An Intelligent Fault Diagnosis Approach for Power Transformers Based on Support Vector Machines

by

Hao Xu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Mechanical Engineering

University of Alberta

Abstract

Power transformers are essential for the operations of industrial systems such as metal production plants, and for the transmission and distribution of electricity to end users. Power transformer failures can cause huge loss of production, expensive downtime, significant costs for repair or replacement, and disruptions to city and community operations. Transformers are desired to operate at a high-reliability level, and they should be maintained carefully through effective condition monitoring and fault diagnosis, for evaluating transformer health conditions based on condition monitoring data and performing suitable maintenance actions.

Dissolved gas analysis (DGA) is a primary way of monitoring the health conditions of transformers by analyzing the insulation oil via periodic sampling. Different gases can be decomposed from the insulation material and the liquid oil under certain thermal, electrical, or mechanical stresses, and these gases will dissolve into the transformer oil. Existing transformer fault diagnosis methods mainly include rule-based methods documented in IEEE Standards, which are based on analyzing key gases, gas concentration ratios, or certain gas proportions. In addition, artificial intelligence (AI)-based methods were proposed, based on artificial neural network, fuzzy logic or support vector machine (SVM) tools. However, the existing rule-based and AI-based methods suffer from limited and imbalanced datasets and the

capability to deal with low concentration DGA data, and the fault diagnosis accuracy needs to be further improved.

In this thesis, a new intelligent approach based on SVM is proposed for condition monitoring and fault diagnosis of power transformers based on DGA data. The proposed method integrates a gas concentration filter and a plurality-voting SVM model. Low concentration data are typical for new transformers, but existing ratio-based methods are generally not effective in utilizing such data. A gas concentration filter is proposed to process low gas concentrations data, and it is combined with the SVM model to generate fault diagnosis results. The plurality-voting SVM model is designed with a new plurality-voting structure and integrates the synthetic minority over-sampling technique (SMOTE) to overcome the problem of imbalanced data, where the dataset sizes are significantly different for different health conditions. A parameter optimization approach based on genetic algorithm is employed. The proposed SVM-based approach is compared with existing DGA-based power transformer diagnosis methods, including rule-based methods and various AI methods. The comparative study results demonstrate the effectiveness of the proposed SVM-based power transformer fault diagnosis approach.

有志者，

事竟成。

Where there is a will, there is a way.

Acknowledgements

I would like to express my enduring gratitude to the faculty, staff, my fellow students and friends at the University of Alberta, who have inspired me to continue my work during the two-year studying for my degree.

I owe particular thanks to my supervisor, Dr. Zhigang (Will) Tian, who continually encouraged and guided me in the project, research and thesis writing. He enlarged my vision of science and provided coherent answers to my questions. I also want to acknowledge the financial support from him and my department.

During the study, I have learned more knowledge from presentations by professors, scholars and other fellow students in the bi-week group meetings organized by Dr. Ming J. Zuo, so I wish to thank Dr. Ming J. Zuo and all the members in the Reliability Research Lab as well.

Special thanks are owed to my parents and sister, who have supported me throughout my years of education, both morally and financially.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

DGA	Dissolved Gas Analysis
AI	Artificial Intelligence
ANN	Artificial Neural Networks
SVM	Support Vector Machine
PSO	Particle Swarm Optimization
GA	Genetic Algorithm
T1	Low-temperature Thermal Fault
T2	High-temperature Thermal Fault
PD	Partial Discharge
D1	Discharge of Low Energy
D2	Discharge of High Energy
NF	No Fault Condition
ND	Fault Not Detected

Chapter 1: Introduction

1.1 Background

It is hard for modern people to live without electricity. To use electricity, power transformers are essential to convert the voltage of electricity to satisfy the demand of users. Power transformer can increase voltage to reduce the energy loss in electricity delivery process, and decrease voltage to meet the demand of daily use of electricity. Therefore, we can find power transformers near power plants as well as the places where people live.

Figure 1.1 An illustration of a general power transformer (ENGie, 2016)

A general power transformer is shown in Figure 1.1 (ENGie, 2016), and it usually consists of iron cores, windings, cooling systems, insulation components, and bushings. The iron cores and windings are the key components to convert voltage. The cooling systems and insulation components are vital to guarantee basic operation environment. Brushings are used to connect input and output wires.

Operating a power transformer usually requires less care than most other power and mechanical equipment. However, since it links power plants to customers and the delivery system, transformer failures can cause huge losses in production, and repair or replacement can also lead to significant costs. Transformers accidents are not rare. According to FM Global, a commercial property insurance company, transformer failures cost its clients (energy-related companies) a combined US$339 million in lost revenue within a five-year period (2008–2013), which ranked third among the top five types of losses (Gulla, 2014). Most recently, the Brazi power plant in Southern Romania, operated by OMV Petrom, southeastern Europe’s largest integrated oil and gas group, suffered a power outage because of a power transformer failure on April 28, 2017 (Wallingford, 2017). Although the failure is currently under investigation, OMV Petrom estimated that the plant would more likely be out of operation for the next three months, which would adversely affect the local industry and people’s daily lives.

Such accidents happen frequently and lead to huge consequence because the origin of a failure is hard to detect and the development of an accident is so fast that local fire services cannot easily stop the immediate damage. The problem always begins with an internal short circuit and an electric arc inside insulation components or windings of transformers, which can lead to the increase of the temperature inside transformers. Besides, the failure in cooling systems or jam of the insulation oil can heat the temperature up as well. The high heating temperature can vaporize and decompose the insulating liquid and greatly increase the internal pressure, which results in a huge explosion of the transformer's outer shell. During the explosion, the insulating liquid oil can be ejected and form fireballs that burn down other combustibles. The liquid oil leaking from the transformer's rupture point can lead to a blazing fire that may spread to adjacent equipment. Figure 1.2 (Henderson, 2016) illustrates this terrible process. Although initially smoke can be seen, a muffled explosion can be heard, and local fire services can be requested immediately, it is hard to prevent substantial damage.

Figure 1.2 Explosion of a power transformer (Henderson, 2016)

The enormous potential damage and severe consequences of major transformer faults require actions to prevent them from occurring. Measures to do so include monitoring transformers, detecting faults, and scheduling preventive and predictive maintenance activities (Li et al., 2013; Muthanna et al., 2006).

Methods for monitoring transformers have drawn much research attention (Dong et al., 2008), especially methods based on dissolved gas analysis (DGA), which have gained worldwide acceptance in recent decades (Duraisamy et al., 2007a). DGA methods detect faults in the transformer by monitoring the insulation oil, which includes liquid transformer oil and solid impregnated cellulose. Such faults are typically derived from deteriorated insulation and aging (Sun et al., 2012). Different gases, such as hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), carbon monoxide (CO), and carbon dioxide (CO_2), will decompose from the insulation material and liquid oil under certain thermal, electrical, or mechanical stresses in the transformer and will then dissolve into the transformer oil. When a transformer is undergoing an abnormal situation, more gases will be released than the case under normal conditions, and their concentrations in the transformer oil will increase. Thus, in other words, DGA is used to analyze the degree of the decomposition process by detecting the gas concentrations in the transformer oil.

Many methods based on DGA data have been developed by experts to detect transformer faults, which have become the dominant methods in electric industries worldwide. These methods use different measures and principles, such as gas concentrations, key gases, key gas ratios, and graphical representations (Sun et al., 2012). Gas concentrations involve directly use data obtained from DGA. Key gases are analyzed to find the dominant gas. Key gas ratios are used to find the relationships between certain gases. And graphical representations are used to plot data into a defined graph under specific rules. The traditional rule-based methods are listed as follows:

- Key gas method,
- Doernenburg ratio method,
- Rogers ratio method,
- IEC ratio method, and
- Duval triangle method.

Apart from these empirical methods, many new approaches and techniques have been proposed in recent decades. They are commonly developed with the support of artificial intelligence (AI). The most commonly used AI methods for fault diagnosis of power transformers include the expert system approach (Styvaktakis et al., 2002; Saha and Purkait, 2004a; Németh et al., 2010), fuzzy logic method (Muhamad et al., 2007; Saha and Purkait, 2004b; Su et al., 2000b), artificial neural networks (ANN) approach (Sarma and Kalyani, 2004a; Seifeddine et al., 2012; Wang et al., 2000a; Zakaria et al., 2012), and support vector machine (SVM) method (Bacha et al., 2012a; Fei and Zhang, 2009a).

1.2 Research Motivation

The conventional methods can be performed easily without computers. However, a key disadvantage is that the accuracy of the diagnosis results from these approaches is quite low and cannot satisfy reliability and safety requirements. Intelligent AI-based methods have shown their effectiveness for fault diagnosis of power transformers. However, many disadvantages still exist. For example, large amounts of data should be known in advance when developing such methods, yet real data are usually limited and the amount of data for each fault type is extremely imbalanced, which can adversely affect the performance of AI-based methods. Besides, existing AI-based methods that directly use ratios as input features ignore to consider low-concentration data so that these methods are hard to give correct diagnosis results for this kind of data. Therefore, the traditional methods and AI-based methods deserve extensive study and further exploration to find the best measure for fault diagnosis of power transformers.

1.3 Objective and Research Contributions

This thesis focus on condition monitoring and fault diagnosis methods of power transformers to support predictive maintenance actions. The existing traditional rule-based methods and selected AI methods are reviewed and discussed. They suffer from limited and imbalanced datasets and the capability to deal with low concentration DGA data, and the fault diagnosis accuracy needs to be further improved. To overcome these problems, a new method should be proposed.

In this thesis, a new intelligent approach that integrates a gas concentration filter and a plurality-voting SVM model is proposed for condition monitoring and fault diagnosis of power transformers based on DGA data, and then this method is compared with existing methods to validate its effectiveness. The main contributions of this thesis are summarized as follows:

- This thesis contains a review of many research works related to fault-diagnosis techniques and methods for power transformers. The strengths and weaknesses of these methods are summarized and commented upon, which can provide meaningful information for further research. Some common problems and challenges in the power transformer industry are also summarized in this thesis.

- The SVM model in the proposed method is designed with a new plurality-voting structure rather than the existing multi-layer structure. The newly designed model does not highly rely on any single binary sub-SVM, and every sub-SVM is equally important.

- The synthetic minority over-sampling technique (SMOTE) is first used to generate DGA data and balance training datasets for SVM modeling since it can overcome the problem of imbalanced data, where the dataset sizes are significantly different for different health conditions. The fault diagnosis performance is improved by using this technique according to the results in this thesis.

- A gas concentration filter is proposed to process low-concentrations data, and it is combined with plurality-voting SVM model to generate fault diagnosis results. This is the first time that an AI-based method is combined with a gas-concentration judgment procedure.

- Comparisons among the results from different methods are presented to show the advantages of the proposed method over the traditional rule-based methods, ANN method, and usual SVM approach. The comparison involves not only the overall diagnosis accuracies from these methods but also the potential costs of misdiagnosis, which is also an important factor that should be considered to validate the effectiveness of different methods.

1.4 Thesis Organization

- **Chapter 1** – Introduction

  This chapter introduces the background, research motivation, objective and the contributions of the thesis. The thesis organization is structured in this chapter here.

- **Chapter 2** – Literature review

  This chapter consists of the literature review on the existing methods for fault diagnosis of power transformers. The traditional methods and intelligent methods are described and discussed, and the observations from the literature are concluded.

- **Chapter 3** – Fundamental knowledge

  In this chapter, the fundamental knowledge of SVM, imbalanced dataset problem, and existing fault diagnosis methods for power transformers are introduced. SVM and SMOTE algorithm are the key components in the proposed method.

- **Chapter 4** – The Proposed SVM-based Approach for Fault Diagnosis of Power Transformers

  In this chapter, the proposed method for the fault diagnosis of power transformers is presented. The details of the proposed approach are discussed too in this chapter.

- **Chapter 5** – Method validation and comparison

  In this chapter, existing methods and the proposed method are compared using the data from the literature. This study gives comparisons of the overall accuracies, specific results, and potential costs of misdiagnosis.

- **Chapter 6** – Conclusion and future work

  Based on the comparison results from Chapter 5, conclusions and suggested future work are presented.

Chapter 2: Literature Review

In this chapter, literature related to condition-monitoring and fault-diagnosis methods for power transformers is presented and discussed. DGA, one of the basic sources of diagnostic methods for power transformers, is reviewed in Section 2.1. The traditional methods that include the key gas, Doernenburg ratio, Rogers ratio, IEC ratio, and Duval triangle method are discussed in Section 2.2. In Section 2.3, the non-traditional methods—in other words, the AI methods—are discussed. Finally, the literature review is summarized in Section 2.4.

2.1 Dissolved Gas Analysis (DGA)

Transformer equipment is so expensive that it should be monitored carefully during their operation. The cost of a 765 KV transformer failure is over $2 million, and this price is only for the equipment itself, without the calculated loss of production (Duval, 1989).

DGA was introduced to monitor the conditions of a specific transformer and gradually gained acceptance among professional experts. Like doctors checking a human body with a stethoscope, DGA can be used to determine the most possible situation inside transformers, give early warnings and diagnoses, and increase the opportunity to act correctly.

It is not difficult to understand why DGA provides useful information for condition monitoring. Under normal circumstances, the insulation oil and cellulose molecules constituting the dielectric insulation do not decompose at a rapid rate. However, if high thermal and/or electrical stresses exist in the transformer, these conditions will increase the chemical breakdown of the insulation oil and solid insulation. These breakdowns generate gases that partially or entirely dissolved in the oil. The dissolved gases can be simply detected at the ppm unit level and can be divided into combustible and noncombustible gases, as listed in Table 2.1.

Table 2.1 Dissolved gases in the insulation oil

Combustible	Noncombustible
Carbon monoxide (CO)	Oxygen (O_2)
Hydrogen (H_2)	Nitrogen (N_2)
Methane (CH_4)	Carbon dioxide (CO_2)
Ethane (C_2H_6)	Vapor (H_2O)
Ethylene (C_2H_4)	
Acetylene (C_2H_2)	

Therefore, transformers should regularly be monitored by periodically sampling the oil in the transformers to collect the gas concentrations as DGA data. With the development of sensors, the gases dissolved in transformer oil can be continuously monitored using a gas chromatography system (De Faria et al., 2015). After collecting the DGA data, the faults can then be diagnosed. These faults roughly include thermal decomposition (overheating), corona (partial discharge), and electric arcing.

2.2 Traditional Methods for Fault Diagnosis of Power Transformers

2.2.1 Key Gas Method

An appearance of a fault can increase the temperature inside the transformer and result in decomposition of the insulation oil. The key gas method (IEEE, 1992) is used to directly measure the concentrations of the gases that are decomposed and dissolved in the insulation oil.

This method identifies faults according to the presence and percentage of gases (Sun et al., 2012). After calculating the percentage of each gas, the most dominant ones can be defined as the "key gases." Based on industry experience, the key gases are used to interpret the DGA data according to a simple set of facts. For example, under low-intensity partial discharge or corona fault, insulation oil mainly produces more H_2, so the key gas for low-intensity partial discharge or corona cases is H_2. In other words, if an oil sample contains a high percentage of H_2, using the key gas method, one can conclude that the potential fault is a low-intensity partial discharge or corona. With this method, only one or two key gases can finish the diagnosis work to determine the final condition results, which is unreliable.

2.2.2 Dornenburg Ratio Method

Unlike the key gas method, with which direct gas concentrations are used, the Doernenburg ratio method (IEEE, 1992) uses ratios. Three types of gas ratios, CH_4/H_2, C_2H_2/C_2H_4, C_2H_2/CH_4 and C_2H_6/C_2H_2, can be used to diagnose thermal faults, corona discharge, and arc. It should be noted that one can not classify the thermal fault into different levels when using this method, but Roger ratio method and IEC ratio method can divide the thermal fault into low-temperature thermal fault and high-temperature thermal fault.

When using the Doernenburg ratio method, one first checks the concentration of each gas. The gas concentrations must exceed the pre-made limits, and then the ratio method can be performed by following some specific rules. To finally get the diagnosis result, each ratio should fall into predetermined ratio ranges to satisfy the requirements for each fault type. However, the Doernenburg ratio method cannot diagnose some conditions, and we will get a result of “fault not identifiable: resample.” Therefore, the Doernenburg ratio method is not applicable for all conditions.

2.2.3 Rogers Ratio Method

The Roger ratio method (IEEE, 1992) is widely used because it can classify more types of thermal faults than the Doernenburg ratio method. This method uses three ratios:

CH_4/H_2, C_2H_2/C_2H_4, and C_2H_4/C_2H_6. The first two ratios are the same as the first two ratios used in the Doernenburg ratio method. However, the Roger ratio method diagnoses the faults using direct ratios without checking the gas concentrations, which is a necessary step in the Doernenburg ratio method. In the Roger ratio method, transformer conditions are classified as normal, low-temperature thermal fault (below 300 °C), medium temperature thermal fault (300 °C to 700°C), high-temperature thermal fault (over 700°C), partial discharges, and high energy arcing. The classes are more detailed and precise than the Doernenburg ratio method.

It should be noted that the Roger ratio method, which only uses ratios, may highly misdiagnose the condition when gas concentrations are low. That is, this method may not perform well on new transformers, which often have low gas concentrations in their insulation oil.

2.2.4 IEC Ratio Method

The IEC method uses the same three gas concentration ratios as the Rogers ratio method. The first version of the IEC ratio method was published in 1978 using a simple coding scheme. The corresponding code for each ratio range can be found in the IEC document. The diagnosis result is determined by a coding scheme. However, the 1999 version of the IEC ratio method uses the ratio ranges directly (IEC, 1999). The next version of the document of IEC added a 3D graphical representation of the ratio ranges. Data can be plotted on a graph if the faults cannot be diagnosed by the ranges alone. The final determination of the fault type is to find which region of fault in the graph is closest to the original ratio's data point. Similar to other methods, faults are typically classified as partial discharges, low or high energy discharges, and thermal faults, whose severity depends on the fault temperature.

2.2.5 Duval Triangle Method

The Duval triangle method (Duval, 2002; Duval and Dukarm, 2005) only uses the percentage values of three gases—CH_4, C_2H_4, and C_2H_2—and their plotted locations on a triangular map (see Figure 2.1) can help to determine the fault type.

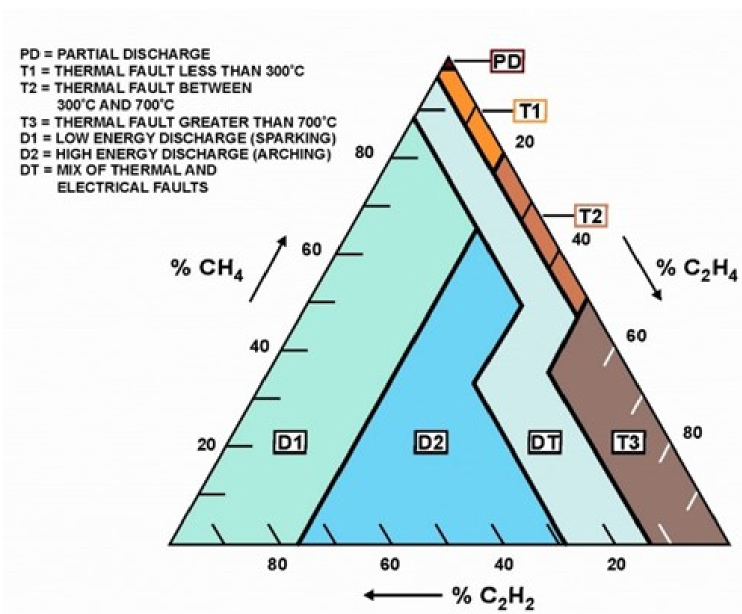

Figure 2.1 The map of the Duval triangle method (Duval, 2002)

In this method, the faults are classified as partial discharges, high and low energy arcing, and thermal faults in three different temperature ranges. Although this approach is easy to perform, this approach can also misjudge the condition in new transformers with low amounts of gases dissolved in the insulation oil.

2.2.6 Summary of Traditional Methods

The conventional methods are easy to implement when following the corresponding rules of each method. Even without using a computer, one of the most important tools in the current century, a person can perfectly finish a diagnosis using DGA data. These methods are rule-based and do not require historical data, which is advantageous. These methods were based significantly on the wealth of experience from monitoring transformers in the industry and have been widely recognized.

The primary goal of a transformer-diagnosis method is to obtain accurate diagnosis results and detect all of the conditions in the transformers. Thus, many problems exist in these methods:

- Except for the Doernenburg ratio method, the other four methods are not strongly applicable to new transformers, which do not have large amounts of gases dissolved in the insulation oil. Ratios obtained from low-concentration data span a wide range, and using ratios alone may not explain the conditions well.

- For the ratio methods and Duval triangle method, if a ratio is just equal to a range's boundary, the diagnosis becomes difficult to decide. If the ratios are close to the boundaries, a small difference in the data can lead to an entirely different result, which is not true in reality.

- The Doernenburg ratio method cannot detect some conditions, and the only solution in this case is to resample the oil (IEEE, 2008). Thus, blind spots exist in the Doernenburg ratio method and can cause more money and time for resampling procedures.

- For one set of data, these five methods can provide different diagnosis result (Mehta et al., 2013), and the accuracy of these five methods is not high (Sarma and Kalyani, 2004b; Su et al., 2000b). Therefore, it is desired to introduce a more reliable method with high fault diagnosis accuracy.

2.3 Non-traditional Methods for Fault Diagnosis of Power Transformers

With the development of AI techniques, researchers have established new approaches for fault diagnosis. These methods rely on the use of computers and are more or less based the experience with the traditional methods.

2.3.1 Fuzzy Logic Method

Fuzzy logic is a way of mapping input to the target output using linguistic rules formed from human understanding, rather than from stringent mathematical calculations. The fuzzy logic method includes three steps: fuzzification, fuzzy inference, and defuzzification (Singh and Joshi, 2015). The first step is to transform the input data into membership grades for linguistic terms of fuzzy sets, during which the membership function is used to associate a grade with each linguistic term. The second step is to find the output results from the knowledge-based rules in the form of the linguistic interpretation. De-fuzzification involves reconverting the fuzzy output back into an output that humans can understand.

