공학박사학위논문

오염된 유중가스분석 데이터에 대한 딥러닝 기반 유입식 변압기 상태예측 연구

Deep Learning Based Health Prognostics of Oil-immersed Transformers for Contaminated Dissolved Gas Analysis Data

2022 년 8 월

서울대학교 대학원

기계항공공학부

서 보 성

오염된 유중가스분석 데이터에 대한 딥러닝 기반 유입식 변압기 상태예측 연구

Deep Learning Based Health Prognostics of Oil-immersed Transformers for Contaminated Dissolved Gas Analysis Data

지도교수 윤 병 동

이 논문을 공학박사 학위논문으로 제출함

2022 년 04 월

서울대학교 대학원

기계항공공학부

서 보 성

서보성의 공학박사 학위논문을 인준함

2022 년 06 월

위 원 장 : 김 윤 영 (인)

부위원장 : 윤 병 동 (인)

위 원 : 김 도 년 (인)

위 원 : 이 호 원 (인)

위 원 : 김 태 진 (인)

Abstract

Deep Learning Based Health Prognostics of Oil-immersed Transformers for Contaminated Dissolved Gas Analysis Data

Boseong Seo
Department of Mechanical and Aerospace Engineering
The Graduate School
Seoul National University

With the acceleration of the energy market, such as smart grids, energy storage systems, and electric vehicles, demand for reliable electrical power systems for safe and continuous power supply is increasing. To meet this, many studies on diagnostic techniques and preventive maintenance for core facilities of transmission and distribution systems have been conducted. Among them, the oil-immersed transformer plays a pivotal role in the electrical power system as a device that changes the voltage according to the user's purpose. Therefore, various tests have been developed for the diagnosis of power transformer, and the dissolved gas analysis (DGA) is the most representative method. DGA is a method of measuring the gas concentrations generated when the internal insulation is decomposed due to

a defect in the transformer. Various international organizations such as IEEE and IEC have established DGA-based transformer diagnostic standards through decades of research and industrial experience. However, this method has a high misdiagnosis rate because it is based on the experience and interpretation of experts. Therefore, this study attempted to develop the superior predictive diagnosis of the transformer based on date-driven approach by using a large amount of DGA data acquired at an actual industrial site.

To improve the diagnosis performance of the transformer, there are three main issues to be addressed: 1) missing data issue in DGA, 2) health feature extraction issue for low-dimensional data, and 3) health prognosis issue for irregular sampling intervals. In order to solve these issues, this doctoral dissertation proposes the following three studies:

The first study proposes iterative denoising autoencoder (IDAE) for multiple missing value imputation. The proposed method can restore the original value of the missing value by iteratively performing denoising autoencoder (DAE). DAE which minimizes the noise estimates the original value of the missing value by making the missing value recognized as noise. The proposed method enables more accurate transformer diagnosis by increasing the reliability of DGA.

The second study proposes a method of extracting health features through semi-supervised autoencoder (SSAE). The proposed method can extract two characteristic features with monotonous degradation behavior by simultaneously performing dimension reduction and health status learning of transformers. Since the correlation between gas concentrations is modeled by learning a vast amount of industrial data,

the performance is more accurate than conventional methods. In addition, the degradation trend can be intuitively understood by visualizing the health feature space consisting of two health features.

Finally, the third study proposes a health prognosis of transformers through the XGBoost regression method. The proposed method can obtain a robust prognosis model on the irregular sampling intervals by learning the irregular time series data using tree-based ensemble learning methods. Since the proposed method learns to minimize errors of the sequential models, it can prevent overfitting and accurately predict the status. It is expected to be of great help in preventive maintenance of transformers because it assures excellent performance for up to 5 years.

It can be used to build a health prognosis framework for transformers by performing the three proposed methods in a continuous process. In addition, it is significant in that it has developed a universal model that can be directly applied to the industry by using a vast amount of data acquired from the real industry.

Keywords: Health prognostics
Data imputation
Oil-immersed transformer
Deep learning
Machine learning
Dissolved gas analysis

Student Number: 2013-20680

Table of Contents

List of Tables

List of Figures

Nomenclatures

IEEE	Institute of electrical and electronics engineers
IEC	International electrotechnical commission
DGA	Dissolved gas analysis
LSTM	Long short-term memory
PdM	Predictive maintenance
RUL	Remaining useful life
GRNN	General regression neural network
ANFIS	Adaptive network-based fuzzy interface system
IDAE	Iterative denoising autoencoder
SSAE	Semi-supervised autoencoder
DAE	Denoising autoencoder
XGBoost	Extreme gradient boosting
ML	Machine learning
FMEA	Failure modes and effects analysis
PD	Partial discharge
TDCG	Total dissolved combustible gas
KNN	K-nearest neighborhood
ELU	Exponential linear unit
KEPCO	Korea electric power corporation
XGB	XGBoost
GBM	Gradient boosting machines
LGB	Light GBM
RF	Random forest
NRMSE	Normalized root mean square error

TP	True positive
TN	True negative
FP	False positive
FN	False negative
HFS	Health feature space
AE	Autoencoder
SC	Softmax classifier
ReLU	Modified linear unit
MSE	Mean square error
CE	Cross-entropy
TCG	Total combustible gas
PPV	Positive predictive rate
TPR	True positive rate
TNR	True negative rate
OP	Orthogonal projection
SVM	Support vector machine
HI	Health index
DNN	Deep neural network
f_θ	Encoder function
$\mathbf{W}$	Weight matrix of autoencoder
$\mathbf{b}$	Bias vector of autoencoder
$\mathbf{h}$	Latent value
s	Activation function
$g_{\theta'}$	Decoder function
$\boldsymbol{\theta}$	Weight parameters of encoder layers
$\boldsymbol{\theta}'$	Weight parameters of decoder layers

L	Loss function
$\hat{\mathbf{x}}$	Estimated output
$q_D(\tilde{\mathbf{x}}\|\mathbf{x})$	Stochastic corruption procedure
$\tilde{\mathbf{x}}$	Noisy input data
fc	Fully connected layer
d	Euclidean distance
Δ	Difference between before and after IDAE update
r	Correlation coefficient
$\bar{\mathbf{x}}$	Mean value
η	Learning rate
σ	Softmax function
$\mathbf{W}_{SC}^{\mathrm{T}}$	Weight matrix of softmax classifier
$\mathbf{b}_{\mathrm{SC}}^{i}$	Bias vector of softmax classifier
L_{CE}	Loss function with cross-entropy
L_{SSAE}	Loss function of semi-supervised autoencoder
L_{AE}	Loss function of autoencoder
L_{SC}	Loss function of autoencoder
$\boldsymbol{\theta}^{\mathrm{en}}$	Weight parameters of encoder layers
$\boldsymbol{\theta}^{\mathrm{de}}$	Weight parameters of decoder layers
$\boldsymbol{\theta}^{\mathrm{sc}}$	Weight parameters of softmax layers
ϕ	XGBoost model
$\mathcal{F}$	Space of regression trees
w	Leaf weight of tree
q	Mapping function to leaf index
T	The number of leaves
l	Loss function of gradient boosting model

Ω	Regularization term
$\gamma,\ \lambda$	Regularization parameters of Ω
$l',\ l''$	First and second order gradient of the loss function of gradient boosting model
N_{layers}	The number of layers

Chapter 1

Introduction

1.1 Motivation

As market opportunities related to smart grids and sustainable electric networks have grown, concerns about the stability of the electric power system have increased. Transformers, one of the main components of the electric system, need careful management. However, as transformers are used over several decades, they degrade and become subject to abrupt accidents that arise for various reasons, such as abnormal voltage, careless operation, and insulation degradation. To avoid transformer failure, global standard organizations, such as IEEE and IEC, suggest maintenance guidelines based on the domain knowledge.

Much research has been conducted towards the goal of accurately diagnosing and maintaining transformers. Among prior methods, dissolved gas analysis (DGA) is the most widely used. When the insulator of a transformer is dismantled by thermal and electrical stresses, combustible gases are generated and dissolved into the oil that is filled inside transformers. The amount and the ratio of these dissolved gases depend on the degradation condition of the transformer. Therefore, the condition of the transformer can be estimated through gas analysis. There are several global

standards on dissolved gas analysis, such as IEEE Std C57.104 [1], IEC 60599 [2], Duval Triangle [3], Doernenburg Ratios [4], Rogers Ratios [5], and Basic Gas Ratios [2]. These methods have been improved based on decades of study; however, they still have low accuracy and a high frequency of false alarms.

As transformer data has accumulated over long periods, and techniques based on big data analysis and artificial intelligence are maturing, much research has been conducted with the aim of inferring the status of a transformer using DGA data. In the early stages of these methods, algorithms using machine learning, including fuzzy logic [6-8], artificial neural networks [9-13], and support vector machine [14-19], are often applied. Recently, state-of-the-art technologies that incorporate deep learning have been adopted to construct diagnosis models for transformers [20-23]. For example, L. Luo et al. [22] proposed a DGA online fault diagnosis method that combines a convolutional neural network and a bidirectional Long Short-Term Memory (LSTM) network. In other work, D. Yang et al. [23] developed a double-stacked autoencoder for fast and accurate judgement of transformer health condition.

Although artificial intelligence methods can greatly enhance the accuracy of diagnosis, and verify various failures of a transformer, they require a large amount of and high-quality data to construct a robust model. Unfortunately, in the case of transformers, obtaining such data is difficult. Transformer data is rarely measured, usually only once or twice a year. Therefore, it is fundamentally difficult to obtain sufficient data to be used to train the diagnostic model. And as the amount of data is small, the imbalance between normal and fault data increases, and even the collected data has a limitation that some of them are unlabeled.

Further, the data may be corrupted for a variety of reasons. The quality of DGA data greatly depends both on the techniques used to extract gas from the insulating oil and the skill of the personnel gathering the data. According to research by Cho et al. [24], the error rate of the round-robin test for each laboratory is 15-30%, on average. In addition, Dukarm reported common issues arising from DGA data, such as low measurement precision, contradictory data, and intermittent gas losses, all of which require special attention when measuring DGA gases [25].

Recently, the world has focused on predictive maintenance (PdM) to prevent safety accidents of facilities and increase the maintenance efficiency. And research and demonstration projects related to remaining useful life (RUL) prediction are already being conducted for many industrial facilities [26-30]. On the other hand, studies of PdM for transformers were relatively slow, and only two studies were found [31, 32]. Study [31] predicted the RUL of transformers from DGA data using General Regression Neural Network (GRNN)-based ensemble learning, and study [32] predicted the concentrations of seven gas types and future status of transformer using Adaptive Network-based Fuzzy Inference System (ANFIS) and rule-based fuzzy logic. However, these studies have limitations in performance in the field due to limited data and low accuracy.

Therefore, in this dissertation, the new deep learning-based framework of fault diagnosis and prognosis for power transformers is proposed to overcome the limitations mentioned above. This dissertation can achieve the following three things:

1) Data reliability and robustness are improved by restoring the original value of contaminated DGA data.

2) By deriving the health features representing the status of transformer, it is possible to grasp the deterioration trend of transformer which changes monotonically.

3) It is possible to predict the status of transformer and the time at which defects occur from the deterioration trend of transformer.

1.2 Research Scope and Overview

The goal of this doctoral dissertation research is to develop the prognostic methods that predicts the status of oil-immersed transformers using deep-learning technologies for incomplete dissolved gas analysis data. The research is composed of three thrusts. First thrust is a data imputation method to recover the multiple missing DGA value using iterative denoising autoencoder (IDAE). Second thrust is a extraction of health features which has a monotonically decreasing degradation pattern. Health features are derived from semi-supervised autoencoder (SSAE). Last thrust is a health prognosis of power transformers based on XGBoost regression. These three thrusts are briefly described below.

Research Thrust 1: Iterative denoising autoencoder (IDAE) for missing data imputation of DGA

In research thrust 1, we propose iterative denoising autoencoder (IDAE), an imputation method to restore the multiple missing values in offline DGA. During the process of extracting and transporting insulating oil, some gas concentrations are

missing as the gas volatilizes into the air. This can lead to a fatal error in not detecting defects of the transformer. Therefore, the industry is presenting various guidelines to pay special attention to DGA, but it is not actually well followed.

Therefore, in this thrust, we seek to restore the original value of DGA data through the data driven approach. This approach is based on autoencoder and designed to overcome the limitations of existing methods. The proposed method consists of three steps: 1) defining the inputs and data normalization, 2) DAE model learning for a single missing value, and 3) IDAE for multiple missing values. The main idea of this study is to enable DAE imputation model for estimation of the original data to replace the incomplete data with multiple missing values by repeating the imputation process until the missing values converge.

The proposed method is verified in this research through three comparative studies that examine field data provided by an electric power corporation. Specific studies provide: 1) a comparison with conventional methods on imputation performance for a single gas, 2) examination of imputation performance between multiple missing values, and 3) documentation of diagnosis accuracy before and after imputation. The results of the case studies show that the proposed method is effective for imputation of the missing DGA data. IDAE can help diagnose the health status of transformers accurately by estimating the missing values of DGA data.

Research Thrust 2: Semi-supervised autoencoder (SSAE) for health feature extraction of power transformers

In research thrust 2, we propose semi-supervised autoencoder (SSAE) based health diagnosis to evaluate the condition of transformer and to identify the degradation trend. Conventional DGA-based diagnostic methods determine the status of transformer by dividing it into three or four grades, so there is a limit to further subdividing and grasping the status of transformer. Therefore, in the real industry, there is a demand to quantitatively judge the status by quantifying it.

Therefore, in this thrust, we try to extract the health features which visually express the degradation condition of transformer. To realize this, it went through a two-step process: 1) defining the inputs and data normalization, and 2) SSAE model learning for extracting health features. Through the SSAE model, monotonically decreasing health features can be extracted.

We evaluate the diagnostic accuracy of proposed model compared to that of conventional methods such as IEEE and IEC. The performance of proposed method is more accurate than others. In particular, the proposed method was able to significantly reduce the false alarm rate because the composition ratio of the total gas was also used as a diagnosis factor.

Research Thrust 3: XGBoost regression for health prognostics of power transformers

In research thrust 3, we propose XGBoost regression model to predict the status or remaining useful life (RUL) of power transformers. The biggest challenge in predicting the transformer status is that the data acquisition interval is not constant.

In the case of DGA data, the measurement period, from 1 month to 2 years, is different depending on the transformer condition according to international guidelines. Most of the methods commonly used for prediction problems are difficult to apply to this case because the data sampling rates must be constant.

Therefore, in this thrust, we try to develop an health prognosis model of power transformer applicable for data with irregular sampling rates. To realize this, it went through a two-step process: 1) orthogonal projection of health features for obtaining health index, and 2) XGBoost regression model learning for health prognosis. Health features can be converted into a one-dimensional health index by projecting them to the orthogonal plane. And by using the sampling interval as an input for the XGBoost learning model, it is designed to obtain the prediction results robust to the measurement cycle.

