination of rough sets and grey theory. *Proc. CSEE* **2008**, *28*, 154–160.

176. Song, S.M.; Wang, Y.N.; Yao, S.X.; Wang, M. An approach to the transformer faults diagnosing based on rough set and artificial immune system. In Proceedings of the IEEE Chinese Conference on Pattern Recognition, Beijing, China, 22–24 October 2008; pp. 1–5.

177. Deng, J.L. Control problems of grey systems. *Syst. Control Lett.* **1982**, *1*, 288–294.

178. Deng, J.L. The grey control system. *J. Huazhong Univ. Sci. Technol.* **1982**, *10*, 11–20.

179. Deng, J.L. Grey system review. *World Sci.* **1983**, *5*, 1–5.

180. Deng, J.L. *Grey Control System*, 2nd ed.; Huazhong University of Science and Technology Press: Wuhan, China, 1985; pp. 5–15.

181. Deng, J.L. Properties of grey forecasting models GM (1, 1). *J. Huazhong Univ. Sci. Technol.* **1987**, *15*, 3–8.

182. Tan, X.R.; Deng, J.L. Grey relational analysis: A new statistical method of multifactorial analysis in medicine. *J. Pharm. Anal.* **1997**, *9*, 59–65.

183. Tan, X.R.; Deng, J.L.; Liu, S.F.; Pan, H.X. Multi-stratum medical grey relational theory and its application research. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Montreal, QC, Canada, 7–10 October 2007; pp. 3932–3937.

184. Liu, S.F. Emergence and development of grey system theory and its forward trends. *J. Zhejiang Wanli Univ.* **2003**, *16*, 14–17.

185. Li, J.; Sun, C.X.; Chen, W.G.; Chen, G.Q.; Cui, X.M. Study on fault diagnosis of insulation of oil-immersed transformer based on grey cluster theory. *Trans. China Electr. Technol. Soc.* **2002**, *17*, 24–29.

186. Song, Q.; Li, F.; Zhang, Y.S.; Xu, L. Research on fault diagnosis of power transformer based on grey entropy relational algorithm. *Transformer* **2010**, *47*, 56–58.

187. Li, S.; Zhao, F. Transformer fault diagnosis method based on weighted degree of Grey incidence of optimized entropy. *Transformer* **2013**, *50*, 48–51.

188. Zhang, Y.Y.; Wei, H.; Liao, R.J.; Wang, Y.Y.; Yang, L.J.; Yan, C.Y. A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers. *J. Electr. Eng. Technol.* **2017**, *12*, 830–839.

189. Qian, G.C.; Yu, H.; Zou, D.X.; Xu, X.W.; Liu, G.Q. Vibration detection method of transformer winding looseness based on entropy weight correlation theory. *Southern Power Syst. Technol.* **2016**, *10*, 45-51, 78.

190. Li, J.P. Study on Power Transformer Fault Diagnosis Technology Based on Dissolved Gases Analysis. Master's Thesis, Jilin University, Changchun, China, 2008; pp. 5–15.

191. Zhao, F.; Li, S. Fault diagnosis for traction transformer based on DGA and improved grey correlation. *High Volt. Appar.* **2015**, *51*, 41–45.

192. Xu, H.J.; Wang, Z.Y.; Su, H.Y. Dissolved gas analysis based feedback cloud entropy model for power transformer fault diagnosis. *Power Syst. Prot. Control* **2013**, *41*, 115–119.

193. Wang, Q.Z.; Wang, X.X. Unified grey relational analysis on transformer DGA fault diagnosis. *Open Mech. Eng. J.* **2014**, *8*, 129–131.

194. Liu, J.F.; Zheng, H.B.; Zhang, Y.Y.; Wei, H.; Liao, R.J. Grey relational analysis for insulation condition assessment of power transformers based upon conventional dielectric response measurement. *Energies* **2017**, *10*, 1526.

195. Zhou, L.W.; Wang, Y.Y.; Wang, F.; Yan, C.Y.; Bi, J.G. A transformer fault diagnosis method based on grey relational analysis and integrated weight determination. In Proceedings of the 2017 IEEE Electrical Insulation Conference, Baltimore, MD, USA, 11–14 June 2017; pp. 491–494.

196. Zhang, K.F.; Yuan, F.; Guo, J.; Wang, G.P. A novel neural network approach to transformer fault diagnosis based on momentum-embedded BP neural network optimized by genetic algorithm and fuzzy C-means. *Arab. J. Sci. Eng.* **2016**, *41*, 3451–3461.