A transformer fault diagnosis system was developed that employed a fuzzy logic approach and showed better performance than the traditional ratio methods (Huang et al., 1997). The defined inputs in their study were the three ratios in the IEC method, and the accuracy of the fault diagnosis was between 70% to 80% based on over 700 datasets from Taiwan Power Company. Instead of the trapezoid membership function used by Huang et al. (1997), Su et al. (2000a) employed a demi-Cauchy distribution function to improve the diagnosis performance. Su (2016), using the same membership function as Su et al. (2000a), extended the IEC's three input ratios into four ratios (Su, 2016). Dhote and Helonde (2014) defined a new fuzzy inference system using a combination of three membership functions for their new fuzzy logic model and obtained higher accuracy than five other methods mentioned in their work.

Although the fuzzy logic method shows advantages over the traditional methods, it still has some drawbacks. Only a fixed mathematical membership function can be used in the modeling when developing a fuzzy logic method, but the function must be dynamic and changeable because it is hard to describe all cases using a single membership function.

2.3.2 Expert System Method

The expert system, emulating the decision-making ability of a human expert, is another branch of AI that has been widely used in many industrial and commercial applications. It can act as an expert and use specific knowledge to deal with real-world problems. Lin et al. made an expert system with rule-based knowledge representation that used a knowledge engineering system integrating the Roger ratio and Doernenburg ratio methods (Lin et al., 1993). The designed expert system has been tested using records from Taiwan Power Company to show its effectiveness in diagnosing transformer faults. Beyond the fuzzy logic method, Wang et al. (2000b) also developed an expert system consisting of an ANN-based normal/abnormal classifier, a knowledge-based normal/abnormal classifier, an ANN-based individual fault detector, and a knowledge-based individual fault detector to recommend maintenance actions (Wang et al., 2000b). The diagnosis accuracy of this expert system is higher than that of the Rogers ratio method. Liao et al. (2001) introduced an expert system comprising many modularization components, including the ANN approach, the fuzzy logic method, the IEC standards, and some expert experiences (Liao et al., 2001), and two case studies showed the effectiveness of their work.

Recently, the number of published studies on the expert system method for diagnosing faults in power transformers has not been growing rapidly, compared with the number of studies on the fuzzy logic, ANN, and SVM methods. The reason is that expert systems depend heavily on known knowledge, which is sometimes complicated and incorrect. Poor knowledge can lead to a bad expert system. In addition, expert systems can neither gain knowledge through self-learning processes with new data nor fit its diagnostic regulations automatically.

2.3.3 Artificial Neural Network (ANN) method

ANNs are powerful tools that can process nonlinear data and has been employed for equipment fault diagnosis and prediction issues (Tian and Zuo, 2010). Many methods based on ANN and DGA were developed to identify transformer faults. To build a good ANN model, researchers should first determine what kind of neural networks to use and then select the proper input features, define the number of layers, and use suitable parameters to develop the model.

Sun et al. (2007) introduced a back-propagation neural network model, in which each weight of neural has an independent learning rate and a momentum coefficient that is adapted through iterations. This approach significantly accelerated learning performance and performed better than the conventional back-propagation algorithm, both with a constant momentum and without momentum, in fault diagnosis for power transformers. Cao et al. (2006) made a probabilistic neural network (PNN) model, in which the parameters of the PNN are determined by genetic algorithms to increase the diagnostic accuracy.

To find the most proper parameters, Illias et al. (2015) combined the ANN and various particle swarm optimization (PSO) techniques to predict transformer faults, which has very reliable diagnosis accuracy. To further evolve their model, they developed a modified model named the particle swarm optimization-time varying acceleration coefficient-artificial neural network (MEPSO-TVAC-ANN) model (Illias et al., 2016). Beykverdi et al. (2016) simulated a transformer fault diagnostic model based on a hybrid approach using the ANN and the neural-imperialistic competitive algorithm (Nero-ICA). Its simulation results validated the Nero-ICA model as being more accurate and efficient than the simple structured ANN model when the number of training datasets becomes larger. Souahlia et al. (2012) developed a multilayer perceptron neural network model that uses a combination of the ratios in the Rogers and Doernenburg ratio methods as inputs. The classification accuracy of the classifier is the highest, compared to the fuzzy logic, radial basis function, K-nearest neighbor, and probabilistic neural network approaches.

It seems that optimized ANN methods can successfully achieve proper diagnosis accuracy, yet difficulty exists in determining the network's structure and the number of nodes in its layers. Also, it is very time-consuming to train the ANN models, compared to the time used to develop other types of models.

2.3.4 Support Vector Machine (SVM) Method

SVM, developed by Vapnik in 1995, is a computational learning method based on statistical learning theory (Vapnik, 2013a), which can develop effective models for classification and reduce the over-fitting problems that occur in ANN methods (Heisele et al., 2003). Based on the procedure of preparing an SVM model, four factors can affect an SVM model's performance, which are the overall SVM model structure, proper parameters, suitable kernel functions, and selection of proper inputs.

The first SVM-based method for fault diagnosis of power transformers was developed through a multilayer approach (Ganyun et al., 2005). A three-layer SVM classifier developed is shown in Figure 2.2 (Ganyun et al., 2005), and the advantages of their three-layer SVM classifier over the back-propagation ANN method were summarized, such as the low requirements of training data and less training time to develop their SVM model. However, Ganyun et al. did not conduct optimization of parameters.

Bacha et al., (2012b) also investigated a multilayer SVM classifier with six layers that elaborates an input vector established by the combination of ratios, and they showed that an SVM with the Gaussian function performed better than an SVM with other kernel functions on diagnostic accuracy. A different multilayer SVM model was established in which the genetic algorithm (GA) was applied to optimize the SVM parameters to prevent over-fitting or under-fitting of the SVM model (Fei and Zhang, 2009b), and this method was proved to perform better than the IEC ratio method, back-propagation ANN, and normal SVM method without using genetic algorithm.

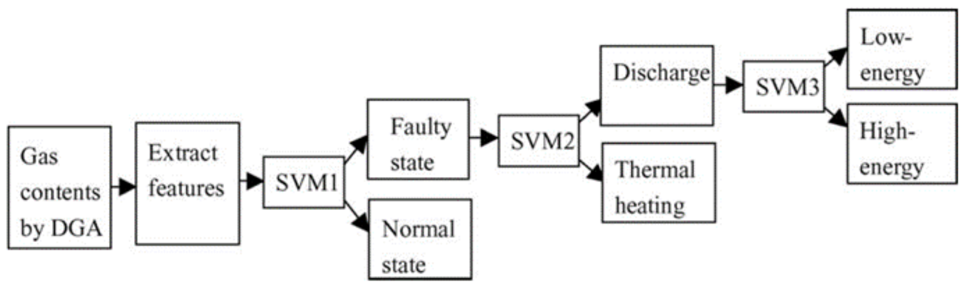

Figure 2.2 The structure of Ganyun's three-layer SVM classifier

In the multilayer SVM fault diagnosis tool proposed by Li et al. (2016), grid search, GA, and PSO were used to find the best parameters and a comparison among these three parameter optimization methods were conducted and they conclude GA could help find the best parameters and can achieve the highest accuracy in fault diagnosis. Liao et al. (2013) developed a one-against-one multiclass SVM classifier based on PSO with time-varying acceleration coefficients for transformer fault diagnosis. Using PSO, the classifier with optimized parameters can achieve the best classification accuracy and generalization performance among other methods. Zheng et al. (2011) presented a multiclass least square support vector machine (LS-SVM)-based classifier for transformer fault diagnosis, and the algorithm of PSO was implemented to select the optimal input features. Yin et al. (2011) developed a multi-kernel support vector classifier that can learn from training samples using the kernel function obtained from a linear combination of several basic kernels. A comparison showed that as the search space of the optimal kernel broadens, the robustness of the classifier is enhanced and its accuracy improves.

The review above can be summarized as follows:

- The multilayer SVM model has been popular among researchers (Bacha et al., 2012b; Fei and Zhang, 2009b; Ganyun et al., 2005). However, if the previous SVM layer classifies a set of data incorrectly, the final result will be wrong.
- In the procedure for optimizing the model, PSO, GA, and other algorithms are widely utilized to find the best parameters for fault diagnosis models (Fei and Zhang, 2009b; Liao et al., 2013; Yin et al., 2011; Zheng et al., 2011), which can significantly improve the performance of the models.
- The choice of kernel functions also plays an important role in perfecting the model. Popular kernel functions, such as the RBF, linear, and Gaussian kernel functions, can be used either independently or jointly (Yin et al., 2011).
- Regarding the procedure of choosing the features as inputs of a model, using different ratios of gas concentrations as the input is the most common procedure. It is feasible to use the GA to find the most related ratios that can be employed as the input (Li et al., 2016).

2.4 Discussion and Summary

In the 1980s and 1990s, the traditional methods for fault diagnosis of transformers, based on historical data and industry experience, prevailed because of their convenience and effectiveness. However, because many accidents still happened after these methods were used for condition monitoring, the requirement of highly reliable diagnosis was adopted. Traditional methods provide more information for the development of AI methods. Most AI methods use ratios derived from traditional ratio methods as input for their models. That is, the traditional approaches have concluded the relevant input features for AI-based methods to determine transformer faults.

Although conventional methods cannot give highly accurate fault diagnosis, some parts of the individual methods are reliable and efficient. For example, the first step of the Doernenburg ratio method, the concentration judgment, matters to deal with the low concentration data in new transformers. Ratios make up almost all the possible inputs in AI methods, but they are not credible in the case of new transformers. If we only use the ratios for a set of low-concentration data from a new transformer, named data A, and another set of data that is precisely ten or more folds of the values in data A, the diagnosis results will be the same. In other words, these two cases are identical for the methods that only use ratios as input without paying attention to the gas concentrations. However, data A may not cause a fault to occur, but another set of data is more likely to give rise to a failure. A new transformer without much gas released and dissolved in the insulation oil should have a normal status, which is true based on real practical cases. This is also true using the Dornenburg ratio method because its first step is to judge the concentration of each gas. If no gas concentrations are over the limit values, the transformer does not have a fault. Thus, the procedure of the concentration judgment in the Doernenburg ratio method is useful and should be saved in developing other new approaches.

As to the AI methods, they are not perfect either. Indeed, they can successfully avoid the boundary problem that a small difference in the data can lead to an entirely different result when the ratios are closer to the condition boundaries set by each traditional method. However, AI methods have shortcomings as well. Developing AI models requires plenty of historical data, but the traditional methods do not. There are few databases available for researchers to use. In the daily transformer management in a company, the frequency of data sampling is from every two to six months, depending on the age of the transformer. This frequency is not high, and the company may not get much data, even over the course of several years.

Meanwhile, since there are many types of faults in the transformers, it is hard to get many data for each of the faults. There are many records for the normal condition because most conditions are normal because of the normal operation of the transformer. Much data also exist on high-energy-discharge faults, because this severe fault occurs at the end under the undetected unhealthy conditions inside the transformers. However, for less serious faults, such as partial discharge, the records are rare and insufficient.

Therefore, usually we can only get an imbalanced dataset with a very significant difference among the total numbers of each type of fault. However, feeding the imbalanced datasets to the SVM model will lead to an imbalanced model with significantly decreased performance (Wu and Chang, 2003). Classifiers, including SVMs, cannot be very effective when based on an imbalanced database. This is because they are designed to generalize from sample data and output the simplest hypothesis that best fits the data based on the principle of Occam's razor, which is embedded in the inductive bias of many machine learning methods (Akbani et al., 2004). In other words, when there is a data imbalance, the classification result is often biased to the majority class. Therefore, to balance the datasets, proper data sampling should be conducted.

From a review of nontraditional methods, we can conclude that AI methods are popular among researchers. The expert system and fuzzy logic models can take DGA standards and other human expertise to form a decision-making system, which can also utilize the influence of objective factors, such as transformer size, manufacturer, volume of oil, and history of diagnosis results. However, both methods require an extensive knowledge base that must be manually constructed. Therefore, they cannot adjust their diagnostic rules automatically and gain knowledge from new data samples through a self-learning process. ANN methods can directly acquire experience from training data, which overcomes the shortcomings of the expert system. However, it still has certain disadvantages in applications, such as local optimization, over-fitting, and difficulties in convergence. Besides, training an ANN model takes longer than making an SVM model (Ganyun et al., 2005). SVM is powerful in tackling the over-fitting problem. It is effective for dealing with small sampling, nonlinear data, and high dimensional input problems. These problems only exist in the problem of transformer fault diagnosis.

Therefore, it is important to propose a new fault diagnosis approach for transformers, which could be integrated with concentration judgment procedures, proper data sampling strategies, and an optimized SVM approach.

Chapter 3: Fundamental Knowledge

This chapter presents the basic knowledge of the mathematical modeling of support vector machines, the imbalanced dataset issue, and some existing fault diagnosis methods mentioned in Chapter 2 and used for comparison purposes in Chapter 5.

3.1 The Basics of Support Vector Machines (SVM)

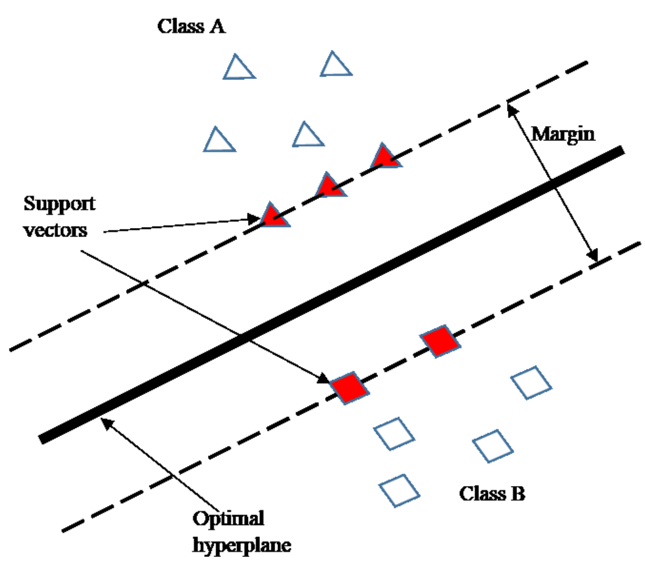

Figure 3.1 Illustration of a binary classification by SVM

An SVM approach tries to find an optimal hyperplane to separate different types of data by obtaining the maximum margin between this hyperplane and the data (Vapnik, 2013b). To make the SVM theorem visually easy to understand, Figure 3.1 shows an illustration of a simple binary classification problem, where the red-filled shapes represent the support vectors and the unfilled and filled shapes represent the training data. The hyperplane can be drawn after obtaining the support vectors.

The mathematical story behind Figure 3.1 can be interpreted as follows. Given a set of data $\mathrm{T} = \{x_k, y_k\}_k^m$, where x_k denotes the input vector, $y_k \in \{-1,1\}$ denotes the output, and m denotes the total sample number, and then $\exists \mathrm{f}(x) = 0$ divides the given data when the two classes are linearly separable.

$$\mathrm{f}(x) = w \cdot k + b = \sum_{k=1}^{m} w_k \cdot x_k + b = 0 \qquad (3.1)$$

where w denotes the weight vector and b denotes the bias term. w and b are used to define the position of the hyperplane, which should satisfy the constraints:

$$y_k f(x_k) = y_k(w \cdot x_k + b) \geq 1, k = 1, 2, \dots, m \qquad (3.2)$$

The positive slack variable ζ_i is the distance between the margin and the vectors x_k that lie on the wrong side of the margin. Therefore, the optimization problem becomes:

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + c\sum_{k=1}^{m} \zeta_i \ , k = 1, 2, \dots, m \qquad (3.3)$$

$$\text{Subject to} \quad \begin{cases} y_k(w \cdot x_k + b) \geq 1 - \zeta_i \\ \zeta_i \ \geq 0 \end{cases} \qquad (3.4)$$

where c is the penalty factor.

According to the Lagrangian principle, the problem transfers to:

$$\text{Maximize} \quad \mathrm{L}(\alpha) = \sum_{k=1}^{m} \alpha_k - \frac{1}{2} \sum_{k,i=1}^{m} \alpha_k \alpha_i y_k\, y_i (x_k \cdot x_i) \qquad (3.5)$$

$$\text{Subject to} \ \sum_{k=1}^{m} \alpha_k y_k = 0, \alpha_k \geq 0, \mathrm{k} = 1, 2, \dots, \mathrm{m} \qquad (3.6)$$

Then, the problem changes to solve the dual optimization problem for linear classification:

$$\mathrm{f}(x) = sign\left(\sum_{k,i=1}^{m} \alpha_k y_k (x_k, x_i) + b\right) \qquad (3.7)$$

An SVM can solve the nonlinear problem as well, using kernel functions to map the original data into a high-dimensional space where the linear separation becomes possible. Eq. 3.7 changes to Eq. 3.8.

$$\mathrm{f}(x) = sign\left(\sum_{k,i=1}^{m} \alpha_k y_k \psi(x_k, x_i) + b\right) \qquad (3.8)$$

where $\psi(x_k, x_i)$ is called the kernel function, $\psi(x_k, x_i) = \phi(x_k)\phi(x_i)$. Figure 3.2 shows an example of mapping two-dimensional data into a three-dimensional space, where the data can be separate linearly. The commonly used kernel functions are shown as follows (Scholkopf and Smola, 2001):

- linear kernel function: $\psi(x_k, x_i) = x_k \cdot x_i$
- polynomial kernel function: $\psi(x_k, x_i) = (x_k \cdot x_i + 1)^d$
- Gaussian radial basis kernel function: $\psi(x_k, x_i) = exp(-\|x_k - x_i\|/2\sigma^2)$
- sigmoid kernel function: $\psi(x_k, x_i) = \tanh(\alpha(x_k, x_i) + \beta)$

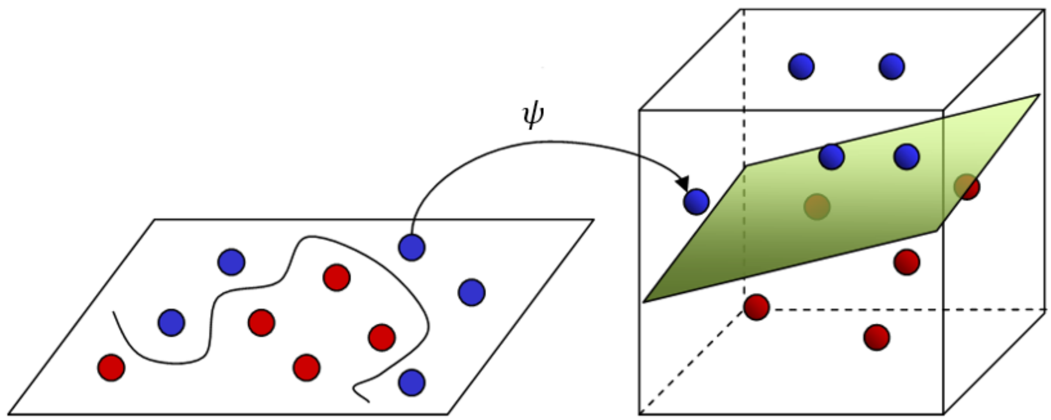

Figure 3.2 An illustration of mapping two-dimensional data into a three-dimensional space

3.2 Imbalanced Dataset Problem and Its General Solution

The problem of the imbalanced dataset in machine learning is a situation where the total size of a class of data (positive) is far larger than the total number of another class of data (negative). This case is ubiquitous in the real world, including in cases of medical diagnosis, optical character recognition, fraud detection, *etc.* For example, if we randomly collected body temperature data from all the children in a primary school, most of the data would be under 37 °C because most of the children would not be suffering a fever, and a dataset containing fewer feverish children and more healthy children would be considered an imbalanced database. That is, the case is considered an imbalanced dataset when the ratio between a class and another class is much higher than one.

Most machine learning algorithms and approaches can work well when the number of instances of each class is roughly equal. However, if the number of cases of one class far exceeds the number in the other, it can give rise to incorrect classification problems. This issue is interpreted in Figure 3.3. In Figure 3.3 (a), we have a hyperplane that separates the two classes of training data (filled in black) with the class labels of A and B. However, here there can be a situation in which a set of testing data (filled in red in Figure 3.3 (b)) labeled in class B is misclassified as class A by the hyperplane. If we have more data (filled in green and black) to balance the data set, the hyperplane can be more exact and make the classification more reliable, as shown in Figure 3.3 (c).

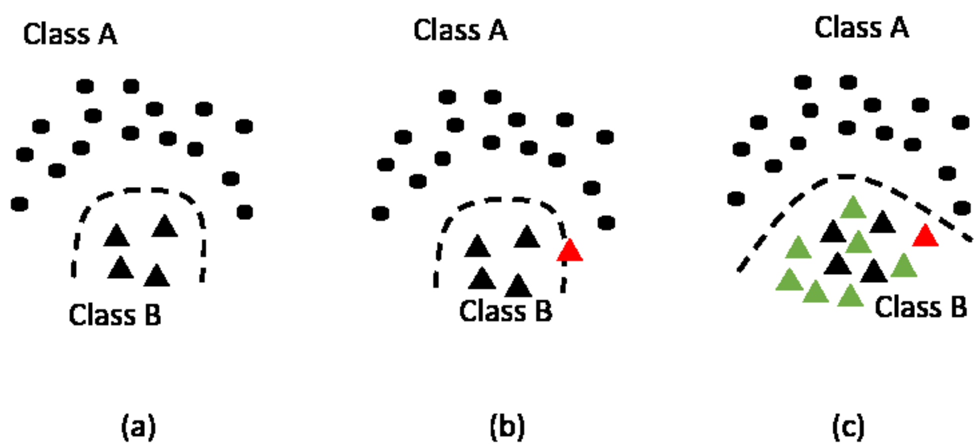

Figure 3.3 An illustration of the imbalanced dataset problem solved by oversampling

It should be noticed that the problem can also be solved if we remove some of the data from class A, shown in figure 3.4. However, this approach will eliminate much real information, so it works well only when we have enough data in the minority class. Otherwise, the removal of information could give rise to a worse classification result.