To evaluate the prognostic performance of the proposed model, two case studies were conducted: 1) a comparison with other machine learning (ML) algorithms on prognosis accuracy and 2) prognosis accuracy degradation due to long prediction period. As a result of studies, XGBoost model has the highest accuracy among the ML models. And even if the prediction period is extended, it maintains an average accuracy of 80%, and has the minor decline.

1.3 Dissertation Layout

The remaining chapters of this dissertation is organized as follows. Chapter 2 provides a literature review including the fault modes of power transformers,

backgrounds of dissolved gas analysis for basic understanding, and conventional fault diagnosis of power transformers. Chapter 3 describes the missing data imputation via iterative denoising autoencoder (IDAE). Chapter 4 suggests the health feature extraction via semi-supervised autoencoder (SSAE). Chapter 5 proposes the health prognosis model of power transformers via XGBoost regression. At the conclusion, Chapter 6 summarizes the result of this research and suggestions for the future research.

Chapter 2

Literature review

In this chapter, to help reader's understanding, we will cover the fault modes of the power transformer and the methods for diagnosing it in general. Subchapter 2.1 provides the failure modes and effects analysis (FMEA) for power transformers. FMEA is process of identifying potential failure modes of components in a system and their causes and effect. Subchapter 2.2 describes the definition and process of dissolved gas analysis (DGA). Lastly, Subchapter 2.3 summarizes the conventional rule-based fault diagnostic methods of power transformers.

2.1 Failure Modes and Effects Analysis (FMEA) for Power Transformers

Transformers are one of the most important components in the electric power system. Transformers connect power systems with different voltage levels. As a transformer experiences enormous electrical and mechanical stresses as it operates under high-voltage conditions, internal components (e.g., insulating paper and windings) undergo degradation. This degradation ultimately results in the abrupt failure of the transformer, which interrupts the power grid that is connected to the

transformer.

A transformer can fail by electrical, thermal, and/or mechanical defects. Electrical defects result from transient over-voltage or winding resonance. Thermal defects may arise from an overload current, local overheating, leakage fluxes and/or failure of the cooling system. Mechanical stress between the conducting material and the winding occurs because of a short circuit of the winding and inrush current.

Such defects in a transformer are caused by a decrease in the mechanical and dielectric strength of the transformer as the internal insulator ages. In general, aged conductor insulation is weakened to the point where it can't withstand the mechanical stresses of a fault. Indeed, the insulation becomes so brittle that even normal operating conditions may cause severe damage. Then, dielectric failure of the turn-to-turn insulation or loosening of the winding clamping pressure occur, which reduces the transformer's resistance to short-circuit forces [33].

According to the case study about failure statistics for large-capacity transformers that are operating for about 1 year and 6 months to 2 years and 6 months, the main failure modes of transformers are shown in Figure 2-1 [34]. As you can see in Figure 2-1, most fault modes are due to insulation defects. Typical major insulation defects occurring in the transformer are as follows: moisture in the cellulose insulation, contamination of oil with water or particles, insulation surface contamination (which occurs mainly due to the adsorption of polar aging products on a cellulose surface or due to deposition of conducting particles and insoluble aging products), and partial discharges in weaker parts of the insulation. Table 2-1 summarizes general insulation faults of transformers based on FMEA [35].

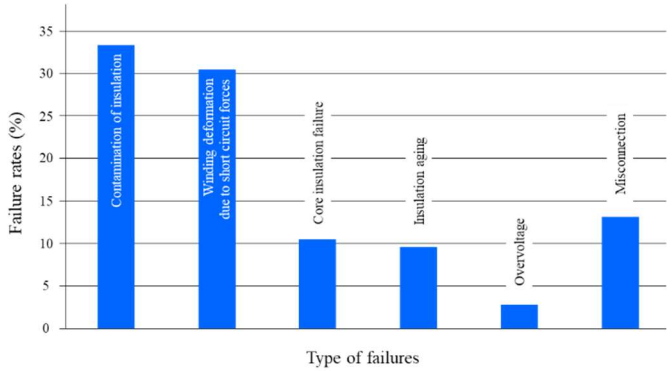


Figure 2-1 Failure rates by major failure modes

Table 2-1 General insulation faults in transformer through FMEA

Root causes of faults	Fault modes	Fault effects
Critical contamination (H_2O) of oil & Rapid change of temperature	PD appearance at rated voltage	Breakdown
Surface contamination & Rapid change of temperature	PD appearance	Flashover
Particle's contamination & Switching surge	Critical PD	Breakdown
Water & Particles contamination	Critical PD (Creeping discharge)	Breakdown
Surface contamination & Lightning impulse	Surface discharge	Flashover
Distortion of winding geometry	PD appearance (Creeping discharge)	Breakdown
Distortion of winding geometry & Switching surge	Flashover between coils	Gas evolution

2.2 Dissolved Gas Analysis (DGA) in Power Transformers

As previously described, there are various failure modes of transformers; thus, developing a diagnosis model for each failure would be costly and ineffective. However, most failures are the reason of the generation of dissolved gas in the oil filled inside a transformer. Thus, by analyzing the dissolved gas, the overall status of the transformer can be monitored, and the failure can be predicted in advance and thus prevented. The examination of dissolved gas is commonly called dissolved gas analysis (DGA).

A high-voltage transformer is usually insulated with insulation paper and oil to safely handle the high voltage. Two levels of insulation are used. First insulation paper is wrapped around the coils and iron core of the transformer. Next, the inside of the transformer where the coils and iron core are located is filled with oil for further insulation. When there is local heating, arc, or thermal or electrical stress, the insulation material is decomposed and combustible gases (e.g., H_2, C_2H_2, C_2H_4, C_2H_6, CH_4, CO, CO_2, N_2, O_2, and C_3H_8) are generated as shown in Figure 2-2. The ratio of these gases varies according to the cause of the decomposition. Therefore, DGA is usually utilized to analyze the gas ratio and to infer the status of the transformer.

From a thermodynamic point of view, the severity of transformer fault is very related to enthalpy that changes during the gas decomposition [36]. The gas decomposition that has larger enthalpy change occurs faster by more severe faults. Table 2-2 summarizes the chemical reaction formula for each major gas and the enthalpy change ($\Delta H°$) generated when n-octane (C_8H_{18}), the main component of insulating oil, is pyrolyzed. For example, comparing C_2H_2 with CH_4, since the

enthalpy change in C_2H_2 is 278.3 kJ/mol and the enthalpy change in CH_4 is 77.7 kJ/mol, it can be assumed that more severe faults occurred when C_2H_2 is produced. Accordingly, methods for diagnosing a transformer based on the correlation between the transformer fault and the gas concentration have been developed and widely used.

There are two main ways of DGA: offline and online. The offline is a method in which an operator directly collects insulating oil and measures the gas concentration using gas chromatography equipment in the laboratory. The online method is to measure gas concentration in real time by an online sensor installed on a transformer. The offline method has a disadvantage in that data errors may occur depending on the skill of the operator, and data sampling interval is too long, from 1 months to 2 years. So, in recent years, there is a trend of switching to an online method. However, verification of sensor performance is still weak, and equipment that can measure the concentration of six or more major gases is very expensive. So, the penetration rate in the industrial sites is very low. Therefore, in this study, we conducted on more practical offline data in consideration of the current state of industrial sites.

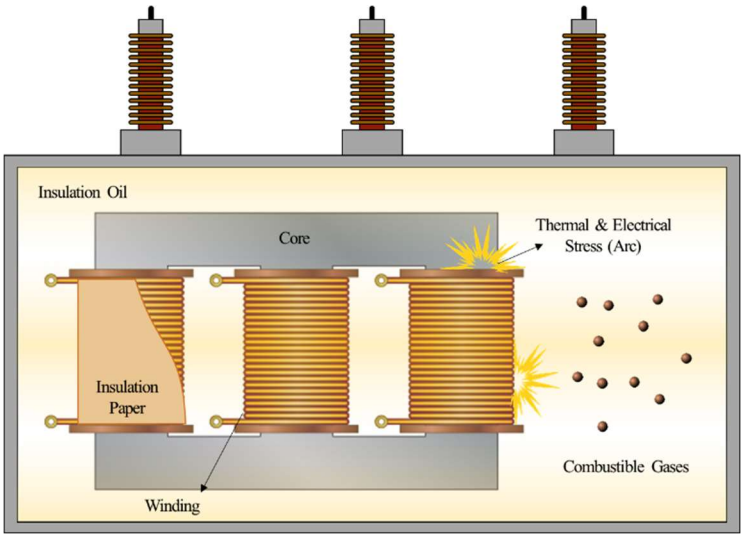


Figure 2-2 Generation of combustible gases by thermal and electrical stress

Table 2-2 Chemical reaction and enthalpy change of major gases

Gas	Reaction Formula	ΔH° (kJ/mol)
$CH_4(g)$	$C_8H_{18}(l) = CH_4(g) + C_7H_{14}(l)$	77.7
$C_2H_6(g)$	$C_8H_{18}(l) = C_2H_6(g) + C_6H_{12}(l)$	93.5
$C_2H_4(g)$	$C_8H_{18}(l) = C_2H_4(g) + C_6H_{14}(l)$	104.1
$H_2(g)$	$C_8H_{18}(l) = H_2(g) + C_8H_{16}(l)$	128.5
$C_2H_2(g)$	$C_8H_{18}(l) = C_2H_2(g) + C_6H_{14}(l) + H_2(g)$	278.3

2.3 Conventional Fault Diagnosis of Power Transformers

There are various rule-based diagnostic methods with DGA. These methods have been mainly established from empirical hypotheses or know-how of field experts for a long time. Nevertheless, a fault that is still difficult to identify occurs, and different results may be derived depending on the interpretation method.

Therefore, they are not acceptable for reliable diagnosis methods yet. There are a method of determining the severity based on the gas concentration and a method of determining the failure mode based on the gas composition ratio. The six most commonly used methods are presented below:

(1) IEEE Standard C57.104 [1]: A four-level criterion of 'IEEE Std C57.104™-2008 - IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers' is provided to classify risks to transformers. The content includes threshold of the dissolved gas concentrations for the individual gases and TDCG from Condition 1 to Condition 4. The condition for a target transformer is determined by finding the highest level for individual gases or the TDCG in Table 2-3.

(2) IEC 60599 [2]: IEC 60599 is a DGA interpretation guide provided in 'IEC 60599 – Mineral oil-filed electrical equipment in service – Guidance on the interpretation of dissolved and free gases analysis.' It is similar to IEEE C57.104, but it classifies the status in three grades (Normal, Caution I, Caution II) and excludes TDCG. Table 2-4 summarizes the specific threshold by the condition.

(3) Dornenburg ratio method [4]: Doernenburg ratios is an evaluation of possible fault type method provided in IEEE Std C57.104™ for diagnosis of fault mode. As shown in Table 2-5, R1 (Ratio1, CH_4/H_2), R2 (Ratio2, C_2H_2/C_2H_4), R3 (Ratio3, C_2H_2/CH_4), R4 (Ratio4, C_2H_6/C_2H_2) are compared to limiting values, suggests corresponding fault mode.

(4) Rogers ratio method [5]: The Roger ratios method follows the same general procedures as the Doernenburg method, except only three ratios (R1 =C_2H_2/C_2H_4, R2=CH_4/H_2, R5=C_2H_4/C_2H_6). But, as with the Doernenburg method, the Rogers ratios can give ratios that do not fit into the diagnostic codes. Table 2-6 gives the values for the three key gas ratios corresponding to suggested diagnosis.

(5) Basic gas ratio method [2]: Except Rogers Ratios methods provided in IEEE Std C57.104™-2008, IEC 60599 also guides fault identification method using different three gas ratios (R1=C_2H_2/C_2H_4, R2=CH_4/H_2, R3=C_2H_4/C_2H_6). Each of six board classes of faults leads to characteristic pattern of hydrocarbon gas composition, which is described in Table 2-7.

(6) Duval triangle method [3]: Duval triangle is a fault identification method provided in IEC 60599. As shown in Figure 2-3, the triangle is divided by six faults mode zones and depending on the three gas ratios values (R1=$C_2H_2/C_2H_2+C_2H_4+CH_4$, R2=$C_2H_4/C_2H_2+C_2H_4+CH_4$, R3=$CH_4/C_2H_2+C_2H_4+CH_4$), it indicates the corresponding fault mode.

Table 2-3 IEEE Standard C57.104

Status	Dissolved key gas concentration limits [ppm]							
	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2	TDCG
Condition 1	100	120	1	50	65	350	2500	720
Condition 2	101-700	121-400	2-9	51-100	65-100	351-570	2501-4000	721-1920
Condition 3	701-1800	401-1000	10-35	101-200	101-150	571-1400	4001-10000	1921-4630
Condition 4	> 1800	> 1000	> 35	> 200	> 150	> 1400	> 10000	> 4630

Table 2-4 IEC 60599

Status	Dissolved key gas concentration limits [ppm]						
	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2
Normal	50	30	2	60	20	4000	3800
Caution I	51-100	31-130	3-20	61-280	21-90	401-600	3801-14000
Caution II	> 100	> 130	> 20	> 280	> 90	> 600	> 14000

Table 2-5 Doernenburg ratio method

Suggested fault diagnosis	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)
1) Thermal decomposition	> 1.0	< 0.75	< 0.3	> 0.4
2) Partial discharge (low-intensity PD)	< 0.1	Not significant	< 0.3	> 0.4
3) Arcing (high-intensity PD)	> 0.1 to < 1.0	> 0.75	> 0.3	< 0.4

Table 2-6 Rogers ratio method

Cases	R2 (C_2H_2/C_2H_4)	R1 (CH_4/H_2)	R5 (C_2H_4/C_2H_6)	Suggested fault diagnosis
0	< 0.1	> 0.1 to < 1.0	< 1.0	Unit normal
1	< 0.1	< 0.1	< 1.0	Low-energy density arcing
2	0.1 to 3.0	0.1 to 1.0	> 3.0	Arcing: High-energy discharge
3	< 0.1	> 0.1 to < 1.0	1.0 to 3.0	Low temperature thermal
4	< 0.1	> 1.0	1.0 to 3.0	Thermal < 700℃
5	< 0.1	> 1.0	> 3.0	Thermal > 700℃

Table 2-7 Basic gas ratio method

Cases	Suggested fault diagnosis	R1 (C_2H_2/C_2H_4)	R2 (CH_4/H_2)	R3 (C_2H_4/C_2H_6)
PD	Partial discharges	Not significant	< 0.1	< 0.2
D1	Discharges of low energy	> 1	0.1 – 0.5	> 1
D2	Discharges of high energy	0.6 – 2.5	0.1 – 1	> 2
T1	Thermal fault t < 300℃	Not significant	> 1 but not significant	< 1
T2	Thermal fault 300℃ < t < 700℃	< 0.1	> 1	1 – 4
T3	Thermal fault t > 700℃	< 0.2	> 1	> 4