197. Yu, H.; Wei, J.; Li, J. Transformer fault diagnosis based on improved artificial fish swarm optimization algorithm and BP network. In Proceedings of the 2010 2nd IEEE International Conference on Industrial Mechatronics and Automation, Wuhan, China, 30–31 May 2010; pp. 99–104.

198. Geng, C.; Wang, F.H.; Su, L.; Zhang, J. Parameter identification of Jiles-Atherton model for transformer based on hybrid artificial fish swarm and shuffled frog leaping algorithm. *Proc. CSEE* **2015**, *35*, 4799–4807.

199. Yu, H.; Wei, J.; Wang, D.D.; Sun, P. *The Application of Improved Artificial Fish Swarm and Support Vector Machine in Transformer Fault Diagnosis*: *Advances in Technology and Management*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 299–308.

200. Hu, H.S.; Qian, S.X.; Wang, J.; Shi, Z.J. Application of information fusion technology in the remote state on-line monitoring and fault diagnosing system for power transformer. In Proceedings of the 8th IEEE International Conference on Electronic Measurement and Instruments, Xi'an, China, 16–18 August 2007; pp. 550–555.

201. Li, Y.W.; Li, W.; Han, X.D.; Li, J. Application of multi-sensor information fusion technology in the power transformer fault diagnosis. In Proceedings of the 2009 International Conference on Machine Learning and Cybernetics, Hebei, China, 12–15 July 2009; pp. 29–33.

202. Gong, Y.S.; Zhang, B. The fault diagnosis model of transformer which based on the technology of information fusion. In Proceedings of the Ninth IEEE International Conference on Hybrid Intelligent Systems, Shenyang, China, 12–14 August 2009; pp. 184–187.

203. Malik, H.; Mishra, S. Extreme learning machine based fault diagnosis of power transformer using IEC TC10 and its related data. In Proceedings of the IEEE India Conference, New Delhi, India, 17–20 December 2015; pp. 1–5.

204. Wang, L.L.; Pei, F.; Zhu, Y.L. Transformer fault diagnosis based on online sequential extreme learning machine. *Appl. Mech. Mater.* **2015**, *721*, 360–365.

205. Du, W.X.; Zhao, X.P.; Du, H.L.; Lv, F. Fault diagnosis of power transformer based on extreme learning machine. *J. Shandong Univ. Sci. Technol.* **2017**, *36*, 29–36.

206. Shi, X.; Zhu, Y.L. Application of deep learning neural network in fault diagnosis of power transformer. *Electr. Power Constr.* **2015**, *36*, 116–122.

207. Ali, M.; Son, D.H.; Kang, S.H.; Nam, S.R.; Sciubba, E. An accurate CT saturation classification using a deep learning approach based on unsupervised feature extraction and supervised fine-tuning strategy. *Energies* **2017**, *10*, 1830.

208. Wang, X.X.; Tao, W. Power transformer fault diagnosis based on neural network evolved by particle swarm optimization. *High Volt. Eng.* **2008**, *34*, 2362–2367.

209. Jia, R.; Xu, Q.H.; Li, H.; Liu, W. Power transformer fault diagnosis via neural network based on particle swarm optimization with neighborhood operator. *High Volt. Appar.* **2008**, *51*, 8–10.

210. Dong, M.; Yan, Z.; Yang, L.; Judd, M.D. An evidential reasoning approach to transformer fault diagnosis. *Proc. CSEE* **2006**, *26*, 106–114.

211. Irungu, G.K.; Akumu, A.O.; Munda, J.L. Transformer condition assessment using dissolved gas analysis, oil testing and evidential reasoning approach. In Proceedings of the 2015 Electrical Insulation Conference, Seattle, WA, USA, 7–10 June 2015; pp. 145–149.

212. Shi, J.P.; Tong, W.G.; Wang, D.L. Design of the transformer fault diagnosis expert system based on fuzzy reasoning. In Proceedings of the IEEE International Forum on Computer Science-Technology and Applications, Chongqing, China, 25–27 December 2009; pp. 110–114.

213. Su, H.S.; Dong, H.Y. Transformer fault diagnosis based on reasoning integration of rough set and fuzzy set and Bayesian optimal classifier. *WSEAS Trans. Circuits Syst.* **2009**, *8*, 136–145.

214. Qian, Z.; Yan, Z.; Luo, C.Y. Fault diagnosis method of power transformer by integrating case-based reasoning with fuzzy theory and neural network. *High Volt. Eng.* **2001**, *27*, 1–2+5.