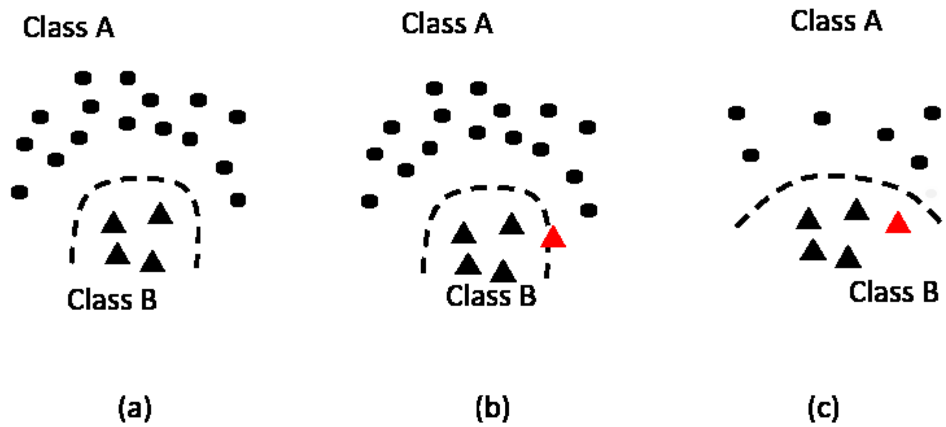

Figure 3.4 An illustration of the imbalanced dataset problem solved by under-sampling

3.3 Existing Fault Diagnosis Methods Based on DGA

This section describes the commonly used DGA approaches for fault diagnosis of power transformers. The key gas method, shown in Section 3.3.1, is related the feature extraction part of the methods proposed in Chapter 4. The Roger ratio method, Doernenburg ratio method, and IEC ratio method are used in the result comparison part in Chapter 5, are presented in Section 3.3.2.

3.3.1 Key Gas Method

The key gas method directly measures the DGA data after evidence of a fault. When the percentage of each gas is calculated, the most dominant gas can be defined as the "key gas." When the key gases can be determined, the corresponding fault type is determined from experience as shown in Table 3.1. This approach is easy to conduct without many calculations, so it was applied the most frequently in industries several decades ago.

Table 3.1 The interpretation of the key gas method

Key gases	Suggested fault types
O_2 and N_2	Non-fault condition
C_2H_6 and C_2H_4	Low temperature overheating
C_2H_4	High temperature overheating
CO and CO_2	Overheating of cellulose insulation
H_2	Corona
C_2H_2	Arcing

3.3.2 Ratio Methods

Unlike the key gas method using the percentage of gas concentration, the ratio method is employed using the ratio values between certain gas concentrations. The commonly used ratio methods are the Doernenburg (Doernenburg and Strittmatter, 1974), Roger (Rogers, 1978), and IEC 60599 ratio methods (IEC, 1999). Ratio methods are also easy to implement by finding the corresponding ratio ranges. The ratios used in these methods are listed below:

- R1: CH_4/H_2
- R2: C_2H_2/C_2H_4
- R3: C_2H_2/CH_4
- R4: C_2H_6/C_2H_2
- R5: C_2H_4/C_2H_6

3.3.2.1 The Doernenburg ratio method

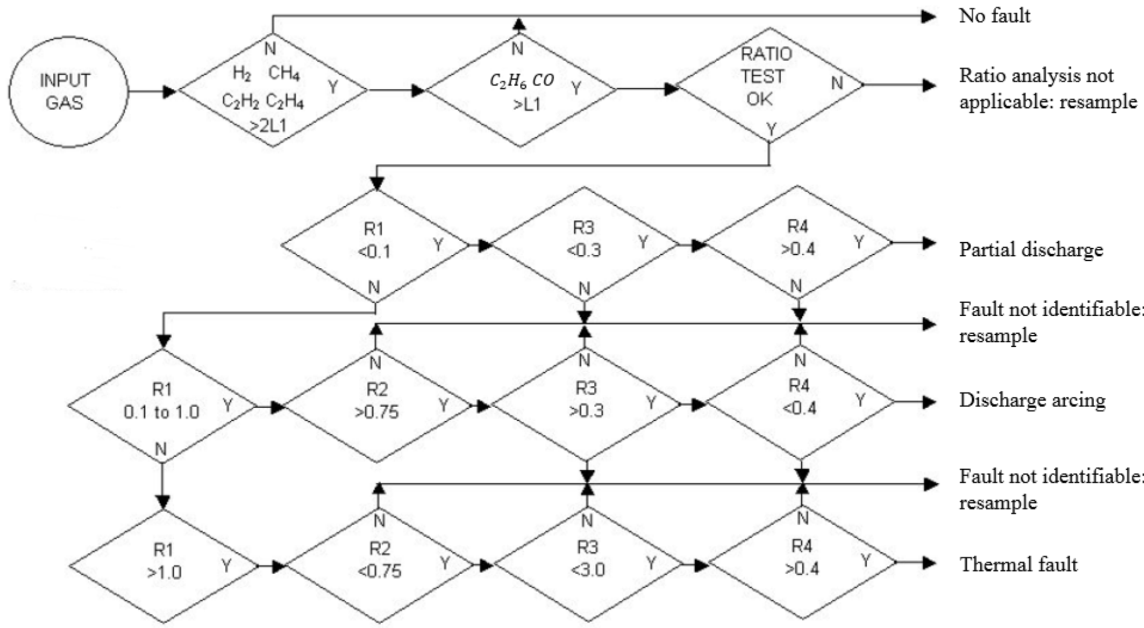

Figure 3.5 Flowchart of the Doernenburg ratio method (IEEE, 1992)

The Doernenburg ratio method uses four ratios, R1 to R4, to make a fault diagnosis for power transformers, and the detailed procedure is as follows (IEEE, 1992):

- **Step 1.** Use a chromatograph to extract the gases and separate them, and then collect the DGA data.
- **Step 2.** If at least one of the gas concentrations for H_2, CH_4, C_2H_2, and C_2H_4 exceeds twice the value of limit L1 and one of the other two gases exceeds the value of limit L1, this case is considered faulty and then proceed to next step. Otherwise, return a non-fault diagnosis result. L1 values are shown in Table 3.2.

- **Step 3.** Calculate the ratios and find the suggested diagnosis result based on Figure 3.5.

Table 3.2 Limit concentrations of dissolved gas

Gas	L1 value (ppm)
H2	100
CH4	120
CO	350
C2H2	1
C2H4	50
C2H6	65

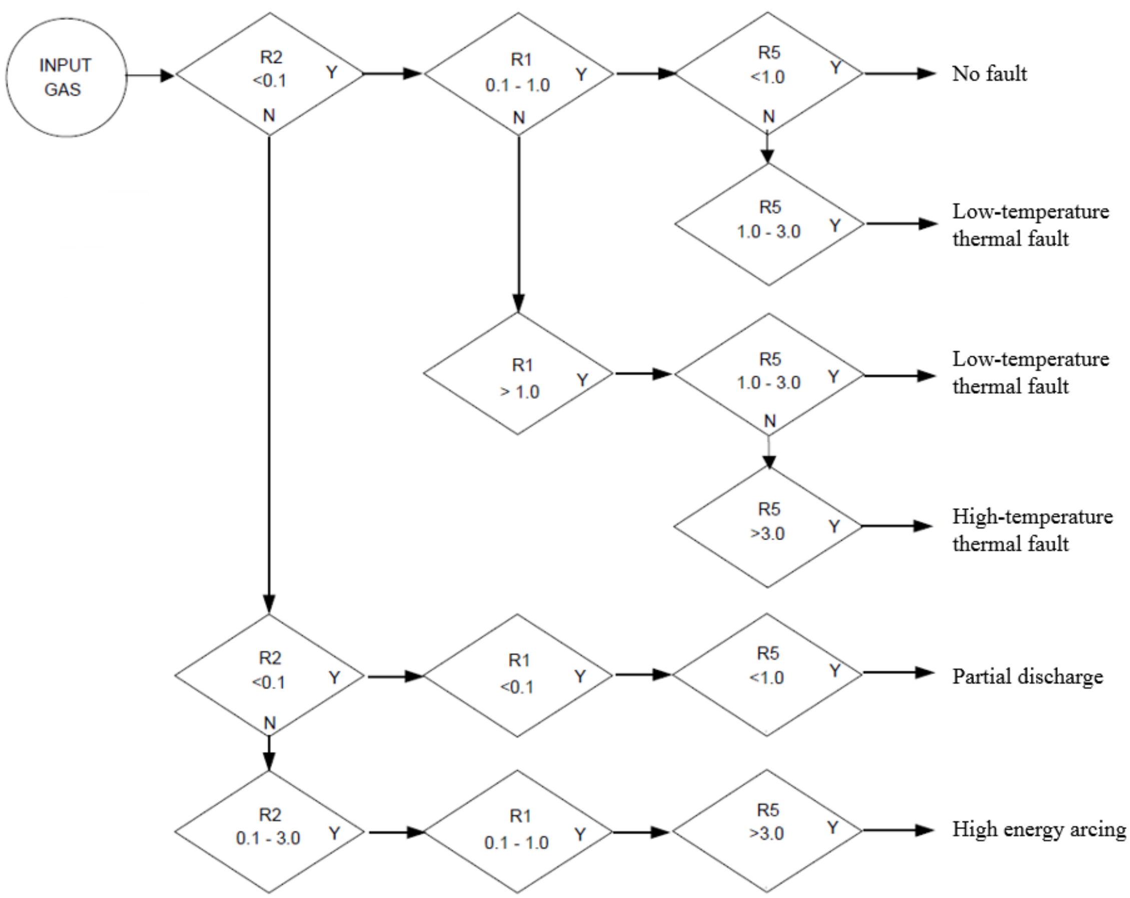

Figure 3.6 Flowchart of the Roger ratio method (IEEE, 1992)

3.3.2.2 Roger ratio method

The Roger ratio method only uses three ratios, R1, R2, and R5 (IEEE, 1992), and it follows a similar procedure as the Doernenburg ratio method. However, this method does not require checking the gas concentrations, and one can calculate the ratios directly. In addition, this method can identify more classes of fault conditions than the Doernenburg ratio method. The flowchart is shown in Figure 3.6.

3.3.2.3 IEC ratio method

The IEC ratio method uses the same three ratios as the Roger ratio method. The latest version uses the graphic rules as shown in Figure 3.7 to classify the specific fault conditions. For each DGA data, the ratios, R1, R2, and R5 should be calculated first, and then map this set of data in Figure 3.7. If the data fall into any of the cuboids labeled with fault names, the diagnosis result can be obtained. This figure can be translated in Table 3.3 so that the procedure for using this method becomes similar with other two ratio methods.

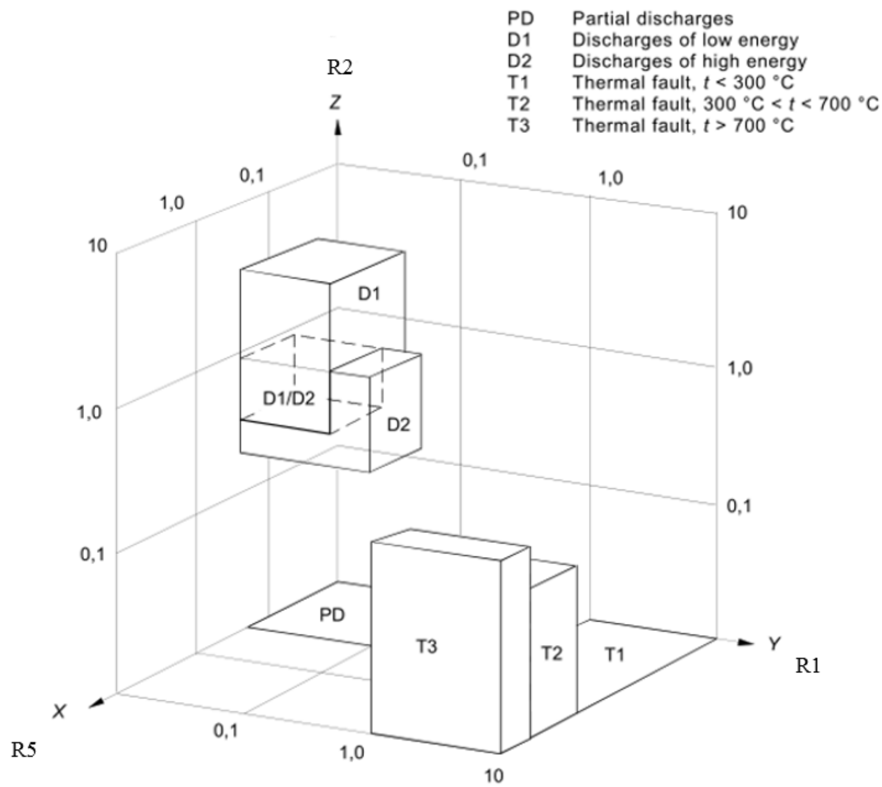

Figure 3.7 The graphic representation of the IEC method (IEC, 1999)

Table 3.3 Diagnosis principle of the ratio method

Fault type	R1	R2	R5
Partial discharge	<0.1	not significant	<0.2
Low energy discharge	0.1-0.5	>1.0	>1.0
High energy discharge	0.1-1.0	0.6-2.5	>2.0
Thermal <300°C	not significant	not significant	<1.0
Thermal 300-700°C	>1.0	<0.1	1.0-4.0
Thermal >700°C	>1.0	<0.2	>4.0

Chapter 4: The Proposed SVM-based Approach for Fault Diagnosis of Power Transformers

In this chapter, the proposed SVM-based approach for fault diagnosis of power transformers is introduced. Section 4.1 describes the challenges that current methods present. To solve these problems, the proposed method is presented in Section 4.2. This method integrates a gas concentration filter and a plurality-voting SVM model developed by using a plurality-voting structure, optimized parameters and balanced datasets from the synthetic minority over-sampling technique (SMOTE). Section 4.3 gives a summary of this chapter.

4.1 Existing Methods and Challenges

The conventional fault diagnosis method cannot show high diagnosis accuracies, and it is a trend to further develop new methods integrated with AI approaches (Bacha et al., 2012b; Ganyun et al., 2005; Illias et al., 2015; Liao et al., 2013; Sarma and Kalyani, 2004a; Zheng et al., 2011). As seen from the literature review, the intelligent methods, like fuzzy logic, the expert system, and ANN, suffer various disadvantages, and a method based on SVM may help to assist fault diagnosis for power transformers.

Table 4.1 Statistic gas concentrations and the time of operation information for healthy transformers

Gas	Under four years	Under ten years	Over ten years
CH_4	70	150	300
C_2H_4	150	200	400
C_2H_6	50	150	1000
C_2H_2	30	50	150
H_2	150	300	300
CO	300	500	700
CO_2	3500	5000	12000

With the increasing investment in electric facilities, many new transformers are being employed. New transformers that have never had a fault occur before do not have many gases generated or dissolved in their insulation oil. As a matter of fact, this condition does not indicate a fault in a new transformer. When SVM methods are used, we do not directly use the gas concentration data as the input for the SVM model because the difference among the concentration values is very large. For example, it is possible for the concentration of H_2 in the insulation oil to be less than 100 ppm in one case, but in another instance, its value could be more than 30,000 ppm. Therefore, instead of directly using concentrations, ratios between gases are commonly used in SVM models, and the ratios are derived mostly from the experience of the conventional ratio methods. However, for most of the data collected in new transformers, the ratios can vary from large to small because the gas concentrations in the new healthy transformers, as shown in Table 4.1, are low (Singh and Bandyopadhyay, 2010). An example is shown in Table 4.2, in which a slight change in the concentration of C_2H_2 can make the C_2H_6/C_2H_2 ratio vary from ∞ to 6, and if this ratio is an attribute of a classifier, the classification results should be questionable. Thus, this situation can make it difficult for classifiers to achieve an accurate diagnosis result.

Table 4.2 An assumption of gas concentrations and ratios

No.	C2H6 (ppm)	C2H2 (ppm)	Ratio
1	60	0	∞
2	60	0.01	6000
3	60	0.1	600
4	60	1	60
5	60	5	12

Several different types of faults can occur in transformers, and we need to be able to diagnose the specific condition inside the transformer. However, the single SVM classifier is only a binary classifier that can only classify different data into two distinct and opposite conditions. This classifier is applicable for problems like classifying a nonzero real number as negative or positive. To classify the faults in detail, almost all of the SVM-based methods currently used in this field are multilayer SVMs (shown in Figure 2.2), where the SVM in each layer is a binary classifier (Bacha et al., 2012b; Fei and Zhang, 2009b). These models consist of several "one-against-rest" SVM layers. The problem is that if the binary SVM in the previous layer classifies the condition incorrectly, no matter how other SVMs classify, the final classification result will be wrong. Therefore, the model relies much on the SVMs in the previous layer, which gives rise to the potential for incorrect classifications.

The third challenge is that typically only limited data with known conditions are available. The truth is that it is hard to detect non-severe conditions inside the transformer, and inspectors always deeply inspect and check conditions inside transformers when their conditions have already become alarming or led to failures. Therefore, we can normally collect very few data for non-severe fault types, such as partial discharge. However, when developing classifiers, a limited amount of data or imbalanced dataset can make the diagnosis accuracy worse.

4.2 Modeling of the Proposed SVM-based Approach for Fault Diagnosis of Power Transformers

Based on the challenges described above, a new fault diagnosis method for a power transformer is presented in Figure 4.1. First, an unknown DGA data should be examined by a gas concentration filter to pick out the low concentration case, i.e. NF (no fault) case. This step is to check whether the set of data satisfies the requirement of low-concentration data that are normally collected from new transformers. If the data belong to an NF case, the gas concentration filter will directly give the final diagnosis result as the NF condition. Otherwise, this information should be diagnosed through a plurality-voting SVM model, which uses a new structure to integrate fifteen binary SVMs. Each binary SVM model is trained by using balanced training datasets from the data oversampling approach, named SMOTE.

To clearly develop the proposed method, Section 4.2.1 shows how to set up the plurality-voting SVM model to avoid the drawbacks of multi-layer SVM models, Section 4.2.2

improved the model developed in Section 4.2.1 by using SMOTE to balance the training data, and Section 4.2.3 introduced the gas concentration filter to deal with the problem of the concentration of data being low.

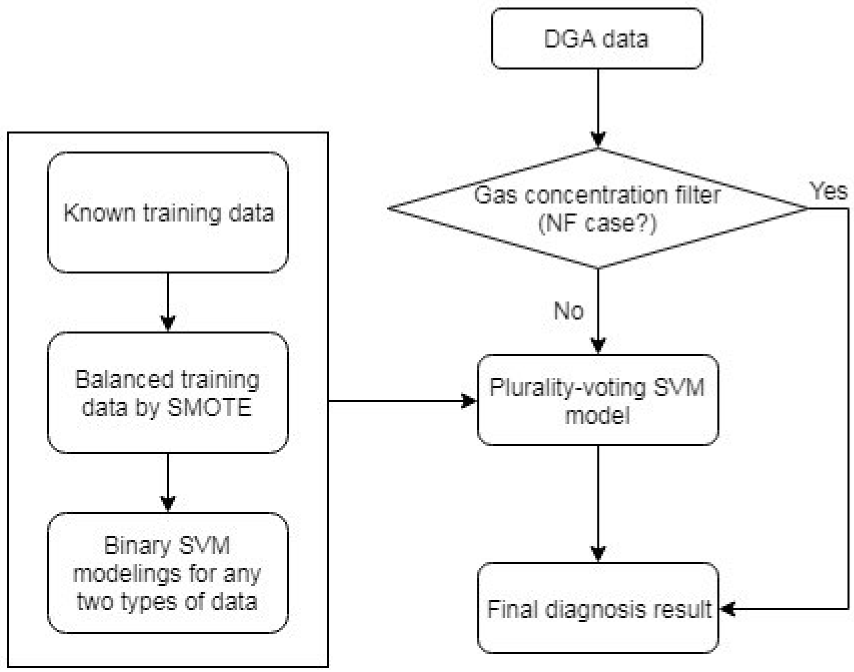

Figure 4.1 Overview of the proposed method

4.2.1 Development of the Plurality-voting SVM Model Using Imbalanced Dataset

In this thesis, we propose to identify total six conditions by using SVM methods, which is according to the IEEE and IEC standard (IEC, 1999; IEEE, 1992). These six cases include five types of fault conditions and the healthy (no fault) condition, as shown in Figure 4.2.

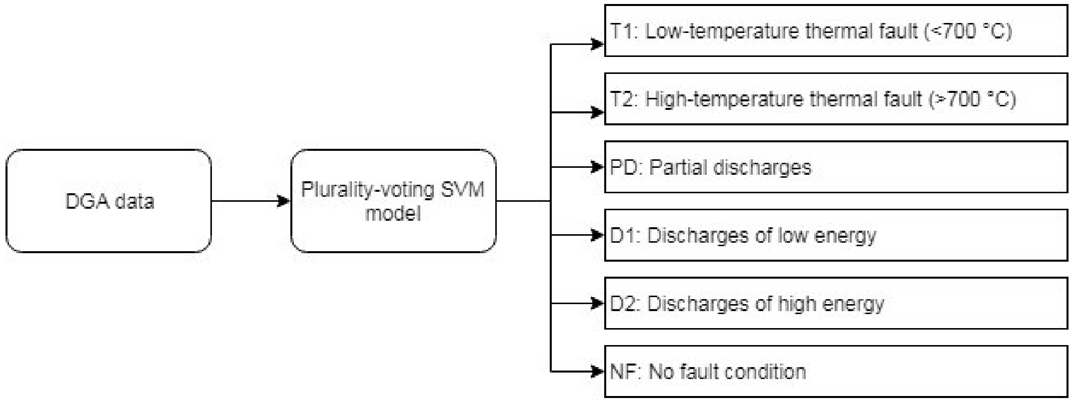

Figure 4.2 Classified fault types by the plurality-voting SVM model

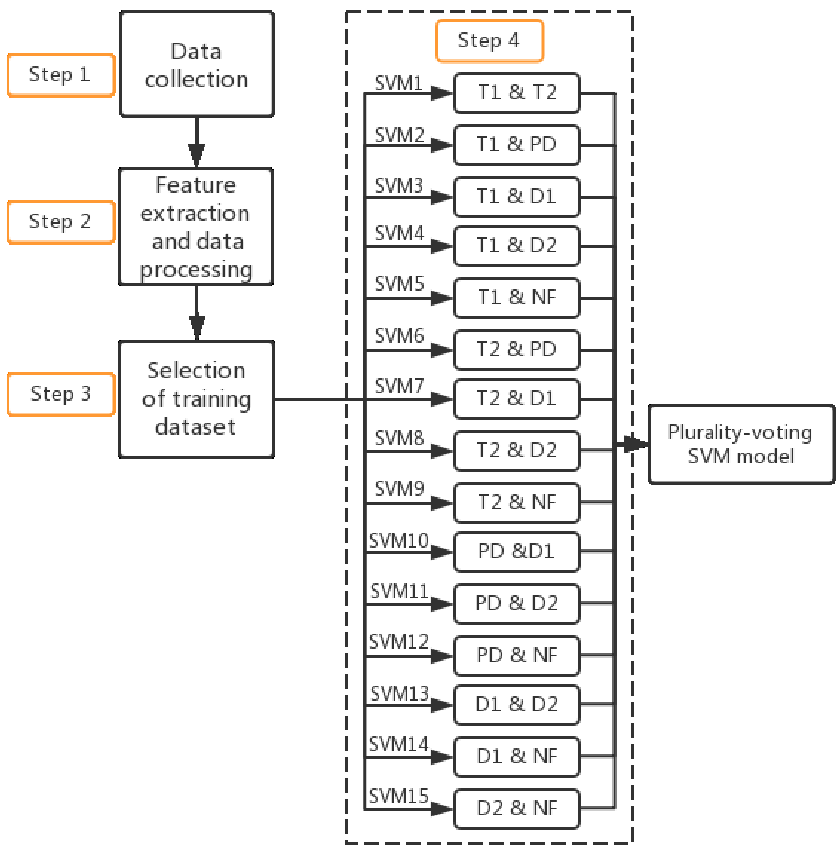

Figure 4.3 Steps for developing the plurality-voting SVM model

The structure of the proposed SVM model is different with multilayer SVM models (see Figure 2.2) and only one layer with fifteen binary SVM models is designed. Thus, this new SVM model does not highly rely on any single binary sub-SVM, and every sub-SVM is equally important.