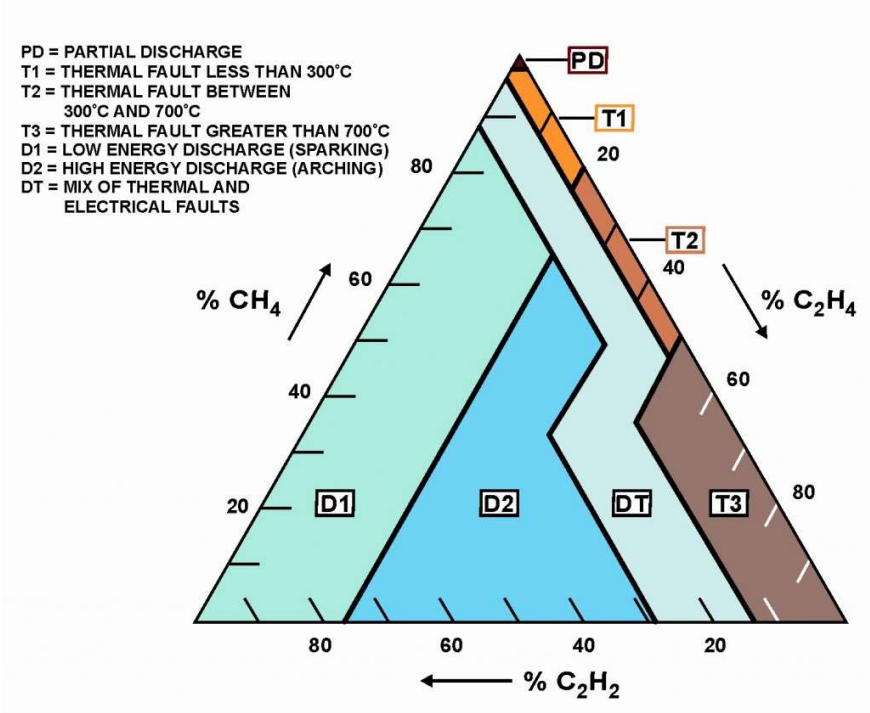


Figure 2-3 Duval triangle method

Chapter 3

Missing Data Imputation via Iterative Denoising Autoencoder (IDAE)

In this chapter, iterative denoising autoencoder (IDAE) is proposed to impute the multiple missing DGA. DGA is missing for various reasons. Therefore, it is important to restore DGA for reliable fault diagnosis. Subchapter 3.1 deals with the data issues arising from DGA. Subchapter 3.2 describes the theoretical background of denoising autoencoder (DAE), which is the basis of this study. Subchapter 3.3 proposes the new imputation methodology consisting of three processes: 1) data preprocessing, 2) construction of DAE model for a single missing value, and 3) construction of IDAE model for multiple missing values. Subchapter 3.4 considers the results of three case studies: 1) a comparison with conventional methods on imputation performance for a single gas, 2) examination of imputation performance between multiple missing values, and 3) documentation of diagnosis accuracy before and after imputation. At the conclusion, Subchapter 3.5 provides the summary and discussions of this study.

3.1 Data Issues in DGA

Although DGA is widely used due to its simplicity and generality in inferring the various failures of a transformer, a problem arises when this method is applied in practice; specifically, it has difficulty gathering reliable data. Dissolved gas data is subject to corruption for various reasons. This is described in reference [25], which notes the importance of careful data analysis. Some cases of unreliable data are as follows. First, the gas ratio can be significantly different between samples. The samples obtained at similar times should have similar quantities; if not, the samples should be considered to be error. A low concentration of H_2 or CO_2 also indicates possible error. Because the solubility of these gases is very low, the oil rapidly loses these gases when exposed to the air. Therefore, a low ratio of H_2 or CO_2 means that the sample is corrupted by exposure to the air. Similarly, an increase in O_2 and N_2 occurs during the exposure, because these gases are abundant in the air and readily dissolved in the oil.

The causes of unreliable data vary. Most of these causes stem from careless handling of samples. In many cases, the samples are manually gathered by workers. While taking the samples, it is easy for the samples to be exposed to the air if they are not completely sealed. Also, the gases can be lost when the transformer has a small crack or rupture in the transformer housing. Some minor cases of unreliable data include mis-labeling, data transcription error, and incorrect analysis. Unreliable data should be excluded from the analysis. However, transformer data is not frequently obtained; it is typically measured only once or twice a year. Therefore, every data sample is valuable and should be fully exploited if possible, rather than being excluded due to reliability concerns. To this end, the proposed method is

developed to enhance the useability of unreliable or incomplete data.

3.2 Backgrounds of Denoising Autoencoder (DAE)

To estimate the missing values in a transformer's DGA data, a denoising autoencoder (DAE), first proposed by P. Vincent and Y. Bengio, is adopted [37]. DAE is the expanded version of an autoencoder, which is used to recover the original data from noise-corrupted data. DAE is based on the fact that the data maintains its essential characteristics, even when partially destroyed. Therefore, the DAE model can recover the original data from the noise-added input. DAE is widely used for image/voice recovery, typo correction, and noise filtering, among other applications [38-43].

The structure of the DAE model is shown in Figure 3-1. It consists of two parts. The first part is the autoencoder, which itself can be decomposed into encoder and the decoder parts. The role of the encoder is so-called manifolding learning, which sequentially reduces the dimension of the input data. As a result, the essence of the original data, called the latent values, which contains enough information about the original data, is obtained. For given data $\mathbf{x} \in \boldsymbol{R}^D$, the encoder function, f_θ is expressed as shown in Eq. (1).

$$f_\theta(\mathbf{x}) = \mathbf{h} = s(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{1}$$

where **W** is the d×D dimensional weight matrix, **b** is the d dimensional bias vector, **h** is the d dimensional latent value, and s(·) is the activation function.

The decoder is the reverse process of the encoder, which is called generative model learning. The decoder uses sequentially increasing layers, and recovers the original data from the output of the encoder. The equation for the decoder, $g_{\theta'}$ is as follows.

$$g_{\theta'}(\mathbf{h}) = \hat{\mathbf{x}} = s(\mathbf{W}'\mathbf{h} + \mathbf{b}') \tag{2}$$

where $\hat{\mathbf{x}} \in \boldsymbol{R}^D$ is the recovered data from the input of the decoder or, equivalently, the encoder of output, $\mathbf{h}$, and weight parameter $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{b}\}$ and $\boldsymbol{\theta}' = \{\mathbf{W}', \mathbf{b}'\}$ are estimated through the learning process.

The autoencoder learns the weight parameter $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ by minimizing the loss function $L(\boldsymbol{\theta}, \boldsymbol{\theta}')$, which measures the similarity between $\mathbf{x}$ and $\hat{\mathbf{x}}$. In this study, the mean-square error function of Eq. (3) is adopted as a loss function to estimate the missing values. Given the training dataset $\{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$, the loss function is minimized by updating $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ through a backpropagation method that is based on the gradient descent algorithm [44].

$$L(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{N}\sum_{k=1}^{N}\|\mathbf{x}^k - \hat{\mathbf{x}}^k\|^2 = \frac{1}{N}\sum_{k=1}^{N}\|\mathbf{x}^k - g_{\theta'}(f_\theta(\mathbf{x}^k))\|^2 \tag{3}$$

The other part of DAE is the addition of noise to the raw data. The noisy data after this step is designated as $\tilde{\mathbf{x}}$ in Figure 3-1. The noisy data is generated using the stochastic corruption procedure $\tilde{\mathbf{x}} \sim q_D(\tilde{\mathbf{x}}|\mathbf{x})$ [45]. Through this process, about half of the input data is randomly substituted with zero. Noisy data is then provided to the autoencoder model, which generates $\hat{\mathbf{x}}$. Since the loss function is defined as the error between the original data $\mathbf{x}$ and the reconstructed data $\hat{\mathbf{x}}$, as the training goes

on, the model takes noisy input $\tilde{\mathbf{x}}$ and gives an output similar to original data **x**. In other words, the model now acts like a noise filter.

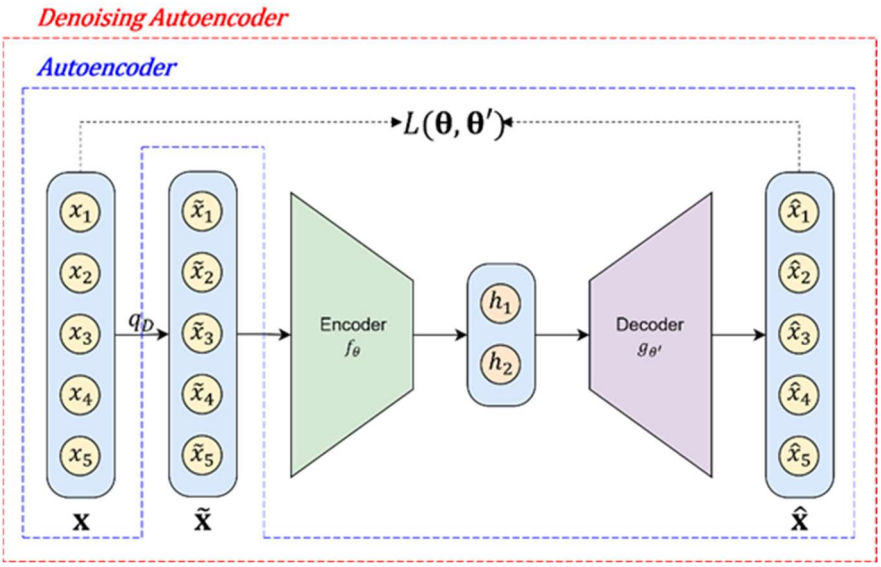


Figure 3-1 The structure of the DAE model

3.3 DAE-based Iterative Imputation for Multi-Missing DGA

The overall process of the proposed method is shown in Figure 3-2. It consists of three main steps; data preprocessing, DAE training, and IDAE. In the data preprocessing step, the input dimension is defined and normalization is performed. In the DAE training step, a DAE model to recover a single missing value is constructed. Then, IDAE, an imputation model for the multiple missing value situation, follows. The initial values for the multiple missing values are first assigned by KNN, and the IDAE updates the initial guesses to robust values. In the following subsections, each step is described in detail.

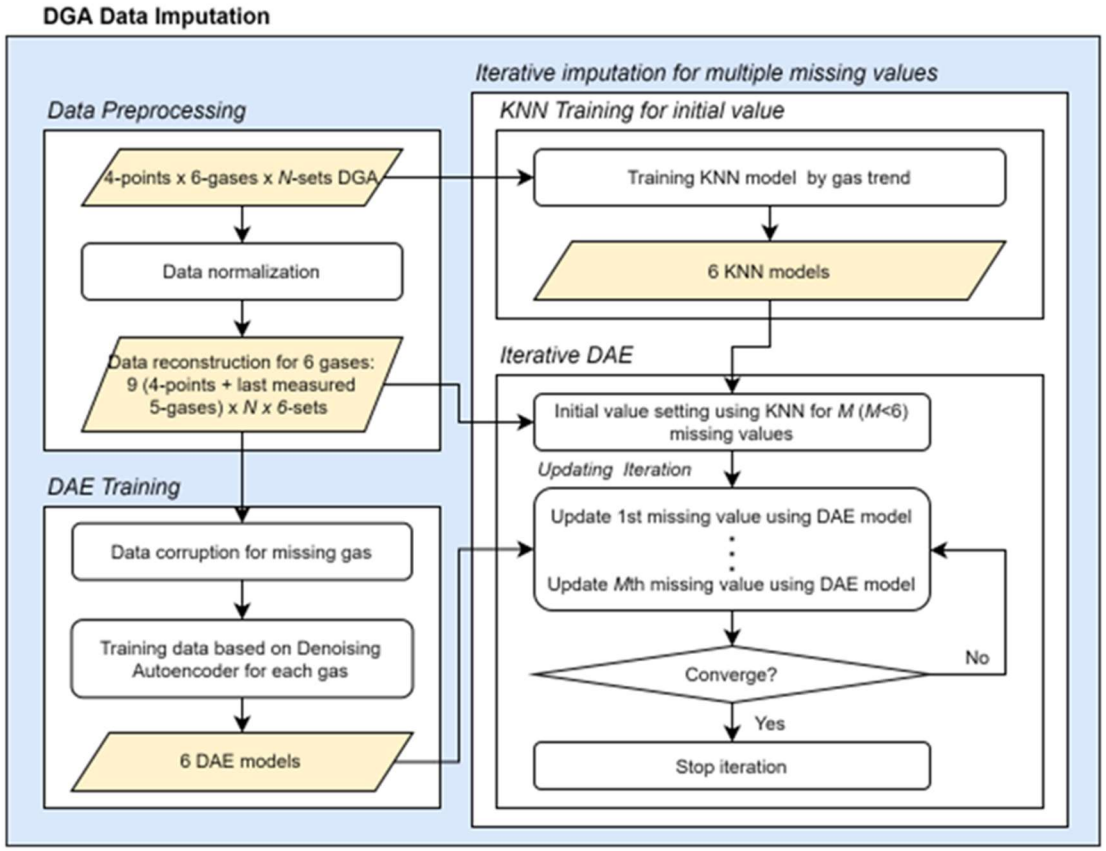


Figure 3-2 Overall process for missing data imputation in DGA

3.3.1 Data Preprocessing

First, to apply the DAE model to the transformer data, the input needs to be defined for accurate reconstruction of the missing data. One way of doing this is to use the time-trend of the DGA data. As the gases are accumulated over time, the amount of each gas monotonically increases. Hence, the missing data may have a value in between the values before and after the missing data. However, DGA data is generally not measured often (i.e., once or twice a year); thus, a gradual change sometimes is not clearly seen because of interruptions by other electrical events. For example, sudden partial discharge could rapidly increase the amount of gases. Therefore, when we estimate the missing gas concentration only using time-series trend information, it will not be able to interpret the case where the gas has abrupt change.

Thus, in this study, in addition to the time-trend data, the gas concentrations are also used to facilitate the determination of the missing value at a certain time. The amount of a gas can be estimated using the other gases, when the failure mode is verified. As a result, the input of the DAE model is designed to include both the time-series data of a gas and the amounts of other gases at a specific time, t. The input is defined as follows.

$$\mathbf{x} = [x_{t-3}^{i}, x_{t-2}^{i}, x_{t-1}^{i}, x_{t}^{i}, x_{t}^{i+1}, x_{t}^{i+2}, x_{t}^{i+3}, x_{t}^{i+4}, x_{t}^{i+5}] \quad (4)$$

where the superscript and subscript of x are gas type and measurement time, respectively. In Eq.(4), x_t^i represents the missing value and will be estimated using both the time-series data, $[x_{t-3}^{i}, x_{t-2}^{i}, x_{t-1}^{i}]$, and gas concentration at time t, $[x_{t}^{i}, x_{t}^{i+1}, x_{t}^{i+2}, x_{t}^{i+3}, x_{t}^{i+4}, x_{t}^{i+5}]$. The six gases include H_2, C_2H_2, C_2H_4, C_2H_6, CH_4,

and CO. The missing value x_t^i is set to be -0.1, which is physically impossible to obtain, so the learning algorithm can recognize it.