215. Liao, R.J.; Zheng, H.B.; Grzybowski, S.; Yang, L.J.; Zhang, Y.Y.; Liao, Y.X. An integrated decision-making model for condition assessment of power transformers using fuzzy approach and evidential reasoning. *IEEE Trans. Power Deliv.* **2011**, *26*, 1111–1118.

216. Irungu, G.K.; Akumu, A.O.; Munda, J.L. Comparison of IEC 60599 gas ratios and an integrated fuzzy-evidential reasoning approach in fault identification using dissolved gas analysis. In Proceedings of the 51st International Universities' Power Engineering Conference (UPEC), Coimbra, Portugal, 6–9 September 2016; pp. 1–6.

217. Xie, L.; Zhou, L.; Tong, X.J.; Chen, M.Y. Fault diagnosis of power transformer insulation based on fuzzy normal partition and logic reasoning. In Proceedings of the IEEE International Conference on Machine Learning and Cybernetics, Hong Kong, China, 19–22 August 2007; pp. 1081–1085.

218. Wang, S.H.; Zhang, Y.D.; Ji, G.L. Survey on theories and applications of swarm intelligence algorithms. *J. Nanjing Norm. Univ.* **2014**, *14*, 31–38.

219. Li, S.; Yuan, Z.G.; Wang, C.; Chen, T.E.; Guo, Z.C. Optimization of support vector machine parameters based on group intelligence algorithm. *CAAI Trans. Intell. Syst.* **2018**, *13*, 70–84.

220. Holland, J.H. *Adaptation in Natural and Artificial Systems*, 2nd ed.; MIT Press: Cambridge, UK, 1992; pp. 126–137.

221. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

222. Chen, W.G.; Yun, Y.X. Fault Diagnosis of power transformers based on genetic algorithm evolving wavelet neural network. *Power Syst. Autom.* **2007**, *31*, 88–92.

223. Mahvi, M.; Behjat, V. Localizing low-level short-circuit faults on the windings of power transformers based on low-frequency response measurement of the transformer windings. *IET Electr. Power Appl.* **2015**, *9*, 533–539.

224. Shi, J.G.; Chen, G.; Gao, Z. Survey on artificial immune algorithm. *Softw. Guide* **2008**, *7*, 68–69.

225. Yuan, J.S.; Lu, W.; Li, Z. Artificial immune algorithm for fault diagnosis of power transformer. In Proceedings of the IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, Wuhan, China, 21–22 December 2008; pp. 352–354.

226. Dorigo, M.; Gambardella, L.M. Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Trans. Evol. Comput.* **1997**, *1*, 53–66.

227. Dorigo, M. Stützle, T. *Ant Colony Optimization*; MIT Press: Cambridge, UK, 2004; pp. 1–10.

228. Mo, L.L. Transformer fault diagnosis method based on support vector machine and ant colony. *Adv. Mater. Res.* **2013**, *659*, 54–58.

229. Liu, Q.; Huang, G.Q.; Mao, C.; Shang, Y.; Wang, F. Recognition of dissolved gas in transformer oil by ant colony optimization support vector machine. In Proceedings of the IEEE International Conference on High Voltage Engineering and Application, Chengdu, China, 19–22 September 2016; pp. 1–4.

230. Niu, W.; Xu, L.F.; Hu, S.G. Fault diagnosis method for power transformer based on ant colony-SVM classifier. In Proceedings of the 2nd IEEE International Conference on Computer and Automation Engineering, Singapore, 26–28 February 2010; pp. 629–631.

231. Tian, B.B.; Liu, N.; Liu, K. The transformer diagnosis data reduction based on improved ant colony algorithm. *Power Syst. Prot. Control* **2011**, *39*, 96–99.

232. Li, A.H.; Sun, X.Y.; Liu, D.H. Application of RBF network with ant colony and fisher ratio to fault diagnosis of oil-immersed transformer. In Proceedings of the IEEE Seventh International Conference on Natural Computation, Shanghai, China, 26–28 July 2011; pp. 549–551.

233. Wei, L.Y.; Cui, X. Study on power transformer fault test sequence optimization based on multi-colony ant colony algorithm. In Proceedings of the IEEE Computer Society Third International Conference on Intelligent Human-Machine Systems and Cybernetics, Zhejiang, China, 26–27 August 2011; pp. 127–130.

234. Qiao, W.D. Transformer intelligent breakdown diagnosis system based on ant colony algorithm. *Jiangsu Electr. Appar.* **2007**, *14*, 34–37.

235. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, the University of Western Australia, Perth, Australia, 27 November–1 December 1995; pp. 1942–1948.

236. Lee, T.F.; Cho, M.Y.; Shieh, C.S.; Fang, F.M. Particle swarm optimization-based SVM application: Power transformers incipient fault syndrome diagnosis. In Proceedings of the IEEE Computer Society International Conference on Hybrid Information Technology, Cheju, Korea, 9–11 November 2006; pp. 468–472.

237. Cheng, S.F.; Cheng, X.H.; Yang, L. Application of wavelet neural network with improved particle swarm optimization algorithm in power transformer fault diagnosis. *Power Syst. Prot. Control* **2014**, *42*, 37–42.

238. Geethanjali, M.; Kannan, V.; Anjana, A.V.R. Bacterial foraging optimization algorithm trained ANN based differential protection scheme for power transformers. In Proceedings of the 2nd International Conference on Swarm, Evolutionary, and Memetic Computing, Part II, Visakhapatnam, India, 19–21 December 2011; pp. 267–277.

239. Gopila, M.; Gnanambal, I. An effective detection of inrush and internal faults in power transformers using bacterial foraging optimization technique. *Circuits Syst.* **2016**, *7*, 1569–1580.

240. Li, X.L.; Shao, Z.J.; Qian, J.X. An optimizing method based on autonomous animals: Fish-swarm algorithm. *Syst. Eng. Theory Pract.* **2002**, *22*, 32–38.

241. Li, X.L.; Qian, J.X. Studies on artificial fish swarm optimization algorithm based on decomposition and coordination techniques. *J. Circuits Syst.* **2003**, *8*, 1–6.

242. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization*; Technical Report-TR06; Computer Engineering Department, Erciyes University, Engineering Faculty: Kayseri, Turkey, 2005; pp. 1–10.

243. Yilmaz, Z.; Okşar, M.; Başçiftçi, F. Multi-objective artificial bee colony algorithm to estimate transformer equivalent circuit parameters. *Period. Eng. Nat. Sci.* **2017**, *5*, 271–277.

244. Krishnanand, K.N.; Ghose, D. Glowworm swarm optimisation: A new method for optimising multi-modal functions. *Int. J. Comput. Intell. Stud.* **2009**, *1*, 93–119.

245. Huang, X.B.; Song, T.; Wang, Y.N.; Liwen, J.Z. Transformer Fault Diagnostic Method Based on Gray Fuzzy Firefly Algorithm Optimization. Patent Application No. CN 103698627 A, 2 April 2014; pp. 2–12.

246. Huang, X.B.; Song, T.; Wang, Y.N.; Liwen, J.Z. Power transformer fault diagnosis based on IGSO optimization algorithm. *Electr. Power* **2014**, *47*, 60–65.

247. Yang, X.S.; He, X.S. Bat algorithm: Literature review and applications. *Int. J. Bioinspir. Comput.* **2013**, *5*, 141–149.

248. Gong, M.F.; Liu, Y.N.; Wang, L.H.; Song, J.; Xie, Y.X. Fault diagnosis of power transformer based on improved BP neural network optimized by bat algorithm. *J. Shandong Univ. Sci. Technol.* **2017**, *36*, 70–74.

249. Zhang, K.F.; Guo, J.; Nie, D.X.; Yuan, F.; Xiao, Z.H. Diagnosis model for transformer fault based on CRO-BP neural network and fusion DGA method. *High Volt. Eng.* **2016**, *42*, 1275–1281.

250. Zang, H.Z.; Xu, J.Z.; Yu, X.D. Power transformer fault diagnosis based on integrated artificial intelligence. *Power Syst. Technol.* **2003**, *27*, 15–17.

251. Huang, W.; Zhao, Y.B. Application of rough set and multiple population genetic algorithm in fault diagnosis of transformer. *Process Autom. Instrum.* **2016**, *37*, 27–30.

252. Yao, H. Power Transformer Fault Diagnosis Based on Improved Artificial Fish Algorithm-RBF Network. Master's Thesis, Guangxi University, Guilin, China, 2016; pp. 8–20.

253. Illias, H.A.; Zhao, W.L. Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimization. *PLoS ONE* **2018**, *13*, e0191366.

254. Wang, X.X.; Wang, T.; Wang, B.S. Hybrid PSO-BP based probabilistic neural network for power transformer fault diagnosis. In Proceedings of the IEEE International Symposium on Intelligent Information Technology Application, Shanghai, China, 20–22 December 2008; pp. 545–549.