There are three steps to develop the plurality-voting SVM model, as shown in Figure 4.3.

1. **Data collection**: gathering enough qualified data with inspected operation conditions of transformers
2. **Feature extraction and data processing**: preparing data for training and testing SVMs
3. **Modeling of the plurality-voting SVM approach**: randomly separate the dataset into training data and testing data, develop total fifteen binary SVM submodels for any two types of training data and organize them as the final plurality-voting SVM model

The modeling procedure is described in detail as follows.

4.2.1.1 Data collection

Collecting enough historical DGA data from the power transformer is not easy. DGA data are normally collected from a transformer every two to six months, depending on the age of the transformer. That is to say, through several years of the operation of a transformer, we may only accumulate less than twenty sets of data, and it is very likely that all the data will represent only normal conditions. Thus, successfully collecting all types of data from only one transformer or only one company, including all kinds of fault types and normal data, is tough. Besides, though many companies have their own databases, they do not know what the real conditions inside the transformers for these data are because no obvious faults have been detected due to the lack of deep inspections. Therefore, the raw data available for public research are very limited.

Thankfully, IEC TC 10 database includes many sets of DGA data, which gathered from global corporations such as LCIE, Asinel, Hydro Quebec, and Enel (Duval and dePabla 2001). The classification of faults in this database includes T1, T2, PD, D1 and D2, which can be reliably identified by visual inspection of the equipment after the fault has occurred in service. The information in each set of data contains the values of dissolved gas concentrations, identified fault type, and places where the fault occurred. It also contains 50 sets of normal-condition data. Most of the data employed in this paper are from this database, and a small portion of the data are collected from credited literature (Duraisamy et al., 2007b; Yadaiah and Ravi, 2011). The detailed data information is shown in Table 4.3.

Table 4.3 The overview of the data employed in this study

Data Sources	Data type						
	T1	T2	PD	D1	D2	NF	Total
Duval and dePabla, 2001	16	18	9	26	48	50	167
Duraisamy et al., 2007b					2	2	4
Yadaiah and Ravi, 2011	3	2				4	9
Total	19	20	9	26	50	56	180

4.2.1.2 Feature extraction and data processing

Feature selection from the raw data is an important procedure for computational classifiers. Any good classifier relies much on the selected features or attributes derived from the raw data (Chandrashekar and Sahin, 2014; Liang et al., 2014; Zhang et al., 2014). The method of extracting features provides a way of maximizing the pattern recognition performance, and good features help the computational classifiers understand more useful knowledge from the data in machine learning applications (Chandrashekar and Sahin, 2014).

The proposed method based on an SVM approach definitely needs good features as well. Therefore, features from data should be selected to use in the SVM modeling procedure. As we mentioned in Chapter 2, almost all the AI-based computational methods use ratios from the conventional methods as the input in their models instead of directly using the gas concentrations. In this study, we use five ratios, R1 to R5, as shown in Section 3.3.2, as the selected features of the modeling. We take the logarithmic transform for the ratios to decrease the great difference among the ratio values, as shown in Eq. 4.1. Using this equation, the raw data collected in step 1 are interpreted from Figure 4.4 to Figure 4.9.

$$LR_i = \log(R_i) \tag{4.1}$$

Where LR is the value of the logarithmic transform of the gas concentration ratios.

Figure 4.4 shows the LR values of the T1 data. As we can see, LR4 is the maximum value among these five LR values, which means the R4 (C_2H_6/C_2H_2) value is very large, and C_2H_2 is not significant compared with the C_2H6 values in the low-temperature thermal fault cases. Also, the C_2H_6-related values, LR5 values, are close to zero, which means that the differences between C_2H_4 and C_2H_6 values are small. Therefore, we can conclude that C_2H_4 and C_2H_6 are the dominant gases for the T1 cases, which also confirms the conclusion from the key gas method, as shown in Table 3.1. Apart from the analysis of the dominant gases, if we only consider the values of the LRs, we can see that the LR2 and LR4 are all negative and almost all LR4 values rank at the top, which is the key feature of T1 cases.

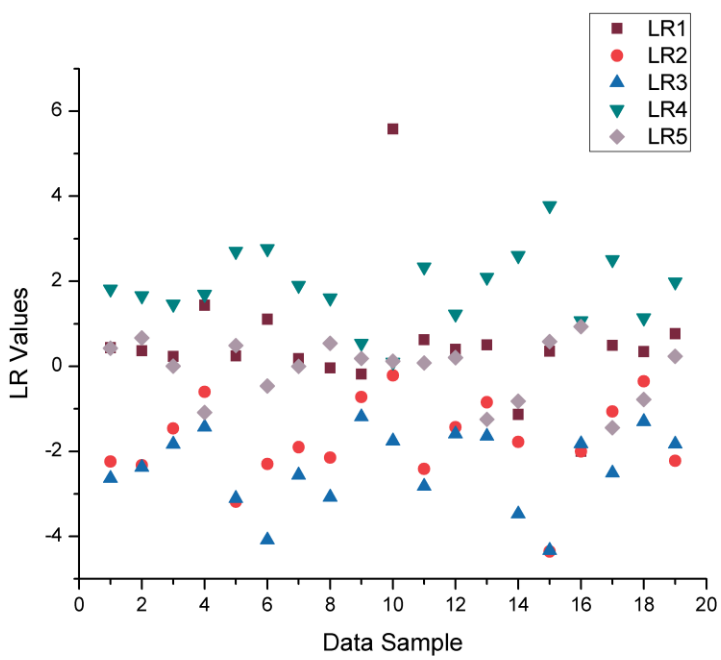

Figure 4.4 LR values of the 19 T1 data

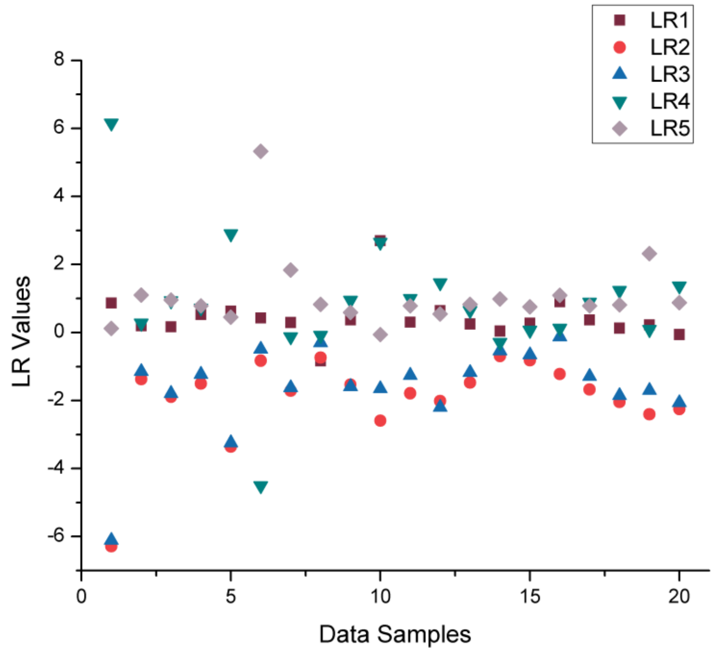

Figure 4.5 LR values of the 20 T2 data

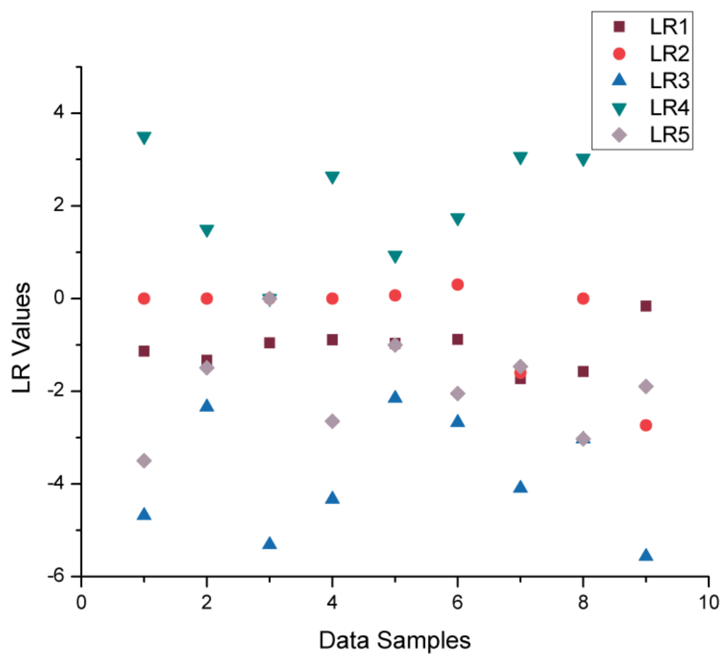

Figure 4.6 LR values of the 9 PD data

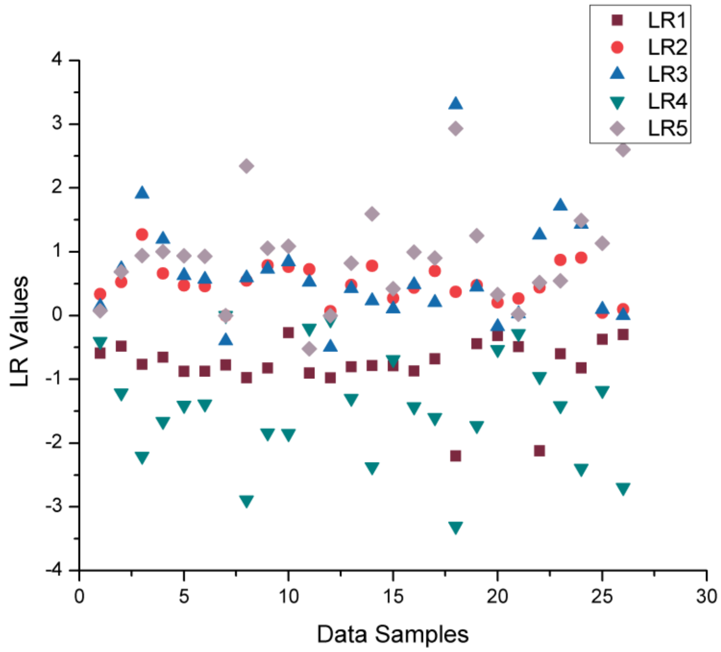

Figure 4.7 LR values of the 26 D1 data

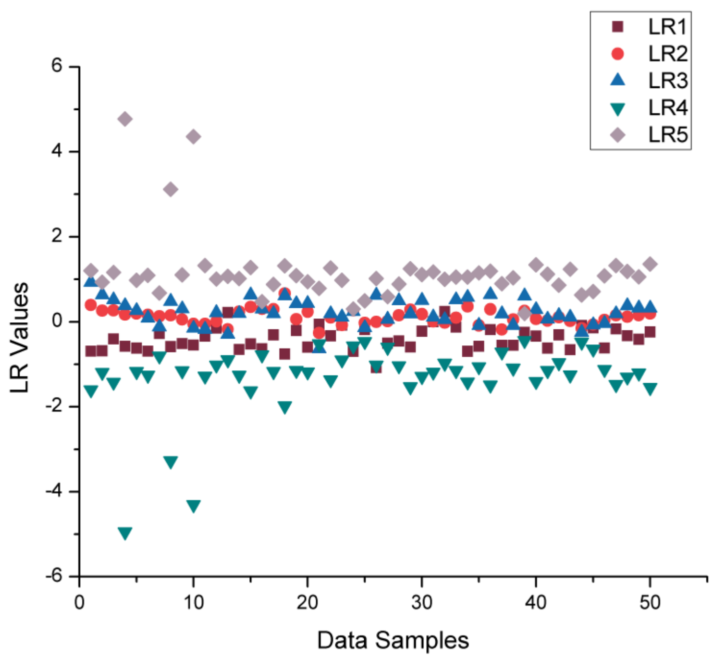

Figure 4.8 LR values of the 50 D2 data

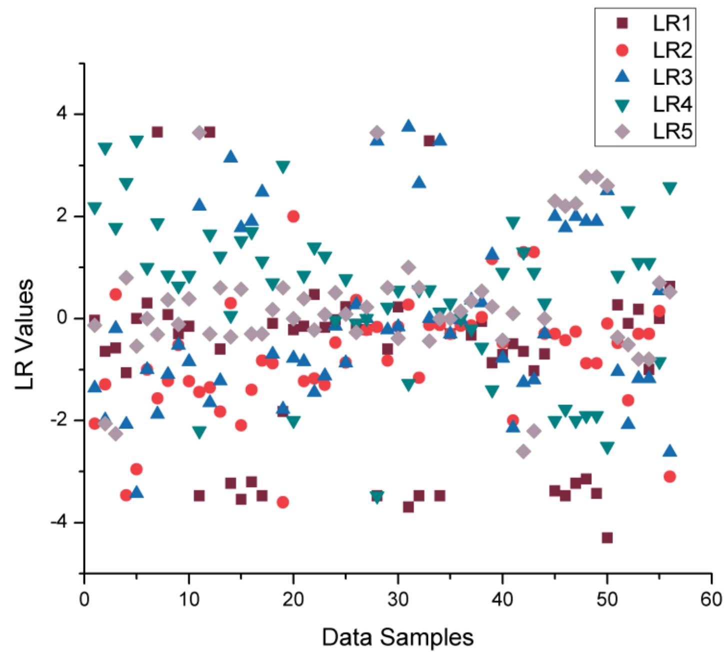

Figure 4.9 LR values of the 56 NF data

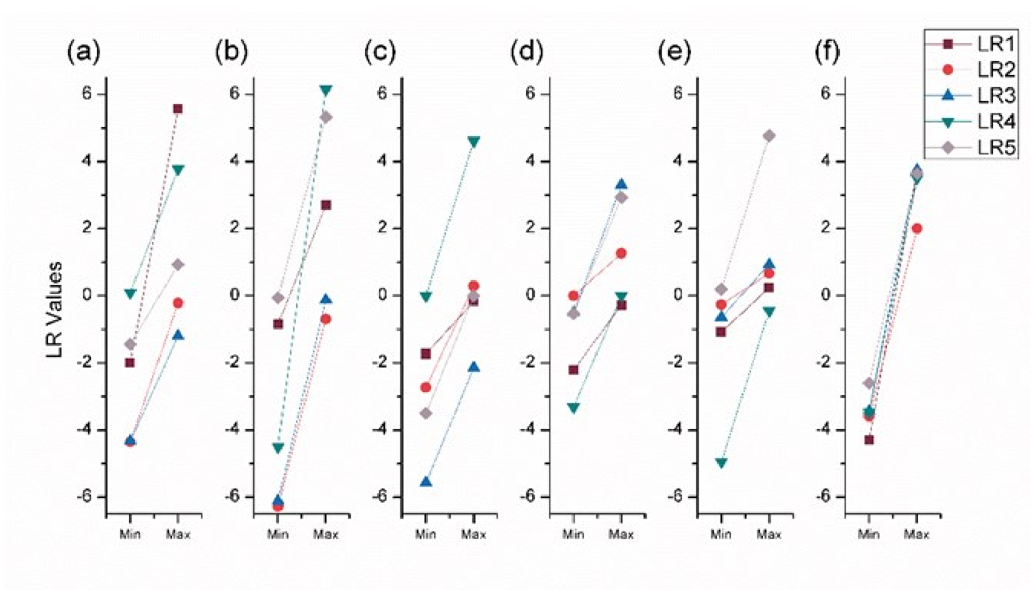

Figure 4.10 LR value ranges for each fault type

Similarly, we can find regularities for some other cases. As we can see in Figure 4.5, LR2 and LR3 are always negative in the case of T2 fault and almost equal to each other. Most of the LR5 values are positive and within the range of [0, 1]. In Figure 4.6, there is a general rule that the order of the LR values, from maximum to minimum, is LR4 > LR2 > LR1 > LR5 > LR3 in each set of data. We can also find that only the LR4 values are positive and the LR1, LR3, and LR5 values are all negative. However, in the case of D1 and D2 fault types, LR4 values are negative and almost the smallest values, as shown in Figure 4.7 and Figure 4.8. LR5 values are the biggest in all D2 cases as shown in Figure 4.8. By contrast, for T1 and PD cases, LR4 values are almost the biggest values among all the other LR values, but the difference between these two cases can be recognized by the values of LR2. For D1 and D2 cases, LR4 has the lowest values, but it is different in thermal fault cases.

Figure 4.10 shows the LR value ranges, which can help compare the differences between the data for different fault types. We can see that if the LR1 value is positive, the fault type is most likely a thermal fault, T1 or T2. Positive LR2s always match a D1 or D2 discharge fault. LR1 and LR2 are both low in thermal fault cases. LR3 is the smallest value in the DP fault. LR4 is low compared to other LR values in discharge fault cases. In D2 situations, the largest LR value is more likely to be LR5. All the LR values seem in a similar range from around -4 to 4 in the NF cases. As for the NF case, there is not a general rule that we can summarize, but it is always the case in the ratio method that if we cannot identify whether a case is in a fault condition, it is an NF case.

Therefore, some rules from different ranges of LR values in each case are somehow obvious. These LR values can contribute as input features to help classify the various situations.

4.2.1.3 Modeling of the plurality-voting SVM approach

For building SVM classifiers, known data should be split into training sets and testing sets. The training data are used to develop the SVM model, and the testing data test the performance of the model. In the 180 total sets of data, there are only nine records of PD faults and fewer T1 and T2 fault records. To guarantee enough data for testing, one-third of the data was used to test the model. If only a small portion of the data were used to train the model, the model would be inadequate for classifying other data accurately. However, if a large portion of the data were used to train the model, the remaining data could not be used to assess the quality of the model. The information for the training and testing samples is shown in Table 4.4, including 37 training samples for the NF case, over six times the amount of PD fault datasets. This difference exists because transformers operate normally most of the time and because non-serious PD cases are difficult to detect. Thus, companies do not have much data on PD cases compared to NF cases.

Once the datasets are ready, it is time to model SVMs. As discussed in Section 4.1, an issue with the multi-layer SVM model is that when the binary SVM in the previous layer classifies the condition incorrectly, the final classification result will also be incorrect. To overcome this problem in the multi-layer SVM model, a plurality-voting SVM structure for using the binary SVM classifiers is proposed in Figure 4.11.

Table 4.4 Overview of the training and testing samples

Data type	Total samples	Training samples	Testing samples
T1	19	13	6
T2	20	13	7
PD	9	6	3
D1	26	17	9
D2	50	33	17
NF	56	37	19
Total	180	119	61

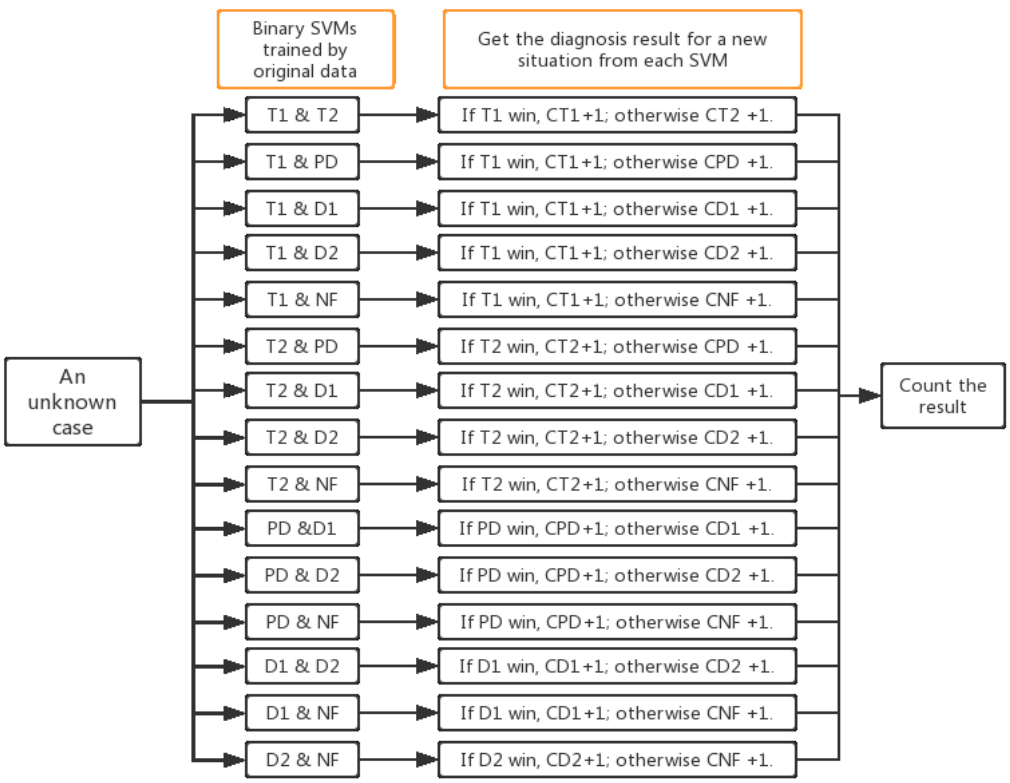

Figure 4.11 A voting system by binary SVMs

The first step is to use any two types of processed feature inputs to train the sub-SVM models. Because there are total six types of data, the models should have total $C_6^2 = 6\times(6-1)/2 = 15$ binary SVMs. For example, we use T1 and T2 training samples to develop SVM1 so that total 26 sets of data are trained. Similarly, we use T1 and PD training samples (19 in total) to develop SVM2. Using SVM1, we can only get the classification result of T1or T2, and, likewise, we can only get the classification result of T1or T2 from SVM2. Unlike the multilayer models, in this model, every one of the fifteen SVMs plays the same role and the diagnosis result does not rely much on any of the SVMs. This approach is similar to many real-world election activities. For example, every citizen should have the same right to select a president, and it may cause problems if citizens select the president based much on the recommendation from only one citizen.