Another preprocessing step is normalization. The scales of the dissolved gases are different. Some gases are more easily generated and dissolved in the oil than other gases because of their low enthalpy. In contrast, some gases such as C_2H_2 require high enthalpy to be formed. In this case, the gas is generated only when there is a fault that releases high energy such as arc or corona. Because this high-energy condition is not readily met, C_2H_2 shows rather low concentrations in the oil. However, a fault with high energy indicates a possible severe fault; thus, it is an important indicator regardless of low concentrations in the oil. Thus, to reduce the scale difference, min-max normalization is adopted, which transforms the data to have a value between 0 and 1.

3.3.2 DAE Model for a Single Missing Value

Using the input defined previously, the model is trained. Before dealing with multiple missing values, a model for a single missing value is first trained for each gas, based on DAE. As a result, a total of six models are obtained. The structure of the DAE model is shown in Fig. 3. In the input layer, the data marked '×' represents the missing data. Each input consists of nine elements, including the time trends and gas concentrations, as mentioned earlier. The corrupted inputs are reconstructed through the encoder and decoder layers. The encoder and decoder layers consist of five hidden layers. Since the encoder and decoder have low input dimension and all input elements are meaningful, the layers are designed to have a fully connected

layer denoted as ‘fc’, as shown in Figure 3-3. The number of nodes of each hidden layer are set to 40, 20, 10, 20, and 40, respectively. Every layer except the last uses the exponential linear unit (ELU) function as its activation function, as shown in Eq. (5). The last layer uses the hyper-tangent function described in Eq. (6) to let the input data fall between 0 and 1. The output through the auto-encoder has recovered value, which is denoted as ‘o’ in the figure. The mean square error of Eq. (3) is adopted for the loss function, and training is conducted to minimize the loss function. This trained model is only applicable for data with one missing value. However, many DGA data include multiple missing values. Thus, an iterative method for multiple missing values is proposed in the following section.

$$R(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases} \tag{5}$$

$$\tanh = \frac{1-e^{-x}}{1+e^{-x}} \tag{6}$$

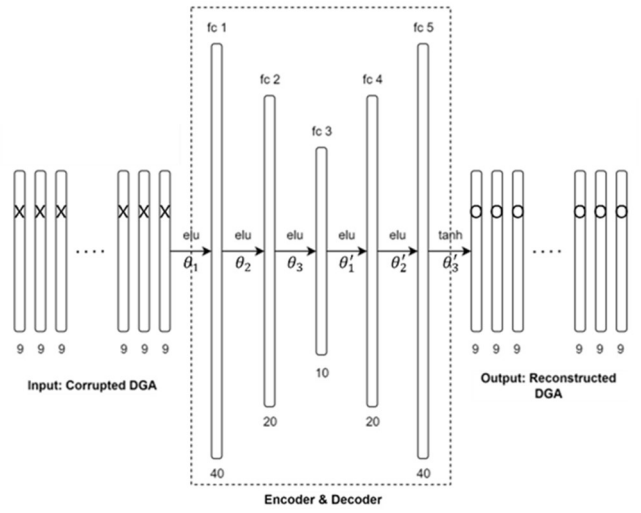


Figure 3-3 Example of the DAE model structure for one gas

3.3.3 IDAE Model for Multiple Missing Values

The previous DAE model works only for a single missing value. To expand the previous DAE model to data with multiple missing values, the missing values are initially estimated and iteratively updated until they converge. For initial estimation, the k-nearest neighborhood (KNN) method is adopted, because it shows quite high accuracy when there is enough data whose dimension is small. In addition, the learning method is simple and requires little effort to estimate parameters.

The KNN method estimates the target value by finding samples that are similar to the target data. If the target data is $x_{t,0}^i$, which is the *i*th gas at time *t*, the input to the KNN model is three time-series data before $x_{t,0}^i$, which is represented as $\mathbf{z}_0^i = [x_{t-3,0}^i, x_{t-2,0}^i, x_{t-1,0}^i]$. Then, the distances between $\mathbf{z}_0^i$ and the *j*th sample $\mathbf{z}_j^i = [x_{t-3,j}^i, x_{t-2,j}^i, x_{t-1,j}^i]$ are measured as follows.

$$d_j^i(\mathbf{z}_0^i, \mathbf{z}_j^i) = \sqrt{(\mathbf{z}_j^i - \mathbf{z}_0^i)(\mathbf{z}_j^i - \mathbf{z}_0^i)^T}, \quad (j \neq 0) \tag{7}$$

Finally, *K* samples of $\mathbf{z}_k^i$ (k=1,…,K) with the smallest distances from $\mathbf{z}_0^i$ are selected and their corresponding data at time *t*, $x_{k,t}^i$ is used for estimation of $x_{0,t}^i$, as shown in Eq. (8). In this study, we set the number of nearest neighbors as 14 (K=14).

$$x_{0,t}^i = \sum_{k=1}^{K} \frac{x_{k,t}^i}{d_k^i} \Big/ \sum_{k=1}^{K} \frac{1}{d_k^i} \tag{8}$$

Once the missing values are replaced with the output of KNN, the DAE model becomes applicable. The output of the DAE model, $\hat{\mathbf{x}} = g_{\theta'}(f_\theta(\mathbf{x}))$, is expected to be more accurate than the output of the KNN; however, it is different from the target

value because it is trained based on the outputs of KNN, which are roughly estimated initial values. The outputs of the DAE models then replace the outputs of the KNN. This updating process goes on until the difference between the updated value and the previous value is minimized. When there are M missing values, the difference is measured by the following equation, where l is the index of the gas that has been missed.

$$\Delta = \frac{\sum_{l=1}^{M} \Delta^l}{M} = \frac{1}{M} \sum_{l=1}^{M} \frac{\left| \hat{x}_{\text{new}}^l - \hat{x}_{\text{old}}^l \right|}{\left| \hat{x}_{\text{new}}^l \right|} \tag{9}$$

The pseudo-code for the algorithm is shown in Table 1. $\tilde{\mathbf{x}}$ is the missed input data and N is the number of data samples.

Table 3-1 Pseudo-code for the IDAE algorithm

Algorithm Impute missing values with IDAE.

Requirements: Input dataset $\widetilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N\}$ with missing values.

1. Initialize:
2. $\mathbf{l} = \{l_1, \ldots, l_N\}$; Define $\mathbf{l}$ where l_n is missing value index for nth data.
3. $\widehat{\mathbf{X}} = \emptyset$; Dataset of imputed input.
4. **for** n in N **do**
5. Make initial guess $\hat{\mathbf{x}}_{n,\text{old}}$ for missing values using KNN.
6. **while** $\Delta_{\text{new}} < \Delta_{\text{old}}$ **do**
7. **for** l_n in $\mathbf{l}$ **do**
8. Predict $\hat{\mathbf{x}}_{n,\text{new}}$ corresponding to $\tilde{\mathbf{x}}_n$ using DAE model;
9. **end for**
10. Calculate Δ_{new};
11. **end while**
12. Append $\hat{\mathbf{x}}_{n,\text{new}}$ to $\widehat{\mathbf{X}}$;
13. **end for**
14. **return** imputed matrix $\widehat{\mathbf{X}}$

3.4 Performance Evaluation of IDAE

The proposed method was verified by examining actual industrial data. Three comparative studies were conducted. The first and second case studies were explored for verification of the imputation model for single and multiple missing values, respectively. The last case study compared the diagnostic performance with and without imputation.

3.4.1 Description of DGA Dataset

The DGA data used for the case studies in this research was provided by the Korea Electric Power Corporation (KEPCO). The data was gathered from more than 8,000 transformers over 10 years. This data reflects the intact characteristics of the operating conditions and data gathering conditions. The first rated voltage of the transformers in the data set varies from 22.9kV to 765kV; the transformers were manufactured by 49 different companies.

Based on the experts from KEPCO and the literature reports on DGA methods, six primary gases, H_2, C_2H_2, C_2H_4, C_2H_6, CH_4, and CO, were selected for analysis. Other gases, such as O_2, N_2, and CO_2, were excluded from the analysis because these gases can be generated from reasons other than degradation. Erroneous data, data with an obviously incorrect value, and data corrupted by oil filtering, were excluded. The total number of data samples in the final study set is 17,850. About 90% of these were used for training, and the remaining 10% was used for testing.

3.4.2 Performance of Estimation Model for a Single Missing Value

For the first comparative study, the accuracy of single missing value estimation will be compared using the proposed method and other previous methods. The comparison algorithms used alongside the proposed method are k-nearest neighbor (KNN) [46], XGBoost (XGB) [47], light GBM (LGB) [48], and random forest (RF) [49].

The accuracy of each algorithm is measured using two metrics. First is the normalized root mean square error (NRMSE), which is the root mean square error divided by the mean of the square of the original value. It is defined in Eq. (10).

$$\text{NRMSE} = \sqrt{\frac{\sum_{k=1}^{N}\left(\mathbf{x}^{k}-\hat{\mathbf{x}}^{k}\right)^{2}}{\sum_{k=1}^{N}\left(\mathbf{x}^{k}\right)^{2}}} \tag{10}$$

NRMSE alleviates the scale difference of gases, so it allows a fair comparison between gases. However, since NRMSE is vulnerable to outliers, it can give biased results, despite generally good performance. Hence, another metric, correlation coefficient, is adopted to supplement the NRMSE, as shown in Eq. (11). The correlation coefficient measures the linearity between the target values and the estimated values, a higher value indicates better estimation.

$$r = \frac{\sum_{k=1}^{n}\left(\mathbf{x}^{k}-\bar{\mathbf{x}}\right)\left(\hat{\mathbf{x}}^{k}-(\bar{\hat{\mathbf{x}}})\right)}{\sqrt{\sum_{k=1}^{n}\left(\mathbf{x}^{k}-\bar{\mathbf{x}}\right)^{2}}\sqrt{\sum_{k=1}^{n}\left(\hat{\mathbf{x}}^{k}-(\bar{\hat{\mathbf{x}}})\right)^{2}}} \tag{11}$$

where $\bar{\mathbf{x}}$ indicates the mean value.

Using these metrics, the estimation results are shown in Table 3-2. Both NRMSE and correlation coefficient results show that the proposed DAE method is

superior to the other methods in estimating the gases, including C_2H_6, CH_4, and CO. In the case of C_2H_2, DAE shows the lowest NRMSE, and the correlation coefficient is the second highest. H_2 and C_2H_4 gases are concentrated at 0 ppm; thus, it seems that overfitting has occurred due to data imbalance. Although its performance is somewhat poor for H_2 and C_2H_4, it has the best overall performance.

KNN has the lowest performance for more than three gases; it also takes ranks lower for the other gases. This is because KNN is dependent on the nearest data. Therefore, when there is an abrupt increase or decrease around the missing value, which is common in DGA data due to the long intervals between samplings, KNN shows low accuracy. This phenomenon is dominant, especially for gases that are rarely generated, such as H_2, and C_2H_2. These gases require high enthalpy to be generated; thus, in most cases they are absent. However, when the requirements are met, they are generated, which is regarded as an abrupt appearance in the KNN model.

XGBoost, light GBM, and random forest are based on the decision tree. Tree-based methods learn criterion that separate the data into groups with similar properties. The criterion is represented as the branch of the tree, and the more the branch is subdivided, the more elaborate the model obtained. However, with DGA data, the decision tree methods have difficulty subdividing the criterion, because most of the DGA data is close to 0 ppm. However, these tree-based algorithms show generally better results than KNN, because the finely divided models are merged by ensemble learning and avoid the overfitting. The tree-based ensemble models performed generally better than DAE for H_2 and C_2H_4 because of these overfitting prevention characteristics. In particular, RF is not overfitted more than the other

algorithms, because it uses a bagging method. However, when the distribution is relatively evenly spread, the performance of DAE is better; this is confirmed for the rest of the four gases.

Table 3-2 Imputation performance comparison for a single missing value

Metric	Methods	H_2	C_2H_2	C_2H_4	C_2H_6	CH_4	CO
NRMSE	DAE	0.532	**0.39**	0.302	**0.165**	**0.187**	**0.213**
	XGB	0.506	0.525	0.313	0.270	0.372	0.317
	LGB	0.479	0.425	0.262	0.202	0.272	0.222
	RF	**0.459**	0.425	**0.241**	0.202	0.204	0.255
	KNN	0.591	0.579	0.275	0.279	0.330	0.226
Correlation coefficient	DAE	0.807	0.904	0.934	**0.981**	**0.974**	**0.907**
	XGB	0.818	0.855	0.944	0.952	0.913	0.820
	LGB	0.824	0.881	0.957	0.974	0.947	0.900
	RF	**0.834**	**0.907**	**0.961**	0.974	0.969	0.883
	KNN	0.730	0.770	0.945	0.951	0.927	0.899

3.4.3 Performance of Imputation for Multiple Missing Values

Here, the performance results of IDAE are compared as the number of missing values increases from one to four. 80 samples are generated for each possible combination. For example, there are 12 possible combinations for two missing values, so 960 samples are generated. In this way, a total of 2,480 samples are prepared.

All missing samples were imputed by applying the proposed IDAE. In Table 3-3, the NRMSE and correlation coefficient between the imputed value and the actual value are summarized. As the number of missing values increases, the performance gradually decreases. This is a natural result, because errors in the single gas imputation model accumulate as the missing values increase.

Comparing the performance decline levels when there is one missing value and four missing values, the NRMSE increased by 0.096 and the correlation coefficient decreased by 0.04. As the number of missing values increases, the NRMSE increases by 0.032 and the correlation coefficient decreased by 0.013, on average. This means that as missing values are added one by one, the performance declines about 3.2% for NRMSE and 1.3% for the correlation coefficient approach. This seems to be very stable because the performance reduction rate according to the number of missing values is very small.

Table 3-3 Imputation performance of IDAE according to the number of missed DGA values

Number of missing values	1	2	3	4
NRMSE	0.289	0.313	0.330	0.385
Correlation coefficient	0.919	0.904	0.894	0.879

3.4.4 Performance of Diagnostic Model with Recovered Data

Original and recovered data are used to construct the diagnosis models, and the diagnostic performances are compared. For diagnostic models, two models, IEEE Std. C57.104-2008 and semi-supervised autoencoder (SSAE) model which will be introduced in Chapter 4, are used. IEEE Std. C57.104-2008 is the standard rule-based diagnostic method and is widely used. The model accepts H_2, C_2H_2, C_2H_4, C_2H_6, CH_4, CO, CO_2, and TDCG gases as input and gives the health status of the transformer with four severity levels.