255. Sima, L.P.; Shu, N.Q. Fault diagnosis of power transformer based on clustering binary tree SVMs. In Proceedings of the 2011 IEEE International Conference on Electric Information and Control Engineering, Wuhan, China, 15–17 April 2011; pp. 5035–5038.

256. Zheng, N.N. On challenges in artificial intelligence. *Acta Autom. Sin.* **2016**, *42*, 641–642.

257. Li, L.; Lin, Y.L.; Cao, D.P.; Zheng, N.N.; Wang, F.Y. Parallel learning—A new framework for machine learning. *Acta Autom. Sin.* **2017**, *43*, 1–8.

258. Shalev-Shwartz, S. Online learning and online convex optimization. *Found. Trends Mach. Learn.* **2011**, *4*, 107–194.

259. Settles, B. *Active Learning*; Morgan and Claypool Publishers: San Rafael, CA, USA, 2012; pp. 10–25.

260. Chen, X.G.; Yu, Y. Reinforcement learning and its application to the game of Go. *Acta Autom. Sin.* **2016**, *42*, 685–695.

261. Liu, Q.; Zhai, J.W.; Zhang, Z.Z.; Zhong, S.; Zhou, Q.; Zhang, P.; Xu, J. A survey on deep reinforcement learning. *Chin. J. Comput.* **2018**, *41*, 1–27.

262. Wang, Y.; Yuan, J.S.; Shang, H.K.; Jin, S. Optimization strategy research on combined-kernel support vector machine for partial discharge pattern recognition. *Trans. China Electrotech. Soc.* **2015**, *30*, 229–236.

263. Wu, X.H.; Liu, J.; Liang, Y.C.; Wang, X.M.; Li, Y.M. Application of support vector machine in transformer fault diagnosis. *J. Xi'an Jiaotong Univ.* **2007**, *41*, 722–726.

264. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; pp. 985–990.

265. Horata, P.; Chiewchanwattana, S.; Sunat, K. Robust extreme learning machine. *Neurocomput* **2013**, *102*, 31–44.

266. Huang, G.; Song, S.J.; Gupta, J.N.D.; Wu, C. Semi-supervised and unsupervised extreme learning machines. *IEEE Trans. Cybern.* **2014**, *44*, 2405–2417.

267. Xu, S.H.; Mu, X.D.; Chai, D.; Luo, C. Domain adaption algorithm with ELM parameter transfer. *Acta Autom. Sin.* **2018**, *44*, 311–317.

268. Liu, L.F. Research status of power transformer fault diagnosis based on machine learning. *Electr. World* **2017**, *24*, 9–10.

269. Yuan, H.M.; Wu, G.N.; Gao, B. Fault diagnosis of power transformer using particle swarm optimization and extreme learning machine based on DGA. *High Volt. Appar.* **2016**, *52*, 176–180.

270. Zou, J. Fuzzy fault diagnosis of power transformer based on new coding membership function. *Electr. Power Autom. Equip.* **2010**, *30*, 88–91.

271. Li, J.M.; Lu, D.Y.; Li, J.Q.; Mo, C.L. Application of wavelet neural networks in fault diagnosis of transformer. *J. Zhejiang Univ. Technol.* **2001**, *29*, 56–59.

272. Bai, C.F.; Gao, W.S.; Jin, L.; Yu, W.X.; Zhu, W.J. Integrated diagnosis of transformer faults based on three-layer Bayesian network. *High Volt. Eng.* **2013**, *39*, 330–335.

273. Zhu, Y.L.; Wu, L.Z. Synthesized diagnosis on transformer faults based on Bayesian classifier and rough set. *Proc. CSEE* **2005**, *25*, 161–167.

274. Wang, Y.N.; Huang, X.B.; Song, T.; Zhu, Y.C. Fault diagnosis strategy for transformer oil chromatography monitoring of fuzzy neural network based on improved PSO. *Guangdong Electr. Power* **2013**, *26*, 82–86.

275. Wang, K.F.; Gou, C.; Duan, Y.J.; Lin, Y.L.; Zheng, X.H.; Wang, F.Y. Generative adversarial networks: The state of the art and beyond. *Acta Autom. Sin.* **2017**, *43*, 321–332.

276. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489.

277. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Lucas, B.; Lai, M.; Bolton, A.; et al. Mastering the game of Go without human knowledge. *Nature* **2017**, *550*, 354–359.

278. Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Lai, M.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; et al. Mastering chess and Shogi by self-play with a general reinforcement learning algorithm. *arXiv* **2017**, arXiv:1712.01815.