The kernel function should then be determined to use in the SVM modeling. Linear, polynomial, Gaussian radial basis, and sigmoid kernel functions are commonly used in the machine-learning field, and in this study, the sigmoid function served as the kernel function. Aside from the selection of kernel functions, the free parameters that should be defined by users are the penalty factor c and the γ (a parameter in the kernel function). Genetic algorithms (GA) are widely employed to choose machine inputs and parameters (Chen et al., 2014; Fei and Zhang, 2009a; Tewari et al., 2012). GA was used to determine the c and the γ. The flowchart of the GA is shown in Figure 4.12.

Below is the step-by-step explanation of the GA:

1. Randomly generate a chromosome population in which each chromosome is composed of binary numbers, as shown in Figure 4.13. The binary numbers can be decoded into decimal numbers.
2. Obtain the candidate parameters of c and γ by decoding the chromosome.
3. Use the parameters obtained above and some of the data to train the SVM model and test the performance of the model, which is determined by a fitness function.
4. Calculate the fitness function, which is the classification accuracy of training samples in this study.

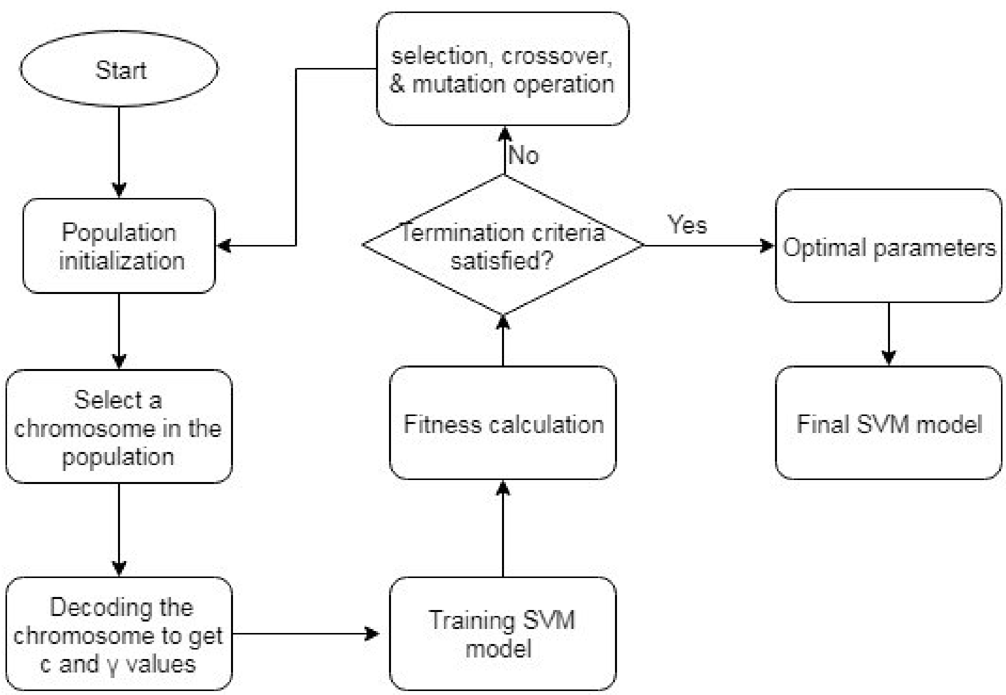

Figure 4.12 Flowchart for the genetic algorithm

5. If performance satisfies a designed stopping criterion (the classification accuracy is 100% for this study), parameters are obtained; otherwise, use selection, crossover, and mutation operators to generate the offspring of the existing population and test a new chromosome against the stopping criterion.
6. Repeat this algorithm until the satisfied fitness accuracy is obtained. If the population is generated over 200 times without finding a satisfactory chromosome, stop using the algorithm and select the chromosome with the best fitness accuracy.

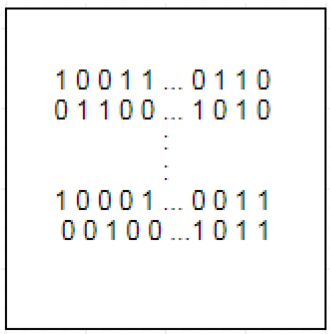

Figure 4.13 An example of chromosome population

Once the parameters are obtained, the binary SVMs can be modeled. To use the model, we can feed the processed under-testing data into the fifteen SVMs and find the corresponding results. For example, if the binary SVM1 (T1&T2 classifier) classifies it as T1 case, it means the SVM votes for T1 as the winner, and the values of the variable CT1 will increase by 1. Moreover, then, if SVM2 (T1& PD classifier) also gives the result as T1case, T1 will get another vote and CT1 will equal with 2; otherwise, CT1 and

CTD both equal with 1. After finding all the results from the fifteen SVMs, the values of CT1, CT2, CPD, CD1, CD2, and CNF will be obtained. The final diagnosis result is determined by finding which case has the most votes, meaning that the maximum value of CT1, CT2, CPD, CD1, CD2, and CNF indicates the diagnosis result from the plurality-SVM model. Here is an example to illustrate how the result is determined. If the output from these 15 SVMs is [T1, T1, T1, D2, T1, T2, T2, T2, NF, PD, PD, NF, D1, D1, D2], T1 gets the maximum votes (4 votes) and T1 is the final diagnosis result by the plurality-voting SVM model.

4.2.2 Development of the Plurality-voting SVM Model Using Balanced Dataset

As we can see, in Table 4.4, we only have a minimum of 6 records of PD cases, but a maximum of 37 records of NF cases. The ratio between the sizes of these two datasets is roughly 1:6, which is highly imbalanced. If these data are used for training, problems caused by imbalanced datasets cannot be avoided. As discussed in Section 3.2.1, the over-sampling and under-sampling approaches both work for imbalanced dataset problems, but over-sampling approach is the only effective approach in situations in which limited data exist for the minority class.

The SMOTE algorithm (Chawla et al., 2002) is an over-sampling method which over-samples minority classes by forming synthetic data samples instead of simply duplicating samples. The method for creating synthetic data is described below:

1. Find a sample, $\tilde{A}$, in the minority class and then calculate the distance, $\|r\|$, between itself and other samples in the same class.
2. From the distance results, find the nearest, k, neighbors to $\tilde{A}$, and randomly select one, A_i.
3. Calculate the difference between $\tilde{A}$ and A_i and multiply the difference by a random number from [0,1].
4. Add the result from the previous step to $\tilde{A}$ to obtain a synthetic sample.

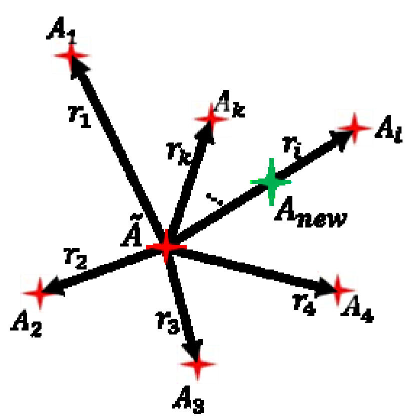

Figure 4.14 The illustration of the SMOTE algorithm

Figure 4.14 visualizes the steps above. Red stars represent the real minority class, and the green stars represent sets of the newly created synthetic data. The mathematical expression is the equation (4.2).

$$A_{new} = \tilde{A} + rand(0,1) \times (A_I - \tilde{A}) \tag{4.2}$$

The SMOTE was used to balance the data. That is, before modeling the sub-SVM models, this data oversampling approach is added. To obtain a reasonable result, none of the testing data should be used for training. In other words, the model should not be allowed access to knowledge related to the testing sets. Only the training sets can be used when generating data through the SMOTE approach. After this approach, the data information can be summarized in Table 4.5.

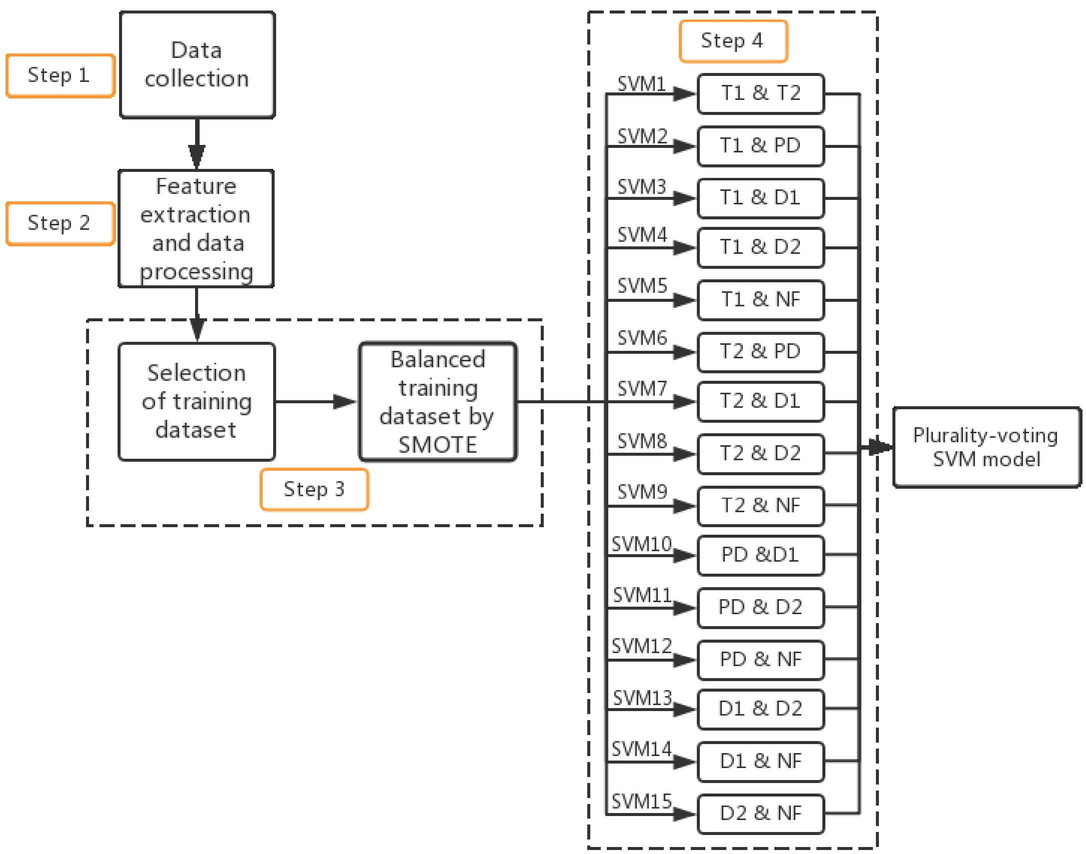

Figure 4.15 Steps for developing the plurality-voting SVM model using balanced dataset

To make it clear, this improved procedure is shown in step 3 of Figure 4.15, and balanced datasets are used in step 4. Other steps are same as presented in Section 4.2.2.

Table 4.5 Overview of the balanced training dataset

Data type	Total samples	Training samples	Testing samples	Training samples from SMOTE	Total training samples
T1	19	13	6	24	37
T2	20	13	7	24	37
PD	9	6	3	31	37
D1	26	17	9	20	37
D2	50	33	17	4	37
NF	56	37	19	0	37
Total	180	119	61	103	222

4.2.3 The Gas Concentration Filter

The developed plurality-voting SVM model is good for core fault diagnosis function for power transformers. However, special cases in which gas concentrations are low are more likely to occur in new transformers. We used the processed ratio-related input to develop the above SVM model, and the diagnostic accuracy was not good when applying the method to low-concentration data.

To solve this problem, we compared the data with the data from the literature and found that gas concentrations for new and healthy transformers are very low. For example, the concentration of CH_4 in a thermal fault can be over 10^4 ppm, but the CH_4 data from the low-concentration data were all under 10^2 ppm. We used the Doernenburg ratio method due to the consideration of gas concentrations. The Doernenburg ratio method used Table 3.2 to check every gas concentration first, and if the concentrations were under the limits, instead of using ratios to make further diagnosis, the diagnosis results were determined as NF.

To validate this gas concentration approach, we found that it is effective for the data summarized in Table 4.6. Although this approach is not applicable for all NF cases, it is necessary to use it in qualified low concentration cases. It is necessary to add this rule before using the SVM model to make diagnosis results precise, especially for data collected from newly employed transformers. The gas concentration limits in the proposed gas concentration filters are shown in Table 4.7.

Table 4.6 Low gas concentration cases

H2	CO	CH4	C2H4	C2H6	C2H2	NF cases?	Source
31	260	6	3	8	1	Yes	Duval and dePabla, 2001
22	180	7	5	4	0.05	Yes	Duval and dePabla, 2001
80	0.05	18	0.05	20	1	Yes	Duval and dePabla, 2001
170	0.05	16	0.05	8	1	Yes	Duval and dePabla, 2001
36.1	85.4	15.5	16.1	2.8	0	Yes	Hong et al., 2015
5	0.05	21	63	19	0.05	Yes	Duraisamy et al., 2007

The combination of the proposed SVM model and the gas concentration filter can also save the time of taking unnecessary maintenance actions, as normal ratio-related SVM models are not very accurate for low concentration data, and the costs for cases of unhealthy conditions are higher than those for cases of low-temperature thermal faults being diagnosed as high-temperature thermal faults.

Table 4.7 Gas concentration limits in the gas concentration filter

Gas	Concentration limit (ppm)
H2	200
CH4	240
CO	350
C2H2	2
C2H4	100
C2H6	65

Therefore, when making the fault diagnosis for a new set of data, it is first necessary to compare the gas concentrations to the limits. If the gas concentrations are under the respective limits shown in Table 4.7, results can be directly diagnosed as NF. Otherwise, the proposed SVM model can be used to identify what is going on in power transformers.

4.3 Summary

In this chapter, we first analyzed the current problems and challenges existing in industrial applications and academic research. These challenges include:

- the limited DGA data with actual and recorded inspected conditions
- the imbalanced DGA data
- the inaccurate diagnosis for low-concentration data through the SVM approach

To solve these problems, we proposed an improved model based on the SVM method (see Figure 4.16). As we highlighted in Figure 4.16, this method combined a gas concentration filter and an improved plurality-voting SVM model. The gas concentration filter solved problems raised by low concentration data. The plurality-voting SVM model was developed using a new model structure and it integrated the SMOTE approach to dealt with imbalanced dataset and limited data problem. The restructured SVM model avoided dependence on any single binary SVM model.

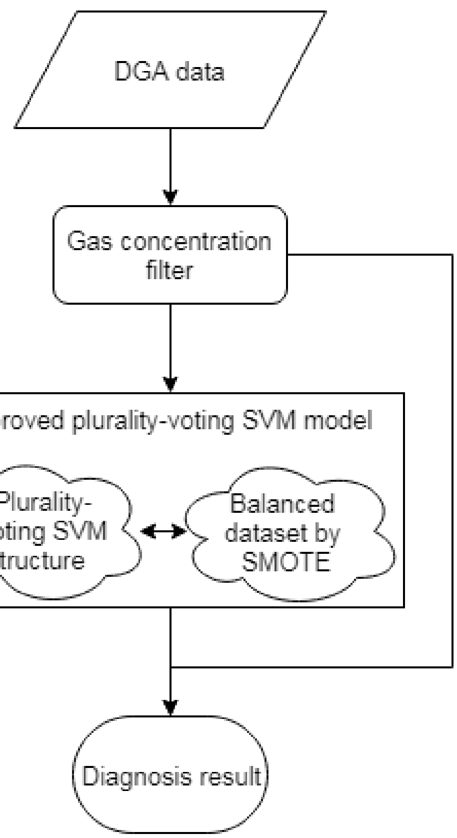

Figure 4.16 Flowchart of the diagnosis procedure of the proposed method

Chapter 5: Method Validation and Comparison

The goal of this chapter is to validate the proposed method through comparisons with existing methods. Section 5.1 shows all the methods used in this chapter. Section 5.2 lists the results from these selected methods and compares diagnosis accuracy. Section 5.3 presents some specific cases with the corresponding diagnosis results. Section 5.4 presents a new way to compare different methods, which compares the cost by misdiagnosis when we consider the loss by maintenance activities.

5.1 Methods Used in the Comparison

The results from the AI-based methods and the traditional ratio-based methods are used to make good comparisons. The ratio-based methods include the Doernenburg ratio method, Roger ratio method, and IEC ratio method. The AI-based methods include the proposed method, existing SVM method, and ANN method.

5.1.1 Explanations of the Results from Ratio Methods

In this subsection, we use the Doernenburg ratio method, the Roger ratio method, and the IEC ratio method for the purpose of comparison. The Doernenburg ratio method can only roughly diagnose faults. For example, with this method, one can only get a label of thermal fault, rather than the thermal fault with a level of temperature. The method was treated as accurate if it roughly correctly diagnosed the data, as shown in Table 5.1. Similarly, the Roger ratio method can only diagnose a total of five conditions. Table 5.2 shows how we treat the conditions in the Roger ratio method. However, we directly use the results from IEC ratio method to compare with the proposed method, as they used the same diagnosis labels.

Table 5.1 Condition labels for the Doernenburg ratio method

The conditions in the Doernenburg	The conditions we used in this study
No fault	NF
Partial discharges	PD
Thermal fault	T1&T2
Discharge arcing	D1&D2

Table 5.2 Condition labels for the Roger ratio method

The conditions in the Roger	The conditions we used in this study
No fault	NF
Partial discharges	PD
Thermal fault < 700 °C	T1
Thermal fault > 700 °C	T2
High-energy arcing	D1&D2

Because ratio methods are all rules-based and do not need training data, we tested these methods by using all the known 180 datasets.

5.1.2 Explanations of the Results from AI-based Methods

To make full use of the data, we introduced a ten-fold cross-validation procedure. We randomly separated the 180 samples into training samples and testing samples for ten times. With this process, we can guarantee use of two-thirds of the data from the real, original 180 samples and then use SMOTE to balance the training dataset so that each fault type has 37 datasets in total. The total number of training samples is $37\times6 = 222$, as shown in Table 4.5. The remaining one-third of real data constitute the testing data. In each round of the ten-fold cross-validation process, there are different training and testing samples that are independent from each other. This cross-validation process has a lower variance than the one-time data sampling procedure, which is meaningful in the case that the amount of data available is limited. Although it is possible for randomly selected data to make a good model, other times they cannot form reliable models. The reason is that data play a key role in classifiers. For example, classifiers can perform well if we only use them to classify children and aged people based on the data of their ages, because the data are easy to be classified and classifiers are reliable based on desired data. Therefore, ten-fold cross-validation can decrease the possibility of performance variance caused by the variance of selected data.

For the purpose of comparison, an ANN model was implemented for transformer fault diagnosis as well. It is a probabilistic neural network (see Figure 5.1) with four layers, an input layer, pattern layer, summation layer, and output layer. In the input layer, there are five neurons since we have five LR values as input features for classifiers defined in this thesis. The pattern layer, including the same number of neurons with the number of training data, calculates the distances between input vectors and row weight vectors, and the distances are measured by radial basis function nonlinearly. The summation layer is to find the summation results from the previous layer for each type of data, and the type of the highest results will be determined as the final output in the output layer. The Matlab neural network toolbox is used to obtain the diagnosis results for this ANN model.

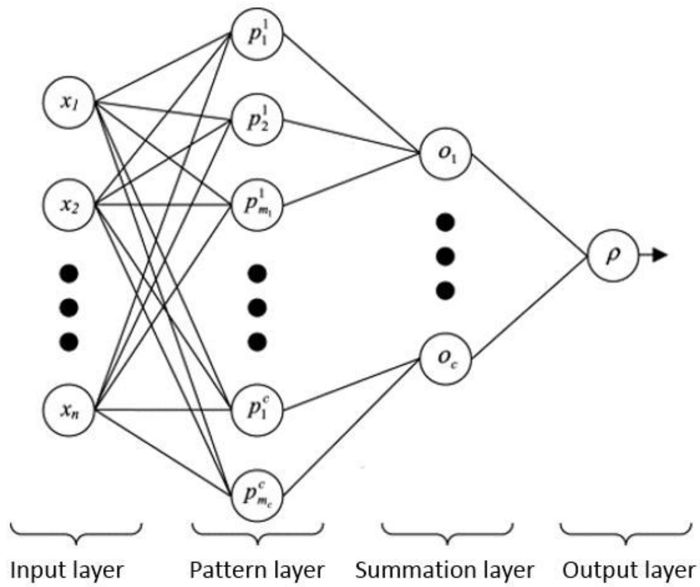

Figure 5.1 The probabilistic neural network structure

In addition, a four-layer SVM model is presented based on Bacha et al.'s model (2012a) and its structure is shown in Figure 5.2. Similar with Bacha’s model, SVM1 is used to separate data into NF cases and thermal/discharge fault cases. SVM2 is employed to divided data into discharge fault and thermal fault. To further classify the discharge fault, SVM3 and SVM5 are used to classify DGA data into PD, D1, and D2 cases. The T1 and T2 thermal fault is classified by SVM4. In Bacha’s model, T1 cases were further classified. However, in this thesis, we do not consider to classify this thermal case in detail due to the availability of known data. Therefore, ignoring this detailed classification procedure, the model in Figure 5.1 should have a better diagnosis accuracy performance than Bacha's original model. GA is also used to find proper parameters for the parameters c and γ in each binary SVM.