For quantitative analysis of the diagnostic performance, the recall metric is used. The recall metric is defined as the ratio of the true positive to the sum of the true positive and false negatives, as shown in Eq. (12).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{12}$$

where TP and FN stand for the true positive and the false negative, which indicate correct and incorrect estimation on the faulty state, respectively. A value close to 1 indicates better performance in recall. Although there are other metrics, recall is considered a suitable metric for a diagnosis model because in diagnosis, knowing the faulty state correctly is more important, even at the expense of false alarm.

Table 3-4 shows the recall scores of the IEEE and SSAE models. The same dataset as in the previous section is used. In both the IEEE and SSAE models, the performance is improved after data imputation. In particular, the results of the SSAE model are significantly improved after imputation. The SSAE model examines the concentration and composition ratio of all gases together. On the other hand, in the IEEE model, a limit is set for each gas to evaluate the health grade. This limit is judged as the most serious result among them. Therefore, as the number of missing values increases, the SSAE model is more likely than the IEEE model to determine the condition to be normal.

Figure 3-4 shows the true data, missing data, and imputed data in the feature space of SSAE. The imputed data is located near the true data around the faulty area; however, the data with missing values departs from it. Also, as the number of missing values increases, the imputed data maintains the data in the faulty area with a slight drift from the true data, but the missed data is greatly affected. As a result, the imputed data is robust to false alarms.

Table 3-4 Recalls of the IEEE and SSAE models before and after imputation

Model	Imputation	N=1	N=2	N=3	N=4
IEEE	No	0.852	0.770	0.714	0.671
	Yes	0.985	0.927	0.937	0.893
SSAE	No	0.359	0.130	0.018	0
	Yes	0.829	0.794	0.828	0.692

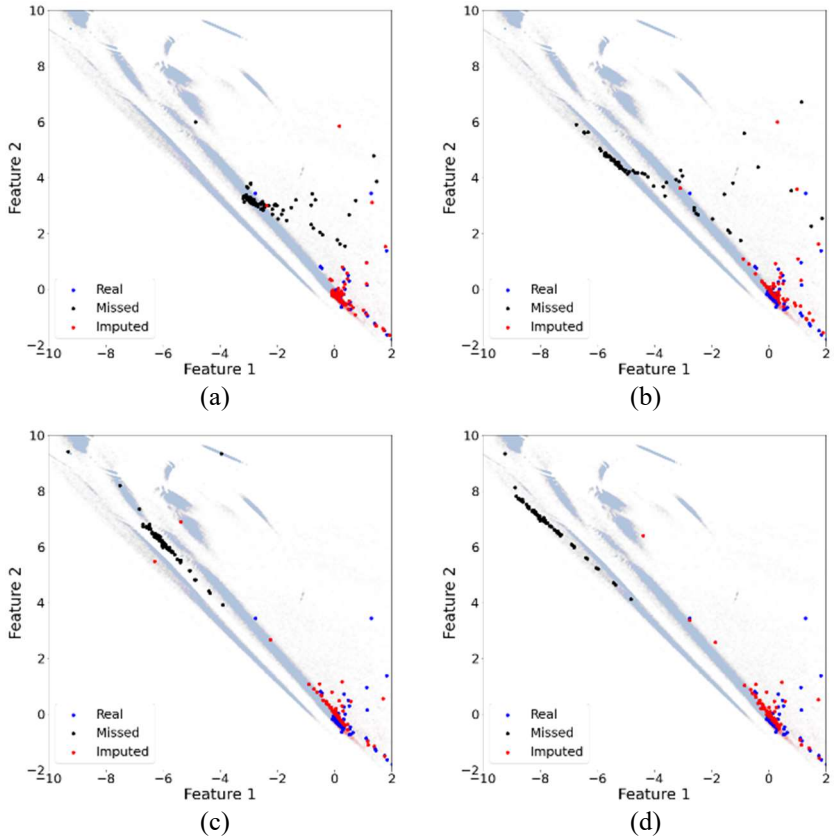


Figure 3-4 The results of SSAE diagnosis before and after the imputation: (a) one missing value, (b) two missing values, (c) three missing values, (d) four missing values

3.5 Summary and Discussion

In this study, we proposed a new methodology that imputes multiple missing values of DGA through DAE-based iterative imputation. The proposed methodology proceeds via three main steps: 1) preprocessing DGA data, 2) learning the DAE imputation model for a single missing value, and 3) iterative imputation of multiple missing values using the DAE model. The model is validated through three case studies that examined a vast amount of real industrial data.

The first study compares the imputation performance for six types of single gases, using conventional algorithms. The DAE model had the best imputation performance for overall gases, so it was adopted as a core imputer in the iterative imputation process. The second study tested the performance of the proposed methodology for imputing multiple missing values. The number of missing values was increased from one to four, and we tested the imputation performance according to the number of missing values. The performance gradually decreased as the number of missing values increased. This is an expected result because the number of unknown variables to be estimated increases. The third study confirmed the difference in diagnostic performance before and after imputation based on existing diagnosis methods for power transformers. The diagnostic methods used in the study are rule-based IEEE Std. C57.104-2008 and an artificial intelligence-based SSAE model. Both methods significantly improved diagnostic accuracy after correction.

Most oil-immersed transformers are managed through DGA. Much research has been conducted to develop accurate diagnosis models; however, prior work has not thoroughly examined the importance of the integrity of the data used in the diagnosis.

In particular, DGA data has a long measurement cycle and a small number of data; thus, each data point is meaningful. Therefore, the methodology proposed in this study is expected to further improve the performance of the diagnosis model and greatly reduce post-mortem maintenance that is required when a wrong diagnosis is provided due to missing data.

In addition, the proposed method has the advantage of being applicable regardless of the data domain to the missing data problem. Traditional imputation algorithms are specialized in single missing problems occurring from time series or sequence-based data or multi missing problems occurring from the data consisting of various variables. However, the method developed in this study can solve both problems at the same time. Various missing values can be imputed for time series data composed of various variables, and it has originality in that it is a method that has not been attempted before.

Due to the limitations of the offline DGA method, attempts have recently been made in industry to gradually convert to an online DGA method. Online methods have the advantage of higher sampling frequency and lower probability of missing data. However, due to the physical limitations of the measurement method, the data accuracy is lower than that of the offline gas chromatography method. In future studies, research for increasing the accuracy of online DGA data will be conducted.

Sections of this chapter have been submitted as the following journal articles:

1) **Boseong Seo,** Jaekyung Shin, Taejin Kim, and Byeng D. Youn, "Missing Data Imputation Using an Iterative Denoising Autoencoder (IDAE) for Dissolved Gas Analysis," *Electric Power Systems Research*, Submitted, 2022.

Chapter 4

Health Feature Extraction via Semi-Supervised Autoencoder (SSAE)

In this chapter, the new health features are developed for diagnosis of power transformers using semi-supervised autoencoder (SSAE). The limitation of the existing standard rules is that the false alarm rate is high because it does not consider the composition cost or combination between gases, but simply considers the gas concentration level. Further detailed diagnosis is difficult because the condition of the transformer is classified into several severities. Because the decomposition reaction of insulation is very complex, it is difficult to build a physical model. Thus, we would like to extract a monotonic health index that enables more accurate and detailed diagnosis based on deep learning. Subchapter 4.1 explains the theoretical background of autoencoder and softmax classifier, which compose the SSAE model. Subchapter 4.2 proposes the new diagnosis methodology consisting of two steps: 1) data preprocessing, and 2) construction of health feature space (HFS) through SSAE. Subchapter 4.3 evaluates the diagnostic performance of the proposed method comparing with conventional diagnostic methods. At the conclusion, Subchapter 4.4 provides the summary and discussions of this study.

4.1 Backgrounds of SSAE

The main purpose of this study is to extract key features representing the health status of the transformers from DGA data. Semi-supervised autoencoder (SSAE) has excellent performance in learning features that are highly correlated with both input data and labeled data. We use the SSAE method because it is the most suitable method for the research purpose. The SSAE proposed in this study consists of a combination of autoencoder (AE), one of the unsupervised learning techniques for extracting important features by reducing the dimension of data, and softmax classifier (SC), one of the supervised learning techniques for classifying the labeled data. Therefore, the corresponding subchapter briefly introduces the theory of autoencoder and softmax classifier that make up the SSAE model.

4.1.1 Autoencoder: Unsupervised Learning for Feature Extraction

AE is the most representative method of unsupervised learning and has excellent performance in dimension reduction and feature extraction of high-dimensional data [50]. The structure of the autoencoder is divided into an encoder part and a decoder part consisting of a hidden layer as shown in Figure 4-1. Encoder acts as a manifolding learning that performs dimension reduction, and decoder acts as a generative model learning that performs data restoration [51].

For a given learning data $\mathbf{x} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$ $\left(\mathbf{x}^k \in \boldsymbol{R}^d\right)$, the encoder function f_θ reduces the dimension to $\boldsymbol{R}^d \rightarrow \boldsymbol{R}^{d'}$ $(d > d')$ through the model parameter $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{b}\}$ ($\mathbf{W}$: $d \times d'$ *dimensional weight matrix* and $\mathbf{b}$: d' *dimensional bias vector*) and the activation function *s*, as shown in Eq. (1). There are many types of

activation functions, and commonly used functions include sigmoid, modified linear unit (ReLU), and exponent linear unit (ELU), as shown in Table 4-1. Conversely, the decoder function $g_{\theta'}$ reconstructs the dimensionally reduced feature $\mathbf{h} = \{\mathbf{h}^1, \ldots, \mathbf{h}^k\}$ $(\mathbf{h}^k \in \boldsymbol{R}^{d'})$ to $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \ldots, \hat{\mathbf{x}}^k\}$ $\left(\hat{\mathbf{x}}^k \in \boldsymbol{R}^d\right)$ through Eq. (2). The feature $\mathbf{h}$ extracted through the optimization process of $\boldsymbol{\theta}$ is called the latent value, and this feature implicitly represents the main characteristics of the input data.

$$f_\theta(\mathbf{x}^k) = h_j^k = s\left(\sum_{i=1}^{d} W_{ji} x_i^k + b_j\right) \tag{13}$$

$$g_{\theta'}(\mathbf{h}^k) = \hat{x}_i^k = s\left(\sum_{j=1}^{d'} \widehat{W}_{ij} h_j^k + \hat{b}_i\right) \tag{14}$$

AE learns $\boldsymbol{\theta}$ to minimize loss function $L(\mathbf{x}, \hat{\mathbf{x}})$. Loss function $L(\mathbf{x}, \hat{\mathbf{x}})$ is represented by a mean square error (MSE) of the input (original data) and output (reconstructed data), as shown in Eq. (15), and updates $\boldsymbol{\theta}$ through a gradient descent-based back-propagation method [44]. Update equation of $\boldsymbol{\theta}$ is described in Eq. (16), and η means the learning rate, the hyperparameter that optimizes the step size at each iteration.

$$L(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}^i - \hat{\mathbf{x}}^i\right\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}^i - g_{\theta'}(f_\theta(\mathbf{x}^i))\right\|^2 \tag{15}$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \frac{\partial L(\mathbf{x}, \hat{\mathbf{x}})}{\partial \boldsymbol{\theta}} \tag{16}$$

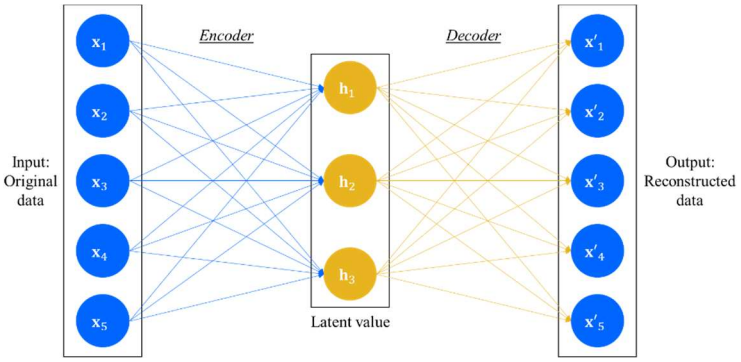


Figure 4-1 The structure of autoencoder

Table 4-1 The typical activation functions of autoencoder

Type	Sigmoid	ReLU	ELU
Equation	$\frac{1}{1+e^{-x}}$	$\max(0,x)$	$\begin{cases} x & x \geq 0 \\ \alpha(e^x-1) & x < 0 \end{cases}$

4.1.2 Softmax Classifier: Supervised Learning for Classification

SC is one of the widely used methods for multinomial classification problems. The softmax function is an activation function originally used to predict the probability distribution for discrete variable with *n* possible values. Therefore, when the softmax function is used as the output unit of the neural network, the output of the neural network is the probability classified into a specific class. The standard softmax function $\sigma: \boldsymbol{R}^n \rightarrow (0,1)^n \ (n > 1)$ is defined by Eq. (17). It normalizes the exponential of each input element z_i by dividing by the sum of all these exponentials. Through the normalization, the sum of all components of the output vector $\sigma(\boldsymbol{z})$ is 1.

$$\sigma(\mathbf{z})_i = \frac{\exp(z_i)}{\sum_j^n \exp(z_j)} \tag{17}$$

Assuming that there are input data $\mathbf{x} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$ $\left(\mathbf{x}^k \in \boldsymbol{R}^d\right)$ and output data $\mathbf{y} = (y_1, y_2, \dots, y_n)$ consisting of n discrete labels, let's calculate the probability $\hat{y}_j^i$ that the given input $\mathbf{x}^i$ belongs to a particular output y_j using the softmax function in Eq. (17). $\hat{y}_j^i$ can be calculated as Eq. (18).

$$\hat{y}_j^i = P\left(\mathbf{y}^i = y_j \middle| \mathbf{x}^i\right) = \sigma\left(\mathbf{z}^i\right)_j = \frac{\exp\left(z_j^i\right)}{\sum_{j=1}^n \exp\left(z_j^i\right)} \tag{18}$$

When the softmax function is used as the output unit of the natural network, $\mathbf{z}^i$ is defined as follows:

$$\mathbf{z}^i = \mathbf{W}_{SC}^{\mathrm{T}} \mathbf{h}^i + \mathbf{b}_{\mathrm{SC}}^i \tag{19}$$

where, $\mathbf{W}_{SC}^{\mathrm{T}}$ is the weight matrix of SC and $\mathbf{b}_{\mathrm{SC}}^i$ is the bias vector of SC, and $\mathbf{h}^i$ is the output vector of the preliminary hidden layer.

If the loss function of the SC model is applied in the same way as the loss function of the linear model, the MSE function, it is expressed in a bumpy form, since the softmax function has the normalized form of the exponential function. Therefore, it is difficult to find the minimum value of loss function by the gradient-descent method. To solve this problem, the log function is used as a loss function of the SC model, and a representative method is the cross-entropy (CE) method. The loss function with CE is presented as Eq. (20), and the exponential function is linearized by taking log to the predicted value of SC. Mathematically, CE means the uncertainty of the difference in similarity between the two probability distributions (in this case, the distribution of actual labeled data $\mathbf{y}^i$ and predicted data $\hat{\mathbf{y}}^i$).

$$L_{CE}\left(\mathbf{y}^i, \hat{\mathbf{y}}^i\right) = -\sum_{i=1}^n \mathbf{y}^i \log\left(\hat{\mathbf{y}}^i\right) = -\sum_{i=1}^n \mathbf{y}^i \log\left(\sigma(\mathbf{z}^i)\right) \tag{20}$$

4.2 SSAE-based Health Feature Extraction

This subchapter describes the process of extracting the health feature for diagnosing and predicting the status of transformers. The process of health feature extraction consists of 2 steps: 1) data preprocessing for SSAE model training, and 2) construction of health feature space using SSAE. In the data processing step, DGA data is normalized to log scale. Then, SSAE model is trained for health feature extraction. The SSAE model has a structure in which the softmax function is implanted into the output layer inside the autoencoder.

4.2.1 Data Preprocessing

In this study, we use the concentration and the composition ratio of the main six gases, H_2, C_2H_2, C_2H_4, C_2H_6, CH_4, and CO, as the input of SSAE model. The enthalpy in which the gas is decomposed varies from gas to gas. Therefore, the amount of gas produced varies greatly depending on the gas as shown in Figure 4-2. Figure 4-2 is an excerpt from the study of S. Bustamante et al. [52], and shows the histogram distribution of DGA data and the positions of 90th and 95th percentiles. For example, CO occurs in thousands to tens of thousands of ppm, while C_2H_2 is only a few ppm. If the difference in the scale of the gas concentration is hundreds of times or more, it is highly likely that the model will be trained to overfit to a gas with high concentration. Therefore, it is necessary to uniformly normalize each gas concentration from 0 to 1. In this study, DGA data is normalized to a logarithmic scale by Eq. (21). The log-scale DGA data might help to avoid the overfitting on the numerical operations.

$$x' = \frac{\log(x) - \log(x_{\min})}{\log(x_{\max}) - \log(x_{\min})} = \frac{\log\left(\frac{x}{x_{\min}}\right)}{\log\left(\frac{x_{\max}}{x_{\min}}\right)} \tag{21}$$

where, x is the original DGA data and x' is the log-scale normalized DGA data.

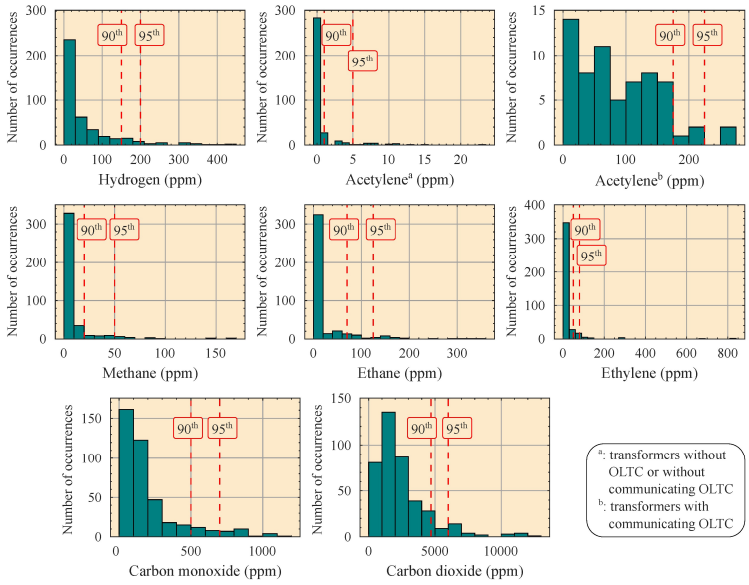


Figure 4-2 Distribution of gas concentrations and positions of 90th and 95th percentiles

4.2.2 Construction of Health Feature Space (HFS) via SSAE

The purpose of this study is to extract features that are highly correlated with the degradation of the transformer from the DGA data. In machine learning, feature extraction refers to a process that analyzes the main components of high-dimensional data and converts them into low-dimensional data. The process of extracting major features from data is to select the most influential features among the various features of the data. Therefore, if a feature is extracted by general methods, there is a possibility that the feature does not include information on the health status of the

transformer. In this study, we developed SSAE method to solve this problem. As illustrated in Figure 4-3, the proposed SSAE has a neural network structure that combine the unsupervised learning, AE, and the supervised classification algorithm, SC, sharing the hidden layers. By sharing hidden layers, it has the advantage of being able to perform feature extraction and health status learning at the same time.

If a feature is extracted by using only the autoencoder, there is a possibility that the feature does not include information on the health status of the transformer. Therefore, in this study, we tried to extract features including health information by applying SSAE to simultaneously train the labeled data for normal and failure. As illustrated in Figure 4-3, the SSAE used in this study has a structure in which the unsupervised learning, AE, and the supervised classification algorithm, SC, share the encoder layer.

For the given training data $(\mathbf{x}, \mathbf{y}) = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^k, \mathbf{y}^k)\}$ ($\mathbf{x}^k$: dissolved gas concentration and ratio data, $\mathbf{y}^k$: status labeled data, $\mathbf{x}^k, \mathbf{y}^{\mathrm{k}} \in \boldsymbol{R}^d$), the loss function of the proposed SSAE model is described in Eq. (22). It is a combination of the loss function of AE (Eq. (15)) and SC (Eq. (20)) which are introduced in Subchapter 4.1.

$$
\begin{aligned}
&L_{\mathrm{SSAE}}\left(\boldsymbol{\theta}^{\mathrm{en}}, \boldsymbol{\theta}^{\mathrm{de}}, \boldsymbol{\theta}^{\mathrm{sc}}\right) = \alpha L_{\mathrm{AE}}\left(\boldsymbol{\theta}^{\mathrm{en}}, \boldsymbol{\theta}^{\mathrm{de}}\right) + (1-\alpha) L_{\mathrm{SC}}(\boldsymbol{\theta}^{\mathrm{en}}, \boldsymbol{\theta}^{\mathrm{sc}}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}^i - \mathbf{x}'^i\right\|^2 - \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}^i \log\left(\mathbf{y}'^i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}^i - g_{\theta^{\mathrm{de}}}\left(f_{\theta^{\mathrm{en}}}(\mathbf{x}^i)\right)\right\|^2 - \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}^i \log\left(\sigma_{\theta^{\mathrm{sc}}}\left(f_{\theta^{\mathrm{en}}}\left(\mathbf{x}^i\right)\right)\right)
\end{aligned}
\tag{22}
$$

where α is the hyperparameter to control the weight between L_{AE} and L_{SC}. SSAE

optimizes the encoder layer parameters $\boldsymbol{\theta}^{\mathrm{en}}$, decoder parameter $\boldsymbol{\theta}^{\mathrm{de}}$, and softmax parameter $\boldsymbol{\theta}^{\mathrm{sc}}$ so that softmax can extract a latent value that can distinguish between normal and failure of a transformer.

As can be seen in Table 4-2, the SSAE used in this study consists of one input layer, three hidden encoder layers, two hidden decoder layers, one autoencoder output layer, and one softmax output layer. ELU is used for unsupervised learning as an activation function, and softmax is used for supervised learning. It is expected that the implicit health features are extracted from shared hidden layers. It is intended to visualize the degradation behavior of the transformer in a 2D space so that it could be intuitively well understood. Therefore, we set the end of the shared hidden layer to have two nodes.

Figure 4-4 shows the health feature space (HFS) reflecting the characteristics of the health feature that deteriorates monotonically. The blue dot and the red dot mean normal and fault respectively, and the decision boundary between normal and fault can be obtained through clustering techniques. HFS can be interpreted that as the degradation of the transformer progresses, the feature moves to the lower right.

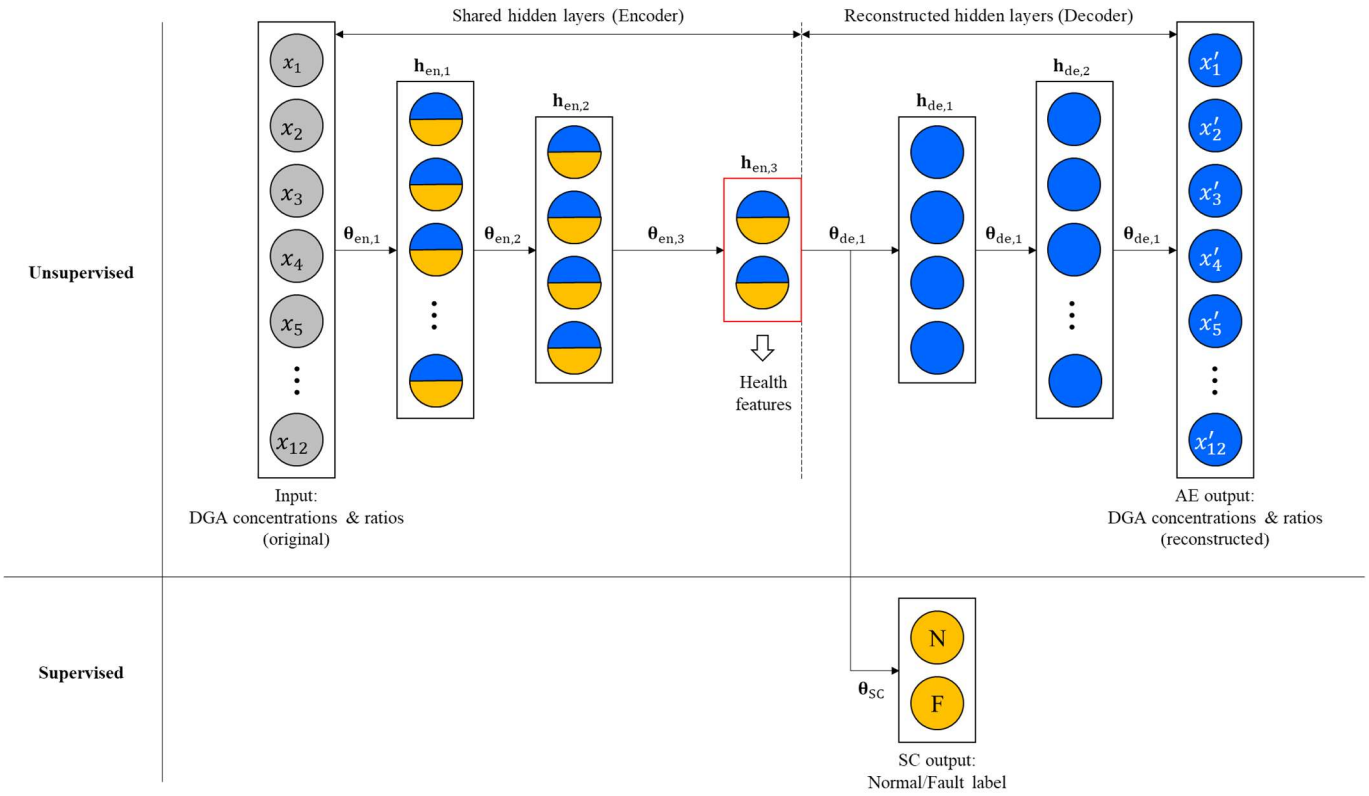


Figure 4-3 Structure of the proposed SSAE

Table 4-2 Parameters in the proposed SSAE

Layer	Activation Function	Number of Nodes	Number of Parameters
Input	-	12	-
Encoder 1	ELU	8	104
Encoder 2	ELU	4	36
Encoder 3	ELU	2	10
Decoder 1	ELU	4	12
Decoder 2	ELU	8	40
Output 1	ELU	12	108
Output 2	Softmax	2	6

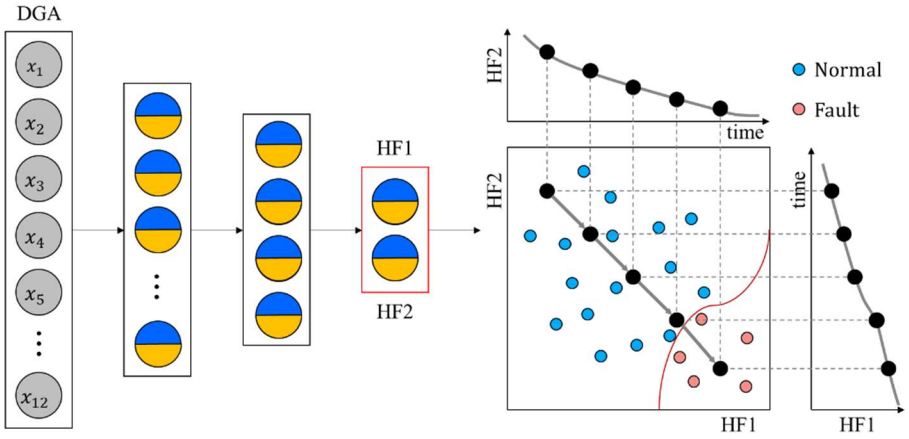


Figure 4-4 Visualization of the health feature behavior in HFS

4.3 Performance Evaluation of SSAE

The SSAE is the AI-based diagnostic model that can classify the normal and fault states of a power transformer by extracting health features that represent the deterioration of the insulation. The SSAE model can visualize the degree of the

degradation, so the effectiveness of the recovered data can be inspected visually. The transformer data in the feature space of the SSAE model is shown in Figure 4-5. It shows the decision boundary between the normal (sky blue) and the fault area (red). When the transformer degrades, the data in the feature space moves toward the red area. The sample degradation data over time is summarized in Table 4-3 and designated in Figure 4-5.

To evaluate the performance of the proposed SSAE model, we compared the diagnosis accuracy of SSAE with that of conventional methods, IEEE and IEC. In Section 4.3.1, dataset used for test is described. And evaluation metrics and experimental results are covered in Section 4.3.2.