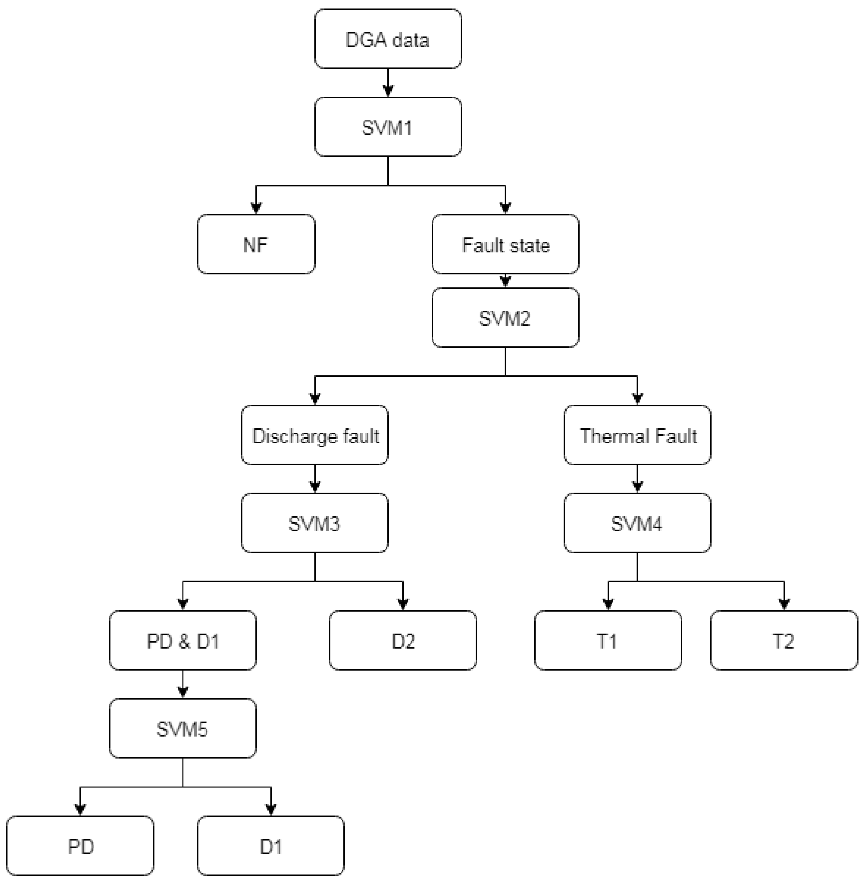

Figure 5.2 The structure of a four-layer SVM model (Bacha et al. 2012a)

Besides, we tested the testing data using an SVM model which is a plurality-voting SVM model and does not include the gas filter or the SMOTE approach (named SVM in Table 5.3) and an SVM* model that included the SMOTE approach but not the gas concentration filter. This information is shown in Table 5.3.

To clarify, we used the same real training data to train the ANN and SVM models and all these models were validated with the same testing data in each round.

Table 5.3 Models used in the comparison

Model name	Structure of SVMs	SMOTE	Gas concentration filter
Multi-layer SVM	Figure 5.1	✗	✗
SVM	Figure 4.3	✗	✗
SVM*	Figure 4.15	✓	✗
Proposed method	Figure 4.15	✓	✓

5.2 Comparison of Diagnosis Results and Accuracies

The results from ratio methods are shown in Table 5.4. The results show that the IEC ratio method provides the most accurate diagnoses, as it correctly finds NF cases. However, it cannot recognize T1 cases, while the Doernenburg ratio method and the Roger ratio method can. The Roger ratio method is better for identifying discharge faults than the other two ratio methods. However, the Doernenburg ratio method performs the best among these three methods to detect the thermal faults. The results from the ten rounds of modeling for the AI-based methods, based on 61 total sets of testing data, are shown in Appendix and the results are summarized in Table 5.5.

Table 5.4 Statistical diagnosis results of the ratio methods

Type	Samples	# of correct diagnosis		
		Doernenburg	Roger	IEC
T1	19	8	8	0
T2	20	13	12	12
PD	9	1	4	1
D1	26	11	15	16
D2	50	27	42	42
NF	56	36	35	53
Total	180	96	116	124

Based on the results in Table 5.4 and 5.5, the diagnosis accuracies of these methods are compared in Figure 5.3. The accuracies from the traditional ratio methods are based on all 180 datasets, whereas the accuracies from the AI-based methods are only based on 61 testing datasets. Using the results above, the performance comparisons of the key components integrated in the proposed method are shown in Section 5.2.1 to 5.2.3.

Table 5.5 Diagnosis results from SVM and the proposed method

Round #	ANN	Multi-layer SVM	SVM	SVM*	Proposed method
1	45	46	44	49	50
2	40	46	49	49	50
3	43	50	46	47	48
4	39	45	42	48	49
5	41	48	46	49	49
6	47	49	50	51	51
7	40	44	46	48	49
8	43	45	46	48	48
9	40	44	49	49	49
10	46	45	50	48	48
Average	42.4	46.2	46.8	48.6	49.1

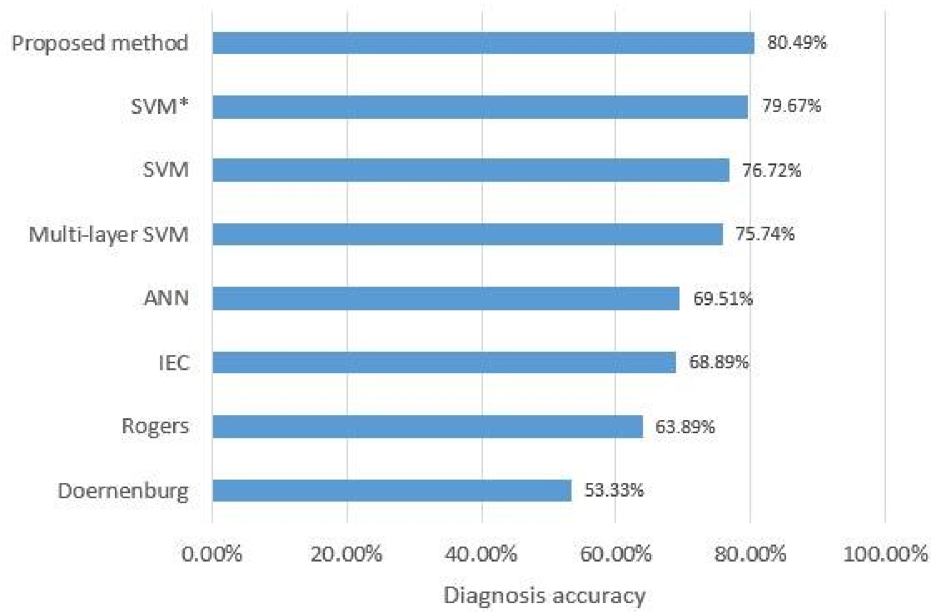

Figure 5.3 Comparison of overall diagnosis accuracy

5.2.1 Comparison Between Multi-layer SVM Model and the Proposed One-layer SVM Model

Table 5.5 shows that the multi-layer SVM model can only correctly diagnose an average of 46.2 out of 61 cases, but the proposed one-layer plurality-voting SVM models can correctly identify an average of 46.8 cases. The multi-layer SVM model can perform better than the ratio methods. The proposed one-layer SVM model can achieve a diagnosis accuracy of 76.72%, which is higher than all the traditional methods, ANN method, and the multi-layer SVM method. This comparison confirms the advantage of the one-layer plurality-voting SVM models over the multi-layer SVM models.

5.2.2 Comparison Between the Methods with/without the SMOTE

As we can see in Figure 5.2, the SVM and SVM* are both better than the traditional ratio methods, ANN method and multi-layer SVM method. By adding the SMOTE approach, the number of correct diagnoses increases from an average of 46.8 to an average of 48.6, and the fault diagnosis accuracy improved to 79.67% from 76.72% of the method without the SMOTE. This result confirms that the SMOTE approach is effective for helping SVM models deal with imbalanced datasets.

5.2.3 Comparison Between the Methods with/without the Gas Concentration Filter

With the gas concentration filter, the proposed method beats the SVM* method by an average of 0.5 correct diagnoses, and diagnosis accuracy improved from 79.67% to 80.49%. This improvement seems not high, which is because there are not many low-concentration data in the known dataset.

To further validate the improvement by the gas concentration filter, 50 sets of data were generated. These data are randomly generated, in which each gas follows a continuous uniform distribution on the interval [0, L], in which L is the gas concentration limit in the proposed gas concentration filter as shown in Table 4.7. Therefore, the six gas concentrations in a set of generated data are subjected to the distributions as follows:

- H_2~*U[0, 200]*
- CH_4~*U[0, 240]*
- CO~*U[0, 350]*
- C_2H_2~*U[0, 2]*
- C_2H_4~*U[0, 100]*
- C_2H_6~*U[0, 65]*

In this way, all the 50 sets of data are guaranteed as the low-concentration data. These data satisfy the requirement of being the low-concentration data collected in transformers under four-year employment (Singh and Bandyopadhyay, 2010). After getting these data, we will test them using the SVM* model and the proposed model. As the proposed method owns the gas concentration filter, it can successfully screen out all the low-concentration data and identify these data as NF cases. To compare with the method without a gas concentration filter, the data should be processed through the same procedures in Step 2 of Figure 4.3, and then test them using the SVM* model as defined in Table 5.3 to obtain the predicted diagnosis results.

Figure 5.3 shows the predicted results of using the generated low-concentration data from the SVM* model that is trained using the balanced datasets from SMOTE approach as shown in Table 4.5. In Figure 5.3, labels from 1 to 6 represent the T1, T2, PD, D1, D2, and NF, respectively. As we can see, most of the data were diagnosed as thermal or PD cases and only 3 out of 50 samples were correctly identified as Label 6. In contrast, with the concentration filter, the proposed method can perfectly deal with these data with a diagnosis accuracy of 100%. This comparison points out a great difference between the methods with/without the gas concentration filter when we only consider the low-concentration data. This kind of data are not more in practice, but the gas concentration filter can perfectly find out this kind of NF cases. Thus, it is effective to improve the diagnosis performance for the proposed plurality-voting SVM model.

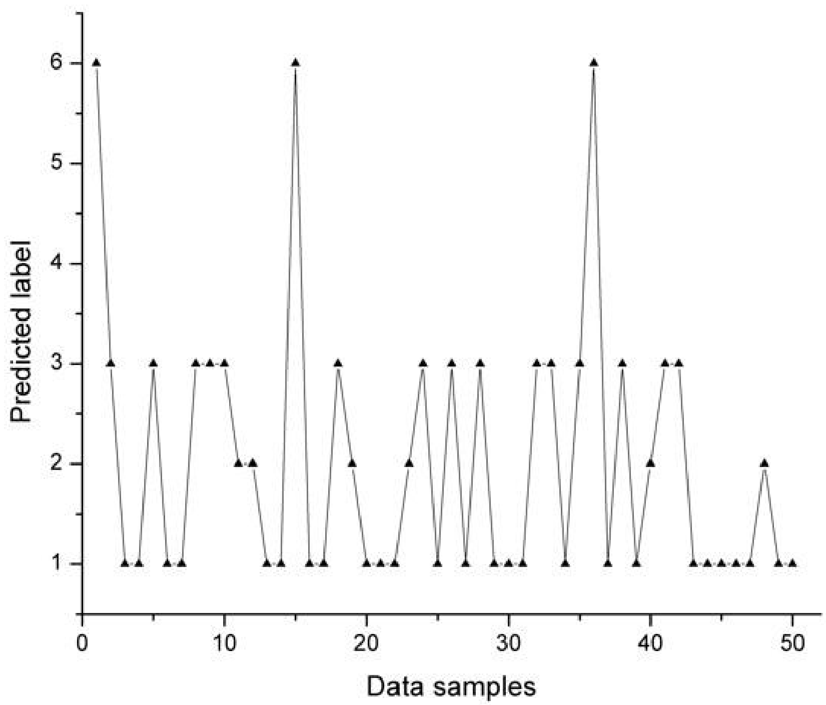

Figure 5.4 Predicted results from the SVM* model

5.2.4 Brief Summary of the Comparison of Diagnosis Accuracies

The results show that the AI-based methods are more reliable than the traditional ratio methods for fault diagnosis of power transformers. The proposed SVM-based approach, which integrates the gas concentration filter and SMOTE approach, increases the diagnosis accuracy to 80.49%. This explains the high effectiveness of the proposed SVM-based intelligent method for the fault diagnosis of power transformers.

5.3 Comparison of Diagnosis Results for Some Specific Cases

This subsection gives some specific cases to further explain the results in details. All the ratio methods and AI-based methods as explained in Section 5.1 are compared by detailed diagnosis results based on selected data.

Table 5.6 Selected gas concentration data (unit: ppm)

No.	Source	H2	CO	CH4	C2H4	C2H6	C2H2
1	Duval	0.05	3900	18900	540	410	330
2	Duval	960	15800	4000	1560	1290	6
3	Yadaiah	24.28	10000	74.59	2.67	74	0.23
4	Duval	1100	0.05	1600	2010	221	26
5	Duval	3910	1800	4290	6040	626	1230
6	Duval	92600	6400	10200	0.05	0.05	0.05
7	Duval	26788	704	18342	27	2111	0.05
8	Duval	60	780	10	4	4	4
9	Duval	6870	29	1028	900	79	5500
10	Duval	5100	117	1430	1140	0.05	1010
11	Duval	310	150	230	610	54	760
12	Duval	150	0.05	0.05	220	0.05	150
13	Duval	150	1000	0.05	200	200	150
14	Duval	80	0.05	18	0.05	20	1
15	Duraisamy	5	0.05	21	63	19	0.05

The procedure for this comparison went as follows:

1. Randomly selected the required number of datasets from the 180 samples to develop models for the ANN method, the normal SVM method, and the proposed method. The remaining samples were the testing sets.
2. Randomly selected 15 datasets from the testing sets to use in this case study.
3. Obtained the diagnosis results from the models developed in the first step.
4. Fed the 15 datasets into the ratio method to get diagnosis results.
5. Summarized and compared the results.

Table 5.7 Diagnosis results from the different methods for the selected data

No.	Actual fault	Doernenburg	Rogers	IEC	ANN	Multi-layer SVM	SVM	SVM*	Proposed method
1	T1	T1&T2	**ND**	**ND**	**NF**	**T2**	**D2**	T1	T1
2	T1	T1&T2	T1	T1	T1	T1	T1	T1	T1
3	T1	T1&T2	T1	**ND**	T1	**NF**	**NF**	T1	T1
4	T2	T1&T2	T2	T2	T2	T2	T2	T2	T2
5	T2	T1&T2	**ND**	**ND**	**D2**	T2	T2	T2	T2
6	PD	**ND**	**ND**	**ND**	PD	PD	PD	PD	PD
7	PD	**ND**	**NF**	**ND**	**NF**	**T1**	**T1**	**T1**	**T1**
8	D1	**ND**	**ND**	**ND**	D1	D1	**NF**	D1	D1
9	D1	D1&D2	D1	D1	**D2**	D1	D1	D1	D1
10	D2	D1&D2	D2	D2	**T2**	D2	D2	D2	D2
11	D2	D1&D2	D2	D2	D2	**NF**	D2	D2	D2
12	NF	**ND**	**ND**	**ND**	NF	NF	NF	**T2**	**T2**
13	NF	**ND**	**ND**	**ND**	NF	NF	NF	NF	NF
14	NF	**ND**	**ND**	**ND**	NF	NF	NF	NF	NF
15	NF	NF	**T2**	**T1**	**T1**	**T2**	**T2**	**T2**	NF

The selected datasets are shown in Table 5.6, and the corresponding results for each method are listed in Table 5.7. As shown, conventional ratio methods are not always active and able to provide results, meaning that faults were not detected (ND). Because faults are not diagnosed in these cases, we treat these cases as NF cases. Even with this assumption, the ratios methods are still not as good as the proposed method. It should be noted that Case #15 is an example that shows the importance of the NF case filter that we integrated into the proposed method. Without the gas concentration filter, the case is diagnosed incorrectly as a thermal fault. Although the Doernenburg ratio method includes the filter, it cannot always detect PD and D1 cases. The proposed method identified Cases #1, #3, #8, and #15 accurately, while the normal SVM model fails to do so, which shows the effectiveness of the proposed method.

5.4 Comparison of the Cost by Misdiagnosis

Generally speaking, if we can find the reason why the SVM misdiagnose a case, we can get a way to correct it so that the diagnosis accuracy can be improved. However, it is not the truth. Original data (low-dimensional data) normally cannot be linearly separated so that we will use a kernel function to remap the data into a high dimensional space. In the high dimensional space, most of the data can be linearly separated by a hyperplane, but some of the data still cannot be linearly separated. Because of this, we set a free factor (also called penalty factor) c to adjust the hyperplane and make the hyperplane bias to these data, and we can get the best diagnosis accuracy. However, if we adjust the c too much to make the misdiagnosed cases being diagnosed correctly, many of the correctly diagnosed cases will be misdiagnosed, which can make the diagnosis accuracy worse. Therefore, the diagnosis result we achieved is the best result, and if we adjust it for the misdiagnosed cases, the performance will be worse. The misdiagnosed results can lead to different consequences, which is worth to be studied.

Power transformers are desired to last longer since it is costly to purchase a new transformer after permanent failures caused by faults. To extend their lives, maintenance activities should be arranged. For mechanical or electrical equipment and systems, common forms of maintenance strategies can be summarized as follows (Garg and Deshmukh, 2006):

- Corrective maintenance,
- Preventative maintenance, and
- Predictive maintenance.

Corrective maintenance can be roughly divided into two categories, unplanned maintenance and planned maintenance. The unplanned maintenance is to correct the failed components/parts directly after their failures; the planned maintenance is to correct the failed components/parts periodically. This kind of maintenance is conducted based on the firm belief that the costs sustained for downtime and repair in case of a fault are lower than the investment required for a maintenance program. Therefore, corrective maintenance is not suitable for the expensive transformers. Preventative maintenance is a strategy to seek to increase the equipment's reliability and availability by reducing the probability of failures and avoiding the need for unplanned corrective maintenance (Narayan, 2004.). This approach is performed at specific time intervals, during which transformer failures can occur, so this approach is not a perfect choice for maintaining transformers.

Predictive maintenance is focusing on predicting potential failures and taking actions before failure occurs. In the maintaining of power transformers, this maintenance approach is conducted by operating staff (Sharma, 1986). Fault diagnosis approach is to assist arranging the predictive maintenance of transformers. Diagnosis results are the guidance for further transformer inspection activities. The inspection activities need to schedule the downtime of devices and require additional labor cost. In Section 5.2, the diagnosis accuracies were compared to treat every case evenly. However, diagnosis accuracy should not be the only consideration, as, in reality, wrong diagnosis results can cause different consequences (Krawczyk et al., 2014; Longadge and Dongre, 2013). Considering the costs of maintenance procedures, a good model should try to decrease downtime caused by inspection activities as much as possible. When a fault is detected, maintenance staff will schedule an inspection. An entirely wrong diagnosis result can waste more time than a different but similar diagnosis decision. For example, a case in which a thermal fault is diagnosed as a discharge fault would cost more than a case in which a low-temperature thermal fault is diagnosed as a high-temperature thermal fault, as a maintenance technician would waste more time checking non-fault-related components in the transformer.

Therefore, it is meaningful to propose another model of performance criteria rather than only considering the overall accuracy to judge the quality of a model or method.

Considering the downtime and the severity of each fault, we modeled the following cost penalty factors to evaluate each diagnosis method. The bigger the factors were, the worse the models performed. Faults can be roughly classified into thermal faults and discharge faults, and high-temperature thermal faults and the high-energy discharges are the most severe faults of each type.

Table 5.8 The cost penalty factors used in the comparison

		True result					
		T1	T2	PD	D1	D2	NF
Diagnosis result	T1	0	0.2	0.4	0.5	0.6	0.2
	T2	0.1	0	0.5	0.5	0.6	0.3
	PD	0.2	0.3	0	0.2	0.2	0.2
	D1	0.3	0.4	0.2	0	0.3	0.2
	D2	0.4	0.5	0.3	0.2	0	0.3
	NF	0.4	0.6	0.4	0.5	0.6	0

The following are some explanations of Table 5.8:

- Smaller factors represent lower costs by the diagnosed results.
- Correct diagnosis results do not increase costs, meaning they have no cost penalties.
- T2 and D2 are the most severe faults, and misdiagnosed cases have higher cost penalties, as the misdiagnosis of these conditions can lead to bad consequences.
- A fault misdiagnosed as another in the same main fault type can lead to lower cost penalties. However, a thermal fault being diagnosed as a discharge fault (or vice versa) will lead to higher cost penalties

- These numbers do not represent actual costs, and they may be adjusted slightly according to different companies' standards But they can reflect the level of losses and evaluate existing methods.

Based on the cost penalty factors above, we used Eq. (5.1) to calculate the total cost penalty factors. The number of 61 in the Eq. (5.1) represents the total number of testing data. We model these methods by randomly selected training data from the 180 total sets of data. The calculated results are shown in Table 5.9.

$$\mathrm{F} = \sum_{i=1}^{61} f_i \tag{5.1}$$

where F is the total cost penalty factors and f_i is the cost penalty factor for each case.

Table 5.9 Comparison of the total cost penalty factor

Label of methods	ANN	Multi-layer SVM	SVM	SVM*	Proposed method
# of incorrect diagnosis	22	18	19	13	12
Total cost penalty	6.6	6.1	6.9	3.9	3.6

As seen in the table, the ANN model made 22 incorrect diagnosis, which is the highest value among these methods. However, its cost penalty factor of 6.6 is lower than that of the normal SVM model's 6.9, even though the SVM model made fewer incorrect diagnosis results. It shows that the ANN method is better for dealing with rough classifications. The cost penalty factors from the ANN, multi-layer SVM, and one-layer SVM method are all over 6.0, which shows the results from these methods give much bad potential to the waste of inspection time for power transformers. By adding the

SMOTE approach, the penalty factor decreases significantly to 3.9, which can show the advantage over other methods.