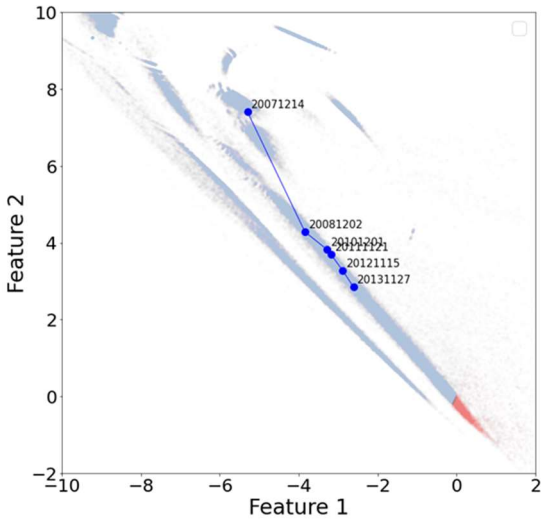


Figure 4-5 Degradation of DGA data in HFS

Table 4-3 Sample DGA data reflecting degradation over time

No.	Acquisition date	H_2	C_2H_2	C_2H_4	C_2H_6	CH_4	CO
1	20071214	0	0	0	0	4	224
2	20081202	2.8	0	12.3	1.2	6.1	1024
3	20101201	6	0	20	1	7	1416
4	20111121	10	0	17	2	10	1302
5	20121115	7	0	16	4	13	1108
6	20131127	3	0	19	7	21	1453

4.3.1 Description of Dataset

We had a total of 141,690 data including 141,573 data provided from KEPCO and 117 IEC TC data, which are public data. Of these, a total of 131,503 data are used for the test after removing all missing data or duplicated data. However, there are several difficulties in using these data as it is. First, there are more than 20,000 unlabeled data, accounting for more than 17% of the total data. Second, the data imbalance problem between normal and failure data is severe. In fact, among the KEPCO data, 121 cases of data marked as failure are very small, but when we checked even this, there were cases where the gas concentration is as low as normal, so the reliability of the handwritten data is low. Finally, the criteria for failure are not clear. Since there is no data recording maintenance action after the failure determination, it is difficult to determine the severity of failure. For example, after determining the failure, it is possible to replace the insulating oil or the internal parts, or entire transformer, but there is no such information. Therefore, we newly established failure criteria and reorganized datasets based on maintenance reports and domain knowledge of substation diagnosis experts.

The criteria for newly defined failure data are as follows:

1) Failure data determined by the detailed inspection in the maintenance report

2) Data with higher numerical values than concentrations of all five gases (H_2, C_2H_2, C_2H_4, C_2H_6, and CH_4) of failure data in the maintenance report

3) Data with C_2H_2 concentration higher than 1 ppm

4) Data with top 50 concentrations of four gases (H_2, C_2H_4, C_2H_6, CH_4) and total combustible gases (TCG)

The number of normal and failure data newly reorganized is 126,664 and 4,839, respectively. By establishing a new failure criterion, it is possible to solve the data imbalance problem to some extent by retaining more failure data and assign the labeling information to non-labeled data. Table 4-4 is an example of DGA data actually used for the test, and as you can see, the concentration and ratio of six gases are used as input of the model, and normal and fault label information is used as output.

Table 4-4 Example of DGA data samples for experiment

Sample	Input						Output
	H_2	C_2H_2	C_2H_4	C_2H_6	CH_4	CO	Status
1	22	0	99	136	122	200	Normal
2	181	0	23	27	67	83	Normal
3	58	3	54	51	54	927	Fault
4	3	0	29	193	121	311	Normal
5	76	7	25	42	31	43	Fault
6	113	13	603	144	237	249	Fault

4.3.2 Diagnosis Accuracy of SSAE for Power Transformers

To evaluate the diagnostic accuracy of the proposed SSAE model, two conventional diagnostic models, IEEE Std. C57.104 and IEC 60599, are used for comparative study. And the following four evaluation metrics are used for quantitative performance evaluation: positive predictive rate (PPV), true positive rate (TPR), true negative rate (TNR), and F1 score. These are mainly used to verify the performance of the classification model. In particular, it is possible to objectively evaluate the model performance for the imbalanced data. PPV refers to the actual positive rate to predicted positive by the model, TPR refers to the predicted positive rate to the actual positive, TNR refers to the predicted negative rate to the actual negative, and F1 score is the weighted average of PPV and TPR. It is closer to 1 as the model performs better. These evaluation metrics are obtained from confusion matrix. Table 4-5 is the confusion matrix of the model, and the evaluation metrics can be mathematically expressed as follows:

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{21}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{22}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \tag{23}$$

$$\mathrm{F1} = \frac{2*\mathrm{PPV}*\mathrm{TPR}}{(\mathrm{PPV+TPR})} \tag{24}$$

Table 4-5 Confusion matrix for model evaluation

		Actual value	
		Normal	Fault
Predicted value	Normal	TN	FN
	Fault	FP	TP

Figure 4-6 represents the confusion matrices of the SSAE, IEEE and IEC models. And four evaluation metrics, PPV, TPR, TNR, and F1, are summarized in Table 4-6. The noticeable difference between SSAE and other models is that PPV is about three to six times higher than other models. That is, the SSAE model has the lowest false alarm rate. In addition, SSAE also shows the best results in TNR and F1 scores.

In the training process, the SSAE model calculates the weight for each gas, representing the correlation between the gas and the status. Then, SSAE comprehensively diagnoses the transformer status based on the gas concentration, gas composition ratio, and gas weight. However, conventional diagnostic methods determine the worst grade based on the thresholds of each gas. So, it is difficult to consider the overall composition ratio and weight. Consequentially, the diagnosis accuracy is highest, and the false alarm rate is lowest in SSAE. In the case of TPR, IEC was 0.035 higher than SSAE because IEC has a greater tendency to determine as failure. Therefore, it is important to evaluate the diagnostic performance using various metrics together, not just TPR.

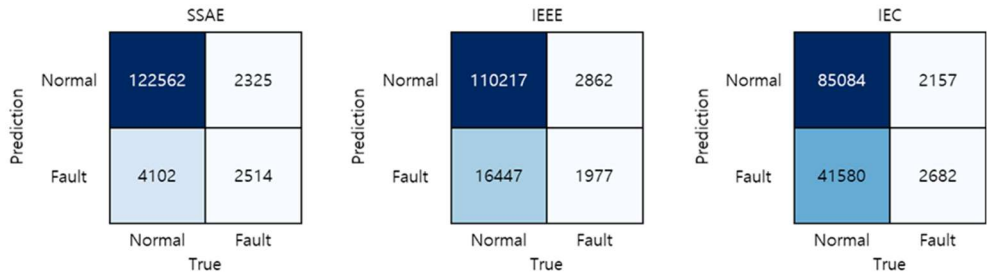


Figure 4-6 Confusion matrices of SSAE and conventional methods, IEEE and IEC

Table 4-6 Health diagnosis performance of SSAE and conventional methods, IEEE and IEC

Model	PPV	TPR	TNR	F1
SSAE	**0.3800**	0.5195	**0.9676**	**0.4389**
IEEE	0.1073	0.4085	0.8702	0.1700
IEC	0.0606	**0.5542**	0.6717	0.1092

4.4 Summary and Discussion

In this study, we proposed a new methodology that extracts the health features for power transformers using DGA data. The proposed methodology proceeds via two main steps: 1) data normalization to log-scale, 2) learning the semi-supervised autoencoder (SSAE) model for health feature extraction. SSAE can construct the health feature space that visualizes the monotonic behavior of health degradation. Thanks to the monotonous behavior of the health feature, workers can understand the health status of the transformer intuitively.

The model is validated through a comparative study that compared the

diagnostic performance of SSAE with conventional methods. Confusion matrix (PPV, TPR, and TNR) and F1 score are used as the performance evaluation index. The performance of the SSAE model is superior in most evaluation indices. Especially, the SSAE model can reduce the false alarm rate compared to other methods. In real industrial sites, frequent false alarms increase worker's fatigue and lose the confidence in the solution. The proposed SSAE helps workers perform active maintenance by accurately diagnosing the status of the power transformers.

The existing semi-supervised learning method solves the problem by additionally performing supervised learning to a pre-trained model through unsupervised learning. However, the method had a limitation that the pre-trained unsupervised model cannot extract the degradation features of the transformer. This is because the dissolved gasses are generated by various reasons including degradation of the transformer. Therefore, the model structure of proposed method was changed in such a way that unsupervised and supervised learning are not sequentially learned but simultaneously learned. In this way, the health features can be extracted effectively. An originality of the proposed methodology is that it can be applied to feature extraction with specific purposes regardless of the data domain.

Dissolved gases in insulating oil may be generated due to various causes, so there is a limitation in diagnosis accuracy of transformer using only DGA data. A further study is needed under the theme of developing a physical model for estimating the amount of dissolved gas generated by thermal decomposition of insulation. It is possible to develop a physical model for reaction enthalpy of insulation thermolysis and thermal and electrical energy required for thermolysis by using DGA, temperature and partial discharge data. If the estimated gas

concentrations and energy obtained from the physical model are additionally used as features of the training model, it will be able to construct a more accurate diagnostic model.

Chapter 5

Health Prognosis Model via XGBoost Regression

In this chapter, XGBoost regression-based health prognostic method is proposed to predict the health status and remaining useful life of power transformers. DGA-based diagnosis or detailed inspection of transformers are not frequently performed. Thus, safety accidents caused by faulty transformers often occur unexpectedly. To prevent this, interest in preventive maintenance of transformers has recently emerged. Subchapter 5.1 introduces the theoretical background of XGBoost, the basis of this study. Subchapter 5.2 proposes the new prognosis methodology consisting of two steps: 1) calculation of health index, and 2) construction of XGBoost model for health prognosis. Subchapter 5.3 describes the results of two case studies: 1) a comparison with other machine learning algorithms on prognostic performance, and 2) examination of prognostic performance according to the prediction period. At the conclusion, Subchapter 5.4 provides the summary and discussions of this study.

5.1 Backgrounds of XGBoost

T.Chen and C.Guestrin introduced XGBoost in 2016 for the first time, which is a scalable machine learning method for tree boosting system [47]. XGBoost is one of the ensemble techniques that uses a combination of weak decision trees, which weights the errors of weak prediction models and sequentially reflects them in the next learning model to create a strong prediction model. We will explain mathematically below how XGBoost learns data.

For a given data set with $n \times m$ (n samples, m features) dimensions, $D = \{(\mathbf{x}_i, y_i)\}$ $(|D| = n, \mathbf{x}_i \in \mathbf{R}^m, y_i \in \mathbf{R})$, a tree ensemble model using K additive functions is expressed as Eq. (23)

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \; f_k \in \mathcal{F} \tag{23}$$

where, $\mathcal{F} = \{f(\mathbf{x}) = w_{q(\mathbf{x})}\}$ $(q: \mathbf{R}^m \rightarrow T, w \in \mathbf{R}^T)$ is the space of regression trees. q is a function that maps a sample to the corresponding leaf index, and w is a leaf weight. T is the number of leaves. XGBoost uses a regularized loss function, Eq. (24), to train the trees.

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \; \Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{24}$$

where, l is a differentiable convex function which calculates the difference between prediction $\hat{y}_i$ and true y_i. And there is a regularization term Ω. This serves to learn that the structure of the model is simple and prevents overfitting. In order to find the best tree from the tree model and loss function, XGBoost adopts an additive manner that increases one branch in the tree for each iteration. The equation below is the loss function in the t-th iteration.

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t) + C \quad (25)$$

Gradient boosting is a method of weighting the parts that were not well learned in the previous iteration in the procedure of ensemble learning. Intuitively interpreting the Eq. (25), the loss function is minimized by adding the current prediction $f_t(\mathbf{x}_i)$ to the previous prediction $\hat{y}_i^{(t-1)}$. So, this explains that XGBoost is the one of Gradient Boosting methods and why is the name extreme gradient boosting (XGBoost).

To optimize the loss function, Eq. (25) can be approximated to Eq. (26) by second-order approximation with Taylor expansion.

$$L^{(t)} \cong \sum_{i=1}^{n} [l\left(y_i, \hat{y}_i^{(t-1)}\right) + l_i' f_t(\mathbf{x}_i) + \frac{1}{2} l_i'' f_t^2(\mathbf{x}_i)] + \Omega(f_t) \quad (26)$$

where, $l_i' = \frac{\partial}{\partial \hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $l_i'' = \frac{\partial}{\partial^2 \hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$, which are first and second order gradient of the loss function. Removing constant term, Eq. (27) becomes the final objective function in *t*-th iteration. As a result, since the objective function depends on l_i' and l_i'', we can optimize XGBoost by entering the first and second derivatives of loss function *l*.

$$\tilde{L}^{(t)} = \sum_{i=1}^{n} [\, l_i' f_t(\mathbf{x}_i) + \frac{1}{2} l_i'' f_t^2(\mathbf{x}_i)] + \Omega(f_t) \quad (27)$$

The goal of XGBoost is to find split points that allow the loss function to be reduced as much as possible. XGBoost can find the split points quickly by using approximate algorithm.

As a result, the advantages of XGBoost over other tree boosting algorithms are as follows: 1) fast execution time through parallelism, 2) overfitting regularization,

and 3) excellent predictive performance in classification and regression problems. Therefore, we want to build a health prediction model for power transformers using XGBoost.

5.2 XGBoost Regression Model for Health Prognosis

This Subchapter introduces the process of constructing a health prediction model for power transformers based on XGBoost regression. A regression model is created by learning the trend of the health index calculated from the HFS in Chapter 4. If the health indices at the past four points are given, it is possible to predict the health index at any future time with this regression model. The process of building a model consists of two main components: 1) health index calculation via orthogonal projection in HFS and 2) XGBoost training. Sections 5.2.1 and 5.2.2 will cover each process in detail.

5.2.1 Health Index Calculation via Orthogonal Projection (OP)

The health index has two significances: 1) quantification of the health status of the transformer and 2) application as a prognosis feature. Since the existing IEEE or IEC diagnostic methods determined the health status of the transformer by dividing it into several grades, it is difficult to distinguish the status if it is determined to be the same grade, even though the degree of degradation is different. Therefore, intuitive and detailed diagnosis is possible by numerically expressing the state of the transformer. In addition, the digitized health index can be utilized as input to the health prediction model of the transformer. If the health index has monotonous

behavior, the future status can be predicted by learning the historical trend of the health index. Therefore, we attempted to calculate the index from the health feature having the monotonous behavior. We suggest applying the concept of orthogonal projection (OP) to calculate the health index.

The proposed process of deriving a health index is described in Figure 5-1. It consists of three steps: 1) finding a linear decision boundary to distinguish between normal and failure using support vector machine (SVM), 2) finding a straight line perpendicular to the decision boundary, and 3) using the x-coordinate value of the axis-rotated coordinate system as the health index. In the process of obtaining the decision boundary, the decision boundary is assumed to be linear because it is actually nonlinear but almost linear. And during axis rotation, the rotation angle θ is the angle between the orthogonal line and the x-axis.

Although there are many different ways dimension reduction, the OP is proposed because it is the most effective way to prevent the reversal of the health index near the boundary between normal and failure. Here, the reversal of the health index means that the health index of a transformer in a relatively poor status is calculated as if it is in a better status.