In this section, the costs by misdiagnosis of AI-based methods are compared. The proposed SVM-based method can achieve the best result when calculating the cost penalty by misdiagnosis. Its cost penalty almost reduced by half compared with the ANN method, multi-layer SVM method, and normal plurality-voting SVM method.

5.5 Summary

This chapter compares the proposed methods with other selected methods, including the ratio methods, ANN method, multi-layer SVM method, and the proposed method without the SMOTE approach. In the comparison of the diagnosis accuracy, the proposed method can achieve the best diagnosis accuracy among all the methods. In the comparison of the cost by misdiagnosis, the proposed method can receive the lowest cost penalty factor, compared with other methods. This chapter validates the effectiveness of the proposed method.

Chapter 6: Conclusions and Future Work

6.1 Conclusions

Monitoring health conditions for transformers is critical for preventing failures. A reliable method for diagnosing health conditions and monitoring transformers can significantly help decrease the probability of transformer failure.

DGA is effective for continuously evaluating transformers' conditions and identifying faults inside transformers without physically opening the devices. It is used to analyze the degree of the decomposition process by detecting the gas concentrations in the transformer oil. Using the DGA approach, existing transformer fault diagnosis methods mainly include rule-based methods that are based on analyzing key gases, gas concentration ratios, or certain gas proportions. In addition, AI-based methods were proposed using DGA data, based on artificial neural network, fuzzy logic, SVM tools, etc.

This thesis summarize existing methods and challenges based on detailed literature review. In AI-based methods, SVM is a functional machine-learning approach for classification and regression problems. It can solve small sampling, nonlinear, and high-dimension practical problems better than the ANN method, which always suffers the over-fitting problem. In the development of the SVM models, gas concentration ratios are always selected as an input feature to achieve good performance. However, there are specific cases in which not much gas is generated and dissolved in new transformers' insulation oil, which can make the gas ratio unsuitable for extraction as an input feature. Gas filters should screen these specific low-concentration cases out and directly draw the final diagnosis results to NF conditions. The inspection frequency of transformers is not high, so power companies and researchers do not normally have much historical data for modeling. In addition, the numbers for data related to severe conditions are significantly higher than those of warning-level conditions, and imbalanced datasets badly influence the performance of SVM methods.

To overcome such challenges existed in the existing method, an improved SVM technique was developed for fault diagnosis of power transformers using DGA data. The proposed method integrates a gas concentration filter and a plurality-voting SVM model. The gas concentration filter can successfully process data on low gas concentrations especially collected from new transformers. The plurality-voting SVM model is designed with a plurality-voting structure and integrates the SMOTE approach and a parameter optimization approach by GA. The new structure reorganizes all binary SVM submodels and is used to avoid the problem that diagnosis results rely much on the performance of one submodel in multi-layer SVM models. SMOTE approach is employed to over sampling known data to balance training datasets, which can overcome the imbalanced datasets problem and help to achieve reliable diagnosis accuracy.

In this thesis, we compare the proposed method with other commonly used methods using the data from literature. Based on the known DGA data, the proposed approach can achieve a diagnosis accuracy of 80.49%, which is higher than that of the existing SVM, ANN, and ratio methods. To show the diagnosis result clearly, some specific cases are compared to show the effectiveness of the proposed method. In this study, we creatively consider to compare the cost-sensitivity by misdiagnosis, and the proposed method can highly reduce the costs of misdiagnosis.

In conclusion, compared with the existing method, the proposed method can not only achieve the best diagnosis accuracy, but also decrease the potential cost by misdiagnosis. Therefore, it can be used to guide the predictive maintenance activities. It is beneficial for modern cities with many electric utilities served and for industries to decrease the potential of equipment breaking off by electric failure.

6.2 Future Work

Based on the discoveries in this thesis, further studies can be conducted in the future.

- A unique kernel function that can determine the hyperplane of SVM modeling can be developed to replace common kernel functions.
- Researchers may improve the data-sampling procedures in the method modeling to get more reasonable "fake" data.

- This SVM model consists of 15 sub-models, which will require more training time than the multilayer SVM models. Therefore, the overall SVM structure could also be optimized to use fewer sub-models and less training time.
- Deep learning, currently a hot topic in the field of AI, may also be applied to the diagnosis of power transformers.
- The research direction can be extended with proper regression methods to predict future gas concentrations so that we can use the proposed model to predict and monitor future conditions in transformers.
- The problem of imbalanced datasets is not unique to transformers. For example, in the medical field, there might be less data from patients with a specific disease than for data for healthy people. The proposed method's solution for this problem may also be applied in other fields.

Bibliography

Akbani, R., Kwek, S., Japkowicz, N., 2004. Applying support vector machines to imbalanced datasets. Mach. Learn. ECML 2004 39–50.

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

Beykverdi, M., Faghihi, F., others, 2016. A new approach for transformer incipient fault diagnosis based on dissolved gas analysis (DGA). Nova J. Eng. Appl. Sci. 3.

Cao, Y.-G., Zhou, L., Ding, X.-Q., Ma, W.-J., 2006. Power transformer fault diagnosis using PNN. Relay 34, 9–11.

Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40, 16–28.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Chen, G., Fu, K., Liang, Z., Sema, T., Li, C., Tontiwachwuthikul, P., Idem, R., 2014. The genetic algorithm based back propagation neural network for MMP prediction in CO2-EOR process. Fuel 126, 202–212.

De Faria, H., Costa, J.G.S., Olivas, J.L.M., 2015. A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. Renew. Sustain. Energy Rev. 46, 201–209.

Dhote, N.K., Helonde, J.B., 2014. Improvement in transformer diagnosis by DGA using fuzzy logic. J. Electr. Eng. Technol. 9, 615–621.

Doernenburg, E., Strittmatter, W., 1974. Monitoring oil-cooled transformers by gas analysis. Brown Brveri Rev. 61, 1836–1841.

Dong, L., Xiao, D., Liang, Y., Liu, Y., 2008. 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. 78, 129–136.

Duraisamy, V., Devarajan, N., Somasundareswari, D., Vasanth, A.A.M., Sivanandam, S.N., 2007a. Neuro fuzzy schemes for fault detection in power transformer. Appl. Soft Comput. 7, 534–539.

Duraisamy, V., Devarajan, N., Somasundareswari, D., Vasanth, A.A.M., Sivanandam, S.N., 2007b. Neuro fuzzy schemes for fault detection in power transformer. Appl. Soft Comput. 7, 534–539.

Duval, M., 2002. A review of faults detectable by gas-in-oil analysis in transformers. IEEE Electr. Insul. Mag. 18, 8–17.

Duval, M., 1989. Dissolved gas analysis: It can save your transformer. IEEE Electr. Insul. Mag. 5, 22–27.

Duval, M., dePabla, A., 2001. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. IEEE Electr. Insul. Mag. 17, 31–41.

Duval, M., Dukarm, J., 2005. Improving the reliability of transformer gas-in-oil diagnosis. IEEE Electr. Insul. Mag. 21, 21–27.

ENGie, 2016. POWER TRANSFORMER | laborelec.

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

Ganyun, L.V., Haozhong, C., Haibao, Z., Lixin, D., 2005. Fault diagnosis of power transformer based on multi-layer SVM classifier. Electr. Power Syst. Res. 74, 1–7.

Garg, A., Deshmukh, S.G., 2006. Maintenance management: literature review and directions. J. Qual. Maint. Eng. 12, 205–238. doi:10.1108/13552510610685075

Gulla, B., 2014. Transformer Failure II [WWW Document]. URL http://www.fmglobal.com/riskessentials/2015/transformer-failure-ii_2015_fmglobalreason (accessed 7.13.17).

Heisele, B., Serre, T., Prentice, S., Poggio, T., 2003. Hierarchical classification and feature reduction for fast face detection with support vector machines. Pattern Recognit. 36, 2007–2017.

Henderson, K., 2016. No injuries in transformer explosion [WWW Document]. WEAR. URL http://weartv.com/news/local/no-injuries-in-transformer-explosion (accessed 7.13.17).

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

IEC, 1999. IEC 60599:1999 | IEC Webstore [WWW Document]. URL https://webstore.iec.ch/publication/2582 (accessed 6.23.17).

IEEE, 2008. IEEE SA - C57.104-2008 - IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers [WWW Document]. URL https://standards.ieee.org/findstds/standard/C57.104-2008.html (accessed 6.12.17).

IEEE, 1992. IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers. Inst. Electr. Electron. Eng. Inc NY.

Illias, H.A., Chai, X.R., Mokhlis, H., others, 2015. Transformer incipient fault prediction using combined artificial neural network and various particle swarm optimization techniques. PloS One 10, e0129363.

Illias, H.A., Chai, X.R., others, 2016. Hybrid modified evolutionary particle swarm optimization-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. Measurement 90, 94–102.

Krawczyk, B., Woźniak, M., Schaefer, G., 2014. Cost-sensitive decision tree ensembles for effective imbalanced classification. Appl. Soft Comput. 14, 554–562.

Li, J., Zhang, Q., Wang, K., Wang, J., Zhou, T., Zhang, Y., 2016. Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. IEEE Trans. Dielectr. Electr. Insul. 23, 1198–1206.

Li, Y., Tang, M.-J., Wu, F.-J., Zhang, G.-J., Wang, S.-H., others, 2013. Aging assessment of power transformer using multi-parameters. Int. J. Electr. Eng. Inform. 5, 34.

Liang, J., Wang, F., Dang, C., Qian, Y., 2014. A group incremental approach to feature selection applying rough set technique. IEEE Trans. Knowl. Data Eng. 26, 294–308.

Liao, R., Yao, C., Sun, C., He, J., 2001. Blackboard expert system of multi-expert cooperating diagnosis for transformers insulation, in: Electrical Insulating Materials, 2001.(ISEIM 2001). Proceedings of 2001 International Symposium on. IEEE, pp. 809–812.

Liao, R.J., Zheng, H.B., Grzybowski, S., Yang, L.J., 2013. A multiclass SVM-based classifier for transformer fault diagnosis using a particle swarm optimizer with time-varying acceleration coefficients. Int. Trans. Electr. Energy Syst. 23, 181–190.

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

Longadge, R., Dongre, S., 2013. Class imbalance problem in data mining review. ArXiv Prepr. ArXiv13051707.

Mehta, A.K., Sharma, R.N., Chauhan, S., Saho, S., 2013. Transformer diagnostics under dissolved gas analysis using Support Vector Machine, in: 2013 International Conference on Power, Energy and Control (ICPEC). Presented at the 2013 International Conference on Power, Energy and Control (ICPEC), pp. 181–186.

Muhamad, N.A., Phung, B.T., Blackburn, T.R., Lai, K.X., 2007. Comparative Study and Analysis of DGA Methods for Transformer Mineral Oil, in: 2007 IEEE Lausanne Power Tech. Presented at the 2007 IEEE Lausanne Power Tech, pp. 45–50.

Muthanna, K.T., Sarkar, A., Das, K., Waldner, K., 2006. Transformer insulation life assessment. IEEE Trans. Power Deliv. 21, 150–156.

Narayan, V., c2004. Effective maintenance management : risk and reliability strategies for optimizing performance /. Industrial Press, New York :

Németh, B., Laboncz, S., Kiss, I., Csépes, G., 2010. Transformer condition analyzing expert system using fuzzy neural system, in: 2010 IEEE International Symposium on Electrical Insulation. Presented at the 2010 IEEE International Symposium on Electrical Insulation, pp. 1–5.

Rogers, R.R., 1978. IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis. IEEE Trans. Electr. Insul. 349–354.

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

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

Sarma, D.V.S.S.S., Kalyani, G.N.S., 2004a. ANN approach for condition monitoring of power transformers using DGA, in: 2004 IEEE Region 10 Conference TENCON 2004. Presented at the 2004 IEEE Region 10 Conference TENCON 2004., p. 444–447 Vol. 3.

Scholkopf, B., Smola, A.J., 2001. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

Seifeddine, S., Khmais, B., Abdelkader, C., 2012. Power transformer fault diagnosis based on dissolved gas analysis by artificial neural network, in: 2012 First International Conference on Renewable Energies and Vehicular Technology. Presented at the 2012 First International Conference on Renewable Energies and Vehicular Technology, pp. 230–236.

Sharma, S.C., 1986. Reliability and maintenance of power systems substations /.

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

Singh, S., Joshi, D., 2015. Computer applications in fault diagnosis of power transformers - a review, in: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom). Presented at the 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1216–1223.

Souahlia, S., Bacha, K., Chaari, A., 2012. 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. 43, 1346–1353.

Styvaktakis, E., Bollen, M.H.J., Gu, I.Y.H., 2002. Expert system for classification and analysis of power system events. IEEE Trans. Power Deliv. 17, 423–428.

Su, C.Q., 2016. A new fuzzy logic method for transformer incipient fault diagnosis, in: Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on. IEEE, pp. 324–327.

Su, Q., Lai, L.L., Austin, P., 2000a. A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer.

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

Sun, H.-C., Huang, Y.-C., Huang, C.-M., 2012. A Review of Dissolved Gas Analysis in Power Transformers. Energy Procedia, 2011 2nd International Conference on Advances in Energy Engineering (ICAEE) 14, 1220–1225.

Sun, Y., Zhang, S., Miao, C., Li, J., 2007. Improved BP neural network for transformer fault diagnosis. J. China Univ. Min. Technol. 17, 138–142.

Tewari, P.C., Khanduja, R., Gupta, M., 2012. Performance enhancement for crystallization unit of a sugar plant using genetic algorithm technique. J. Ind. Eng. Int. 8, 1.

Tian, Z., Zuo, M.J., 2010. Health condition prediction of gears using a recurrent neural network approach. IEEE Trans. Reliab. 59, 700–705.

Vapnik, V., 2013a. The nature of statistical learning theory. Springer science & business media.

Vapnik, V., 2013b. The Nature of Statistical Learning Theory. Springer Science & Business Media.

Wallingford, K., 2017. Transformer failure causes Romanian power plant outage | Power Transformer News.

Wang, Z., Liu, Y., Griffin, P.J., 2000a. Neural net and expert system diagnose transformer faults. IEEE Comput. Appl. Power 13, 50–55.

Wang, Z., Liu, Y., Griffin, P.J., 2000b. Neural net and expert system diagnose transformer faults. IEEE Comput. Appl. Power 13, 50–55.

Wu, G., Chang, E.Y., 2003. Class-boundary alignment for imbalanced dataset learning, in: ICML 2003 Workshop on Learning from Imbalanced Data Sets II, Washington, DC. pp. 49–56.

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

Yin, Y.J., Zhan, J.P., Guo, C.X., Wu, Q.H., Zhang, J.M., 2011. Multi-kernel support vector classifier for fault diagnosis of transformers, in: Power and Energy Society General Meeting, 2011 IEEE. IEEE, pp. 1–7.

Zakaria, F., Johari, D., Musirin, I., 2012. Artificial neural network (ANN) application in dissolved gas analysis (DGA) methods for the detection of incipient faults in oil-filled power transformer, in: 2012 IEEE International Conference on Control System, Computing and Engineering. Presented at the 2012 IEEE International Conference on Control System, Computing and Engineering, pp. 328–332.

Zhang, Y., Wang, S., Phillips, P., Ji, G., 2014. Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. Knowl.-Based Syst. 64, 22–31.

Zheng, H.B., Liao, R.J., Grzybowski, S., Yang, L.J., 2011. Fault diagnosis of power transformers using multi-class least square support vector machines classifiers with particle swarm optimisation. IET Electr. Power Appl. 5, 691–696.

Appendices

Appendix A The diagnosis results by ANN method and multi-layer SVM method

This section gives the original results obtained from MATLAB code of the ANN method and multi-layer SVM method. Each figure compares the results from these two methods and the real conditions.

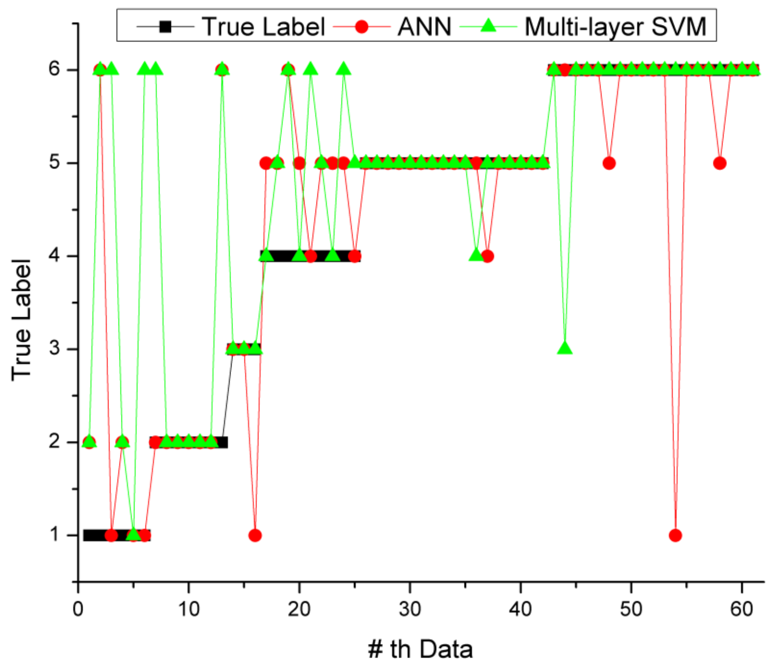

Figure Appendix.1 Comparison of the results from ANN and multi-layer SVM using round #1 data

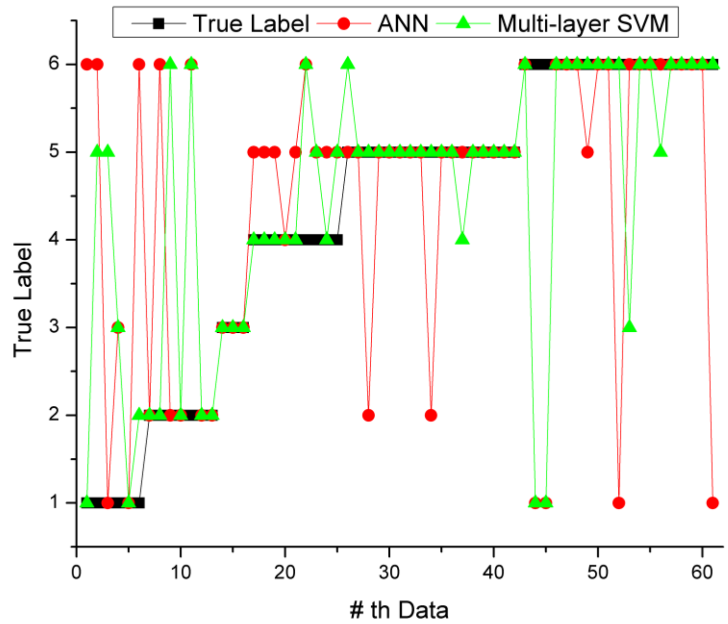

Figure Appendix.2 Comparison of the results from ANN and multi-layer SVM using

round #2 data

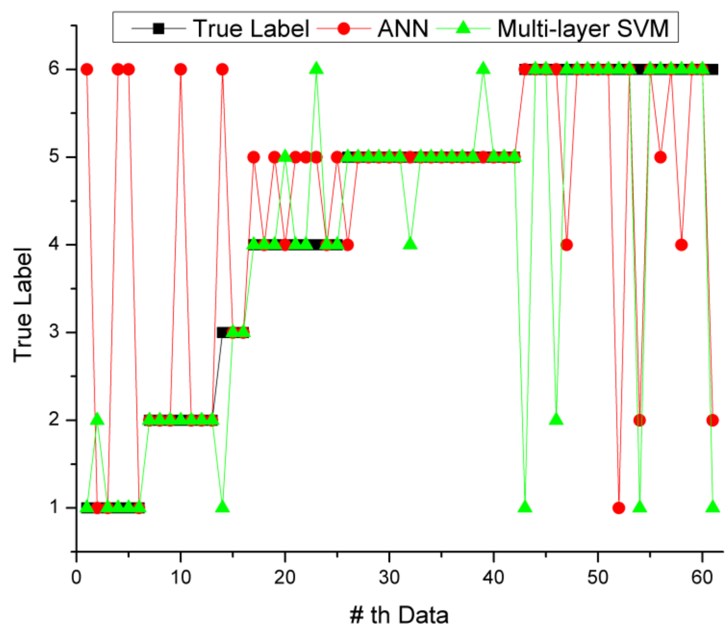

Figure Appendix.3 Comparison of the results from ANN and multi-layer SVM using

round #3 data

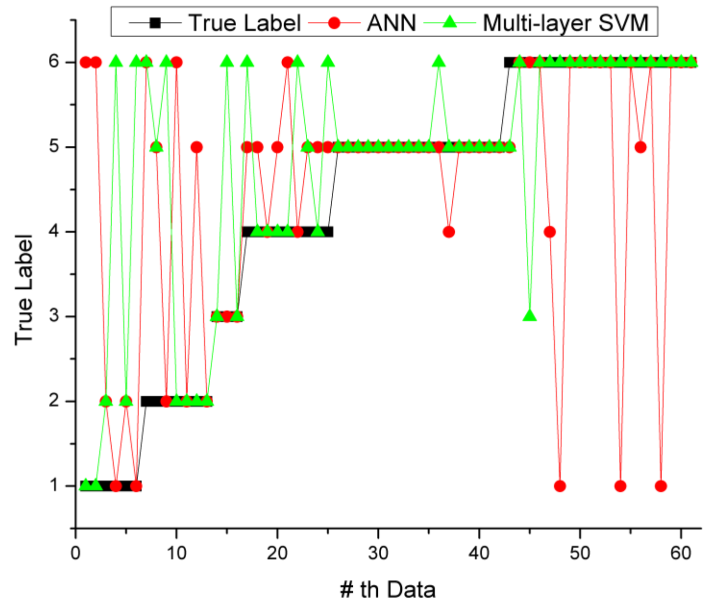

Figure Appendix.4 Comparison of the results from ANN and multi-layer SVM using round #4 data

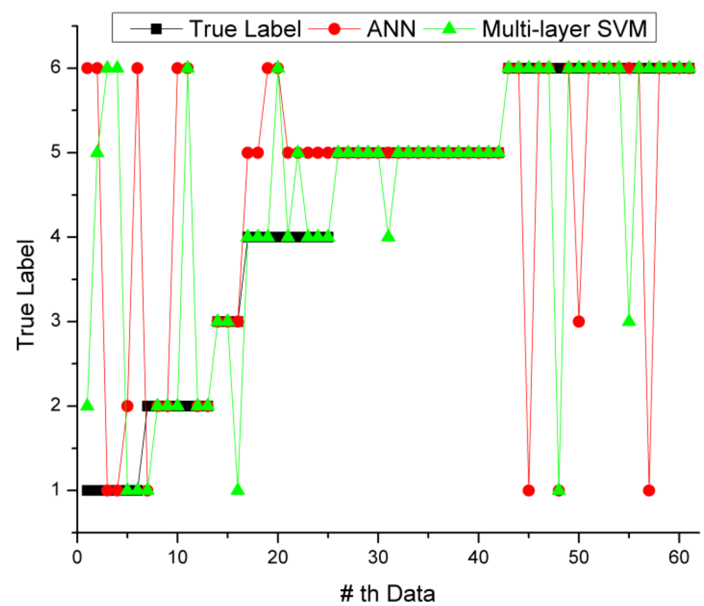

Figure Appendix.5 Comparison of the results from ANN and multi-layer SVM using round #5 data

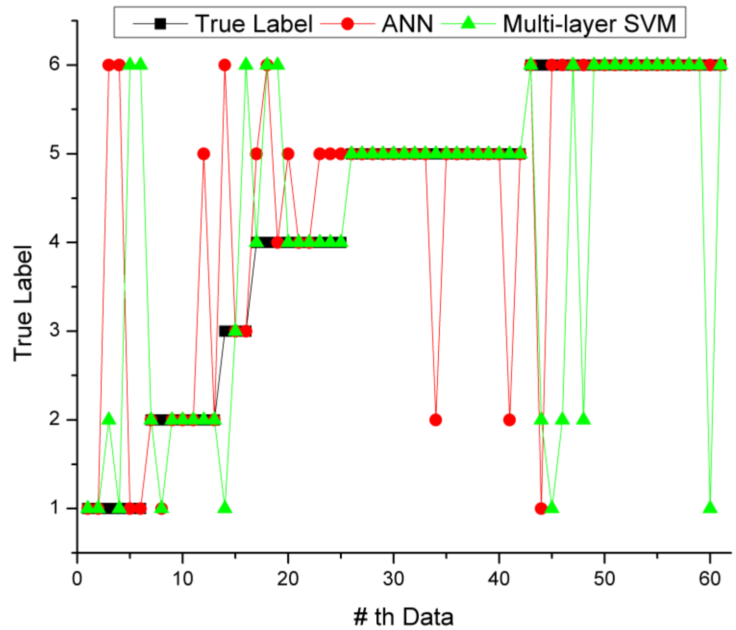

Figure Appendix.6 Comparison of the results from ANN and multi-layer SVM using round #6 data

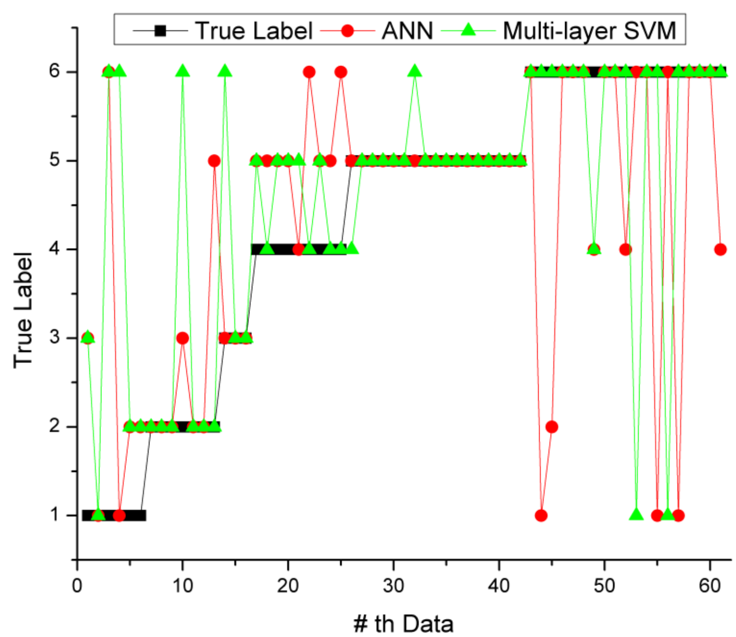

Figure Appendix.7 Comparison of the results from ANN and multi-layer SVM using round #7 data

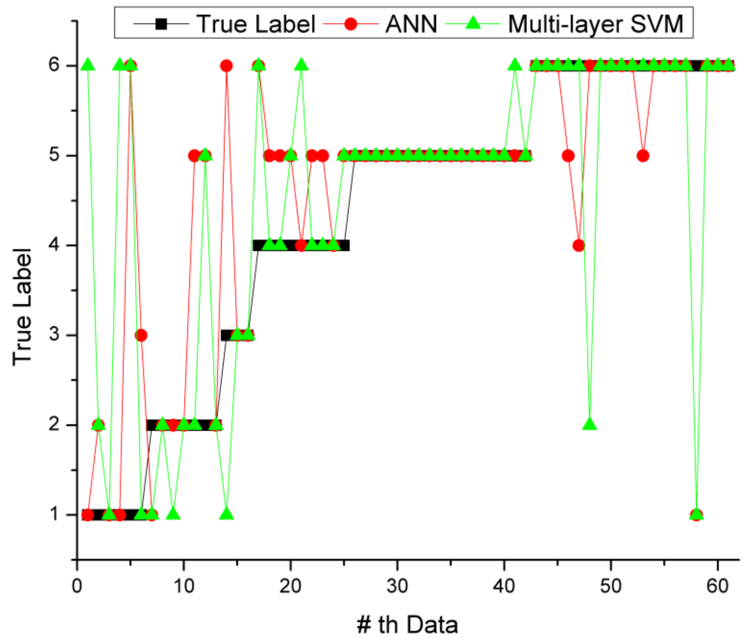

Figure Appendix.8 Comparison of the results from ANN and multi-layer SVM using round #8 data

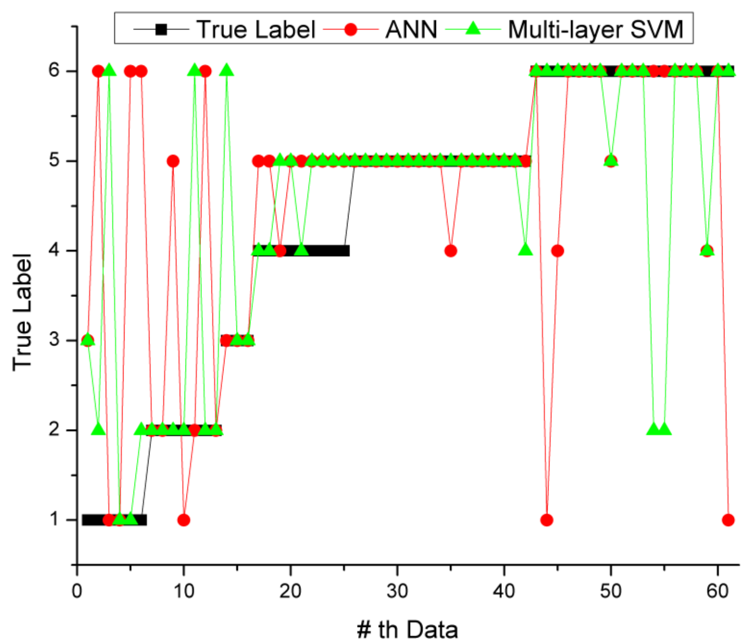

Figure Appendix.9 Comparison of the results from ANN and multi-layer SVM using round #9 data

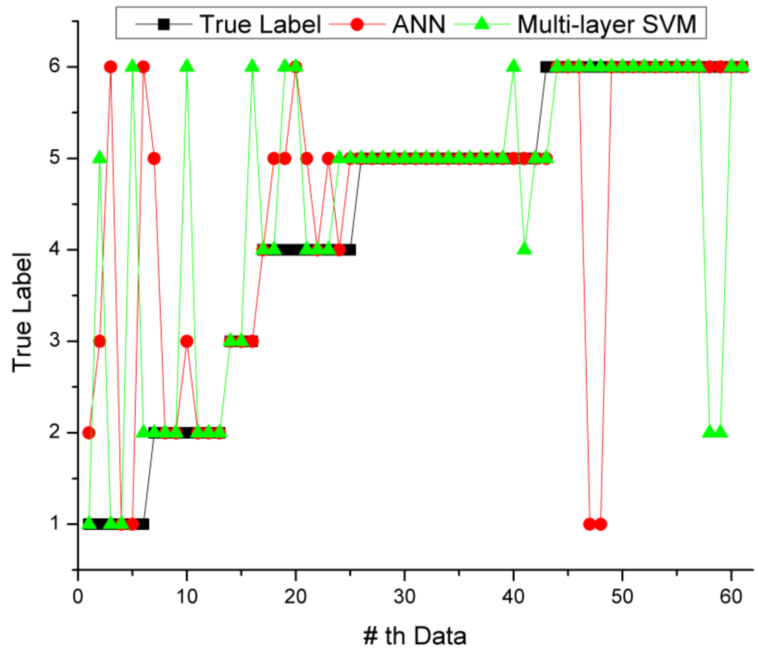

Figure Appendix.10 Comparison of the results from ANN and multi-layer SVM using round #10 data

Appendix B The diagnosis results by the SVM, SVM* and proposed method

This section gives the original results obtained from the MATLAB code of the SVM, SVM* and the proposed method. Each figure compares the results from these three methods and the real conditions.

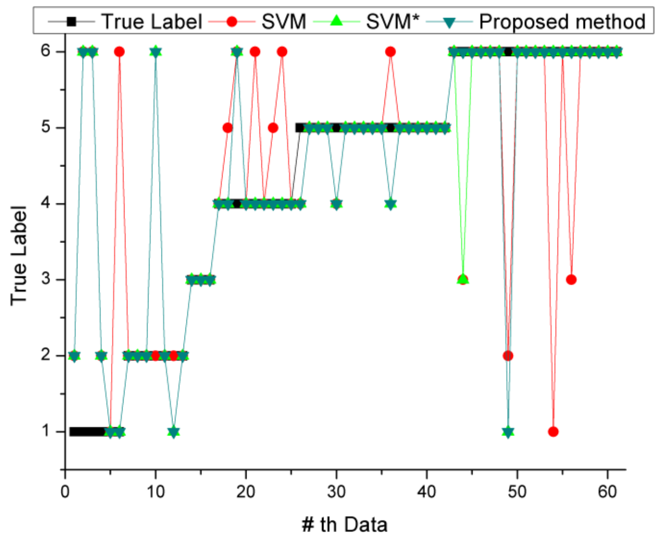

Figure Appendix.11 Comparison of the results from the SVM, SVM* and proposed method using round #1 data

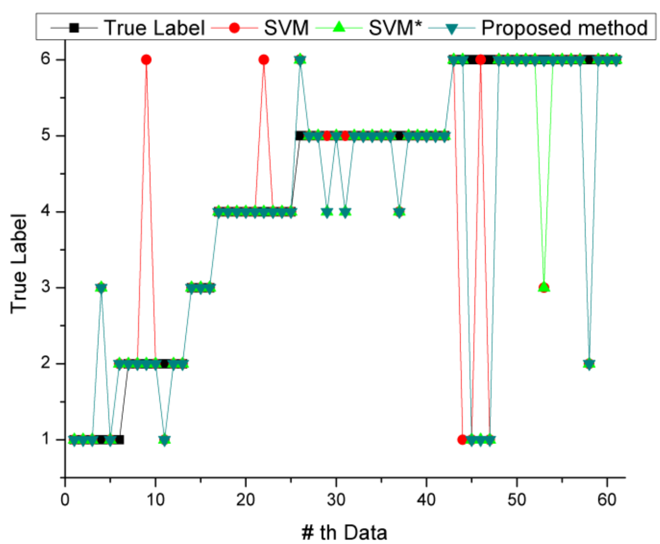

Figure Appendix.12 Comparison of the results from the SVM, SVM* and proposed method using round #2 data

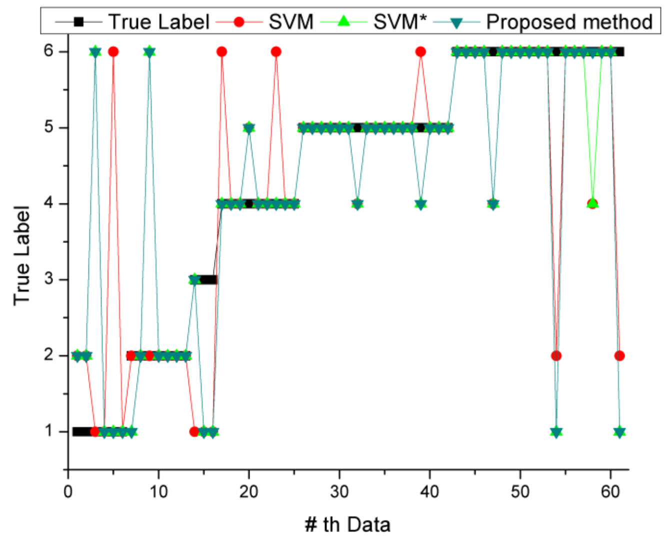

Figure Appendix.13 Comparison of the results from the SVM, SVM* and proposed method using round #3 data

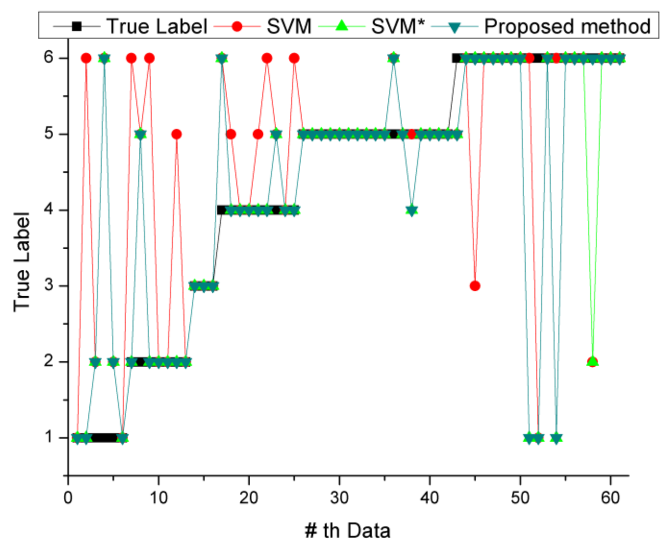

Figure Appendix.14 Comparison of the results from the SVM, SVM* and proposed method using round #4 data

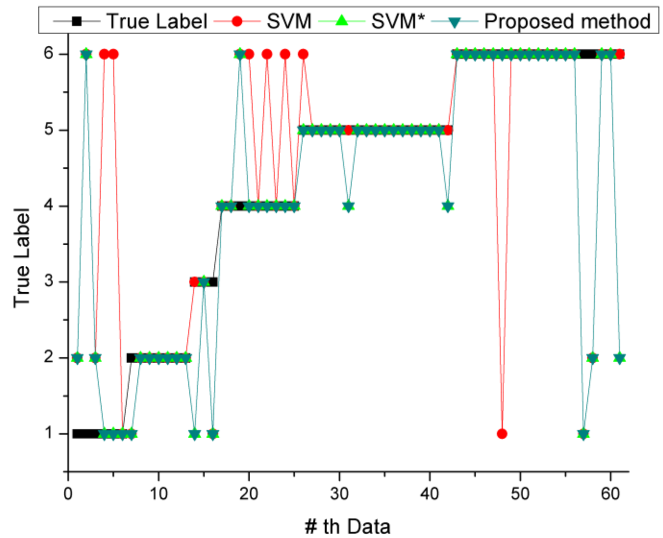

Figure Appendix.15 Comparison of the results from the SVM, SVM* and proposed method using round #5 data

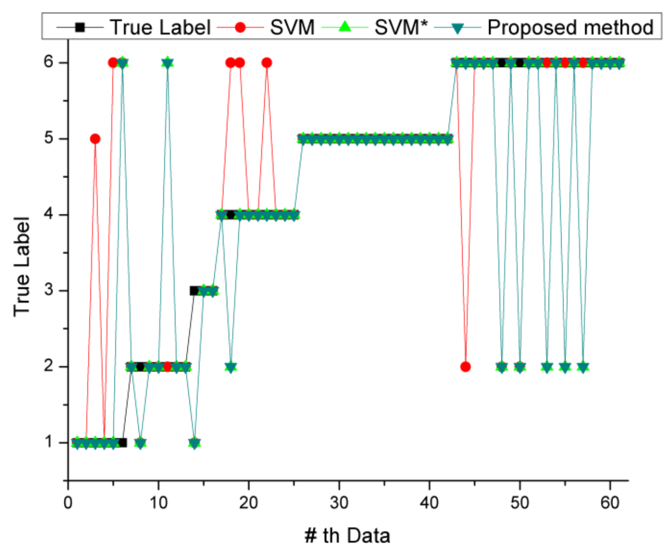

Figure Appendix.16 Comparison of the results from the SVM, SVM* and proposed method using round #6 data

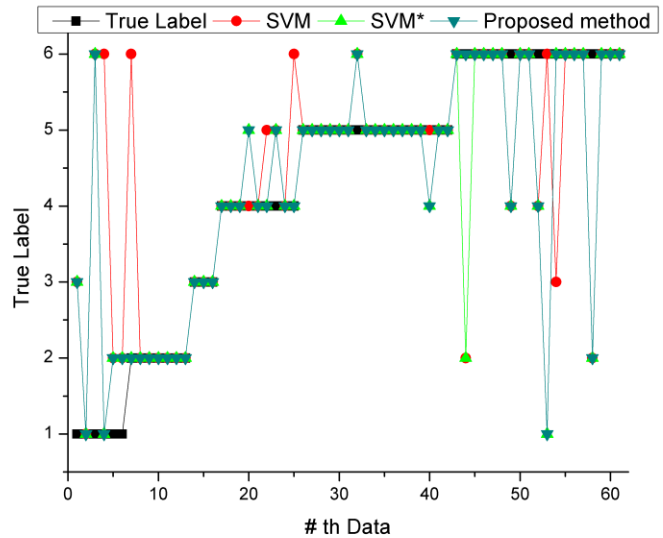

Figure Appendix.17 Comparison of the results from the SVM, SVM* and proposed method using round #7 data

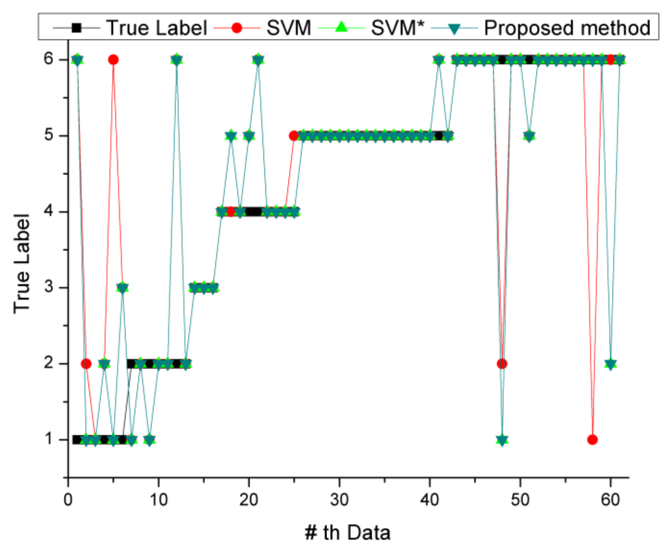

Figure Appendix.18 Comparison of the results from the SVM, SVM* and proposed method using round #8 data

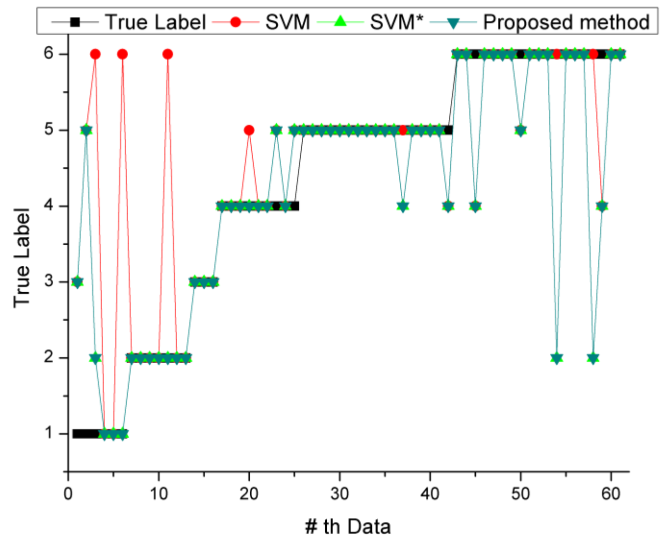

Figure Appendix.19 Comparison of the results from the SVM, SVM* and proposed method using round #9 data

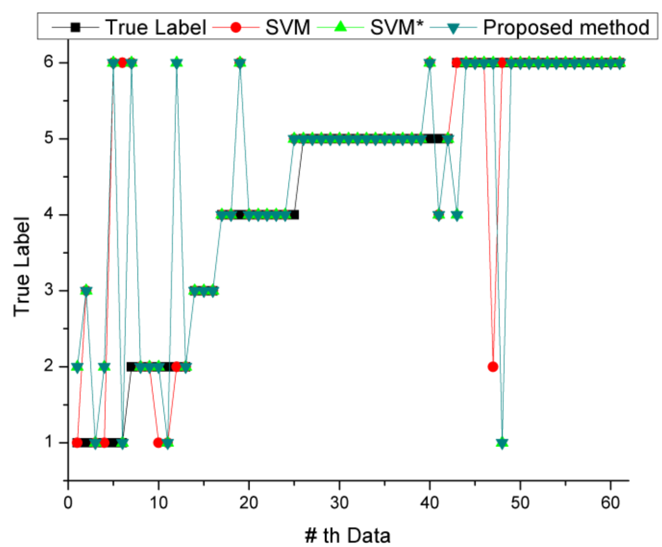

Figure Appendix.20 Comparison of the results from the SVM, SVM* and proposed method using round #10 data