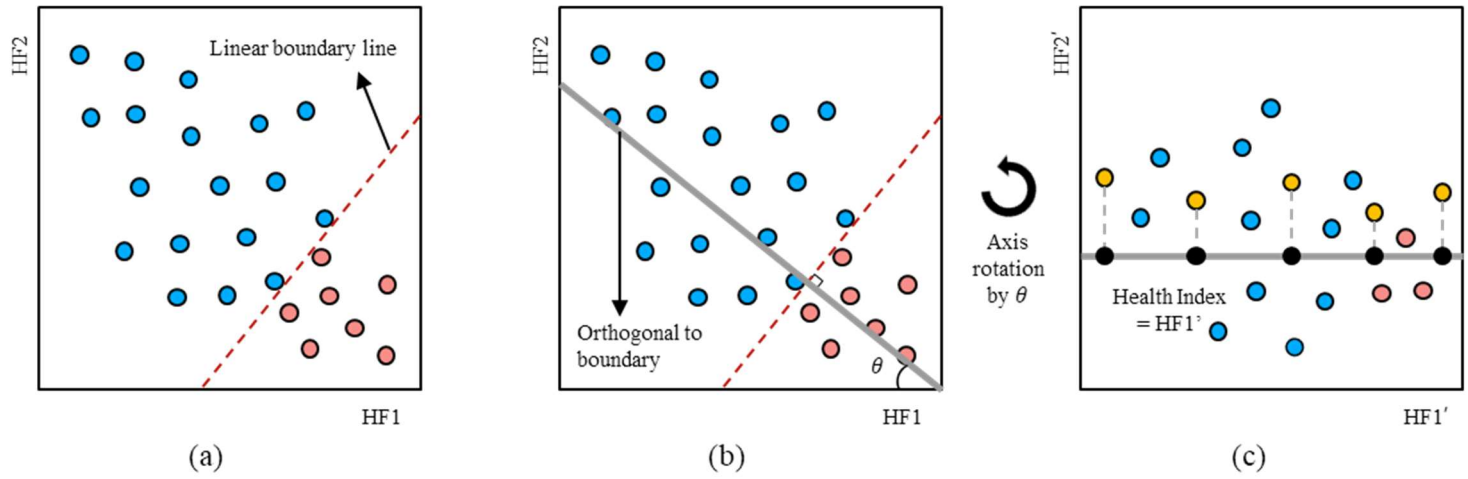
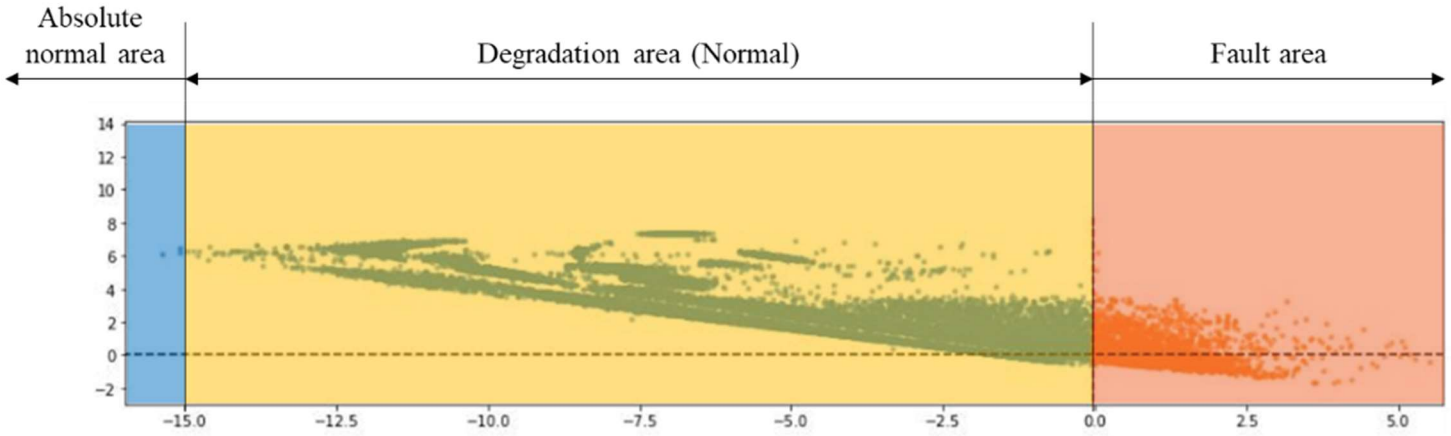


Figure 5-1 Process of deriving a health index in HFS: (a) finding a linear decision boundary using SVM, (b) finding a orthogonal line to the boundary, and (c) health index calculation through axis rotation by θ

5.2.2 Model Training based on XGBoost Regression

(1) Health Index Mapping

Prior to the point, note that this study aims to develop technologies that can be applied in real industrial sites. Therefore, we tried to express the health index in the range [0, 1] for an intuitive understanding. The health index developed in Subchapter 4 has a range of [-∞, ∞], where (-15, 0) is the degradation area. The degradation area is a status in which the transformer can be operated normally, and when the health index exceeds 0, a defect occurs, and repair or replacement is required. An area below -15 means an absolutely normal status. Therefore, the health index is proportionally mapped as shown in the Figure 5-2. The health index is designed to decrease from 1 to 0 as degradation progresses.



Status	Health index before mapping	Health index after mapping
Absolute normal	$x \leq -15$	1
Degradation	$-15 < x < 0$	$1 > x > 0$
Fault	$x \geq 0$	0

Figure 5-2 Health index mapping for intuitive understanding of degradation

(2) Definition of Input and Output Data

Based on the IEEE recommendations in Figure 5-3, it is recommended that the DGA sampling interval be taken differently depending on the condition of the transformer. Typically, DGA is measured on a daily interval for short and on a yearly interval for long. This irregularity in the data acquisition interval makes it difficult to predict the status of the transformer. Therefore, the goal of this study is to create a predictive model, robust to data sampling interval. To achieve the goal, we tried to define the proper input and output to learn the health trend of transformers. We defined the input and output as shown in Figure 5-4. The health index (HI_0, HI_1, HI_2, HI_3) at the past four times, the measurement interval (T_1-T_0, T_2-T_1, T_3-T_2), and the prediction interval (T_p-T_3) are set as inputs, and the health index (HI_p) at the prediction time is set as output. By defining in this way, the health of the transformer and time series information can be learned together. If the historical data and the prediction period are inputted, a future health index may be obtained. Four measurement data are used in consideration of the DGA data acquisition environment in real industrial sites. In order to adjust the scale of input data, the measurement interval and the prediction interval are also normalized by total four measurement periods.

	TDCG levels (μL/L)	TDCG rate (μL/L/day)	Sampling intervals and operating procedures for gas generation rates	
			Sampling interval	Operating procedures
Condition 4	>4630	>30	Daily	Consider removal from service. Advise manufacturer.
		10 to 30	Daily	
		<10	Weekly	Exercise extreme caution. Analyze for individual gases. Plan outage. Advise manufacturer.
Condition 3	1921 to 4630	>30	Weekly	Exercise extreme caution. Analyze for individual gases. Plan outage. Advise manufacturer.
		10 to 30	Weekly	
		<10	Monthly	
Condition 2	721 to 1920	>30	Monthly	Exercise caution. Analyze for individual gases. Determine load dependence.
		10 to 30	Monthly	
		<10	Quarterly	
Condition 1	≤720	>30	Monthly	Exercise caution. Analyze for individual gases. Determine load dependence.
		10 to 30	Quarterly	Continue normal operation.
		<10	Annual	

Figure 5-4 Recommendation for DGA sampling intervals by IEEE Std C57.104

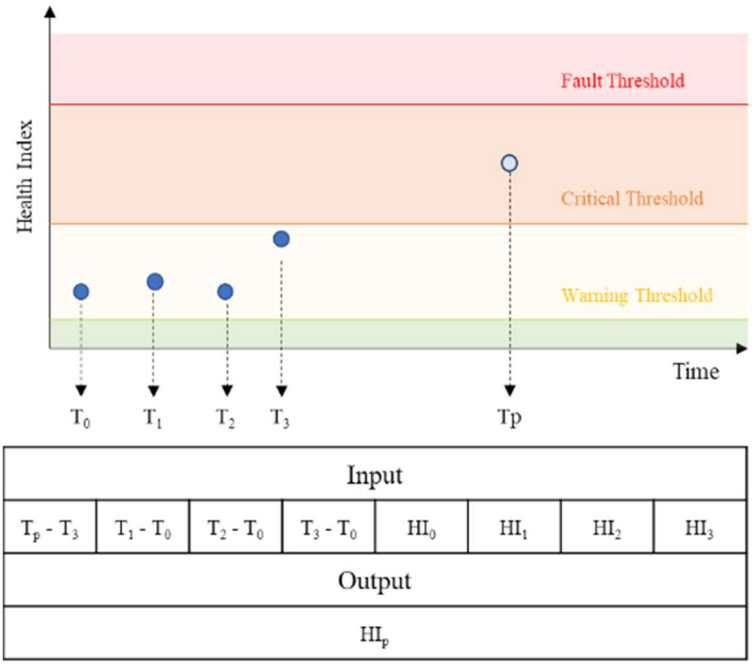


Figure 5-3 Definition of input and output for XGBoost learning

(3) XGBoost Design for Health Index Prediction

As shown in Figure 5-5, the XGBoost used in this study predicts the output as the sum of the results of all tree models for a given input data $\mathbf{x}_i = \{x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}, x_{i7}, x_{i8}\}$ and output data y_i. In this case, the model is learned to minimize the residual of the previous tree, and if the residual is no longer reduced, the tree is not generated. As explained in Subchapter 5.1, the structure of the tree is determined by pruning until the loss function is minimized. In other words, branches are created until the loss of the child branch exceeds the loss of the parent branch.

In this study, the XGBoost library provided by Python was used, and it was applied for regression analysis to solve the problem of predicting future health index. There are various hyperparameters in XGBoost: 1) general parameters for adjusting booster structure and computing power, 2) booster parameters for tree optimization and regularization, and 3) task parameters for setting objective functions and evaluation metrics. In this study, the main parameters were set as described in Table 5-1. The booster structure was set to 'gbtree' for tree-based learning, and the objective function and evaluation metric were set to 'reg:squarederror' and 'rmse', respectively, according to the purpose of regression analysis. XGBoost also has several tricks to prevent model overfitting, such as learning rate η that can control the influence of each tree model and regularization term γ and λ in Eq. (24). The optimal value of η was derived through parameter study. For γ and λ, since the scale of the input data is normalized between 0 and 1, the default of 1 was used as it is. Finally, you can limit the maximum number of trees. Even if the maximum number is not reached, learning is interrupted if the error is no longer improved. Since the number of trees

is correlated with η, the optimal combination was found through parameter study.

Table 5-1 Parameters in the proposed XGBoost

Parameter	Value
Booster	Gbtree
Objective function	Linear regressor (reg:squarederror)
Evaluation metric	Root mean squared error (rmse)
Learning rate (η)	0.1
Regularization 1 (γ)	1
Regularization 2 (λ)	1
# of estimators	10000

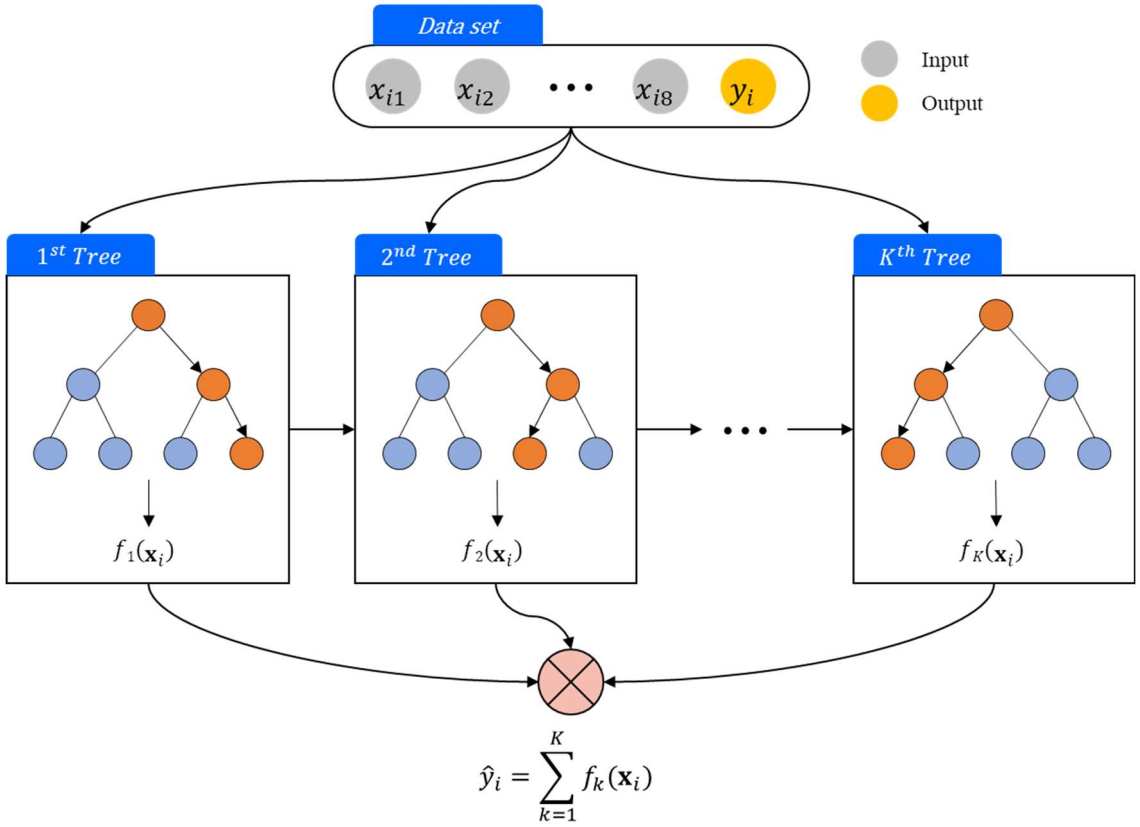


Figure 5-5 Structure of the proposed XGBoost

5.3 Performance Evaluation of XGBoost Regression

The proposed XGBoost-based prognosis model was verified by testing the actual industrial data provided by KEPCO. Two comparative studies were conducted. The first study compared the performance with the proposed algorithm and other tree structure algorithms, light GBM and random forest, and DNN which is based on the neural network. In this study, a parameter study was also performed simultaneously, and performance was evaluated on the optimal parameters. The second study compared the prognostic accuracy according to the prediction period. It was confirmed how much the prediction performance declined for a total period of 5 years on a yearly basis.

5.3.1 Description of Dataset

This study was conducted using DGA data provided by KEPCO. These data have been accumulated by operating more than 8,000 transformers in South Korea for more than 30 years. The amount of historical DGA data is enough to make a prediction model. A total of five historical data are required to learn the proposed XGBoost model, so transformer data less than five times are excluded. For given DGA data $\mathbf{X}^{(n)} = \{\mathbf{x}_1^{(n)}, \mathbf{x}_2^{(n)}, \dots, \mathbf{x}_t^{(n)}\}$ of the n-th transformer, the dataset was reconstructed as shown in Figure 5-6, to get as much data as possible. The total number of reconstructed datasets is 495,196, of which 85% is used for model learning and the remaining 15% is used for model validation. Then, features of input and output as defined in Section 5.2.1 are calculated from DGA data, and are used for model construction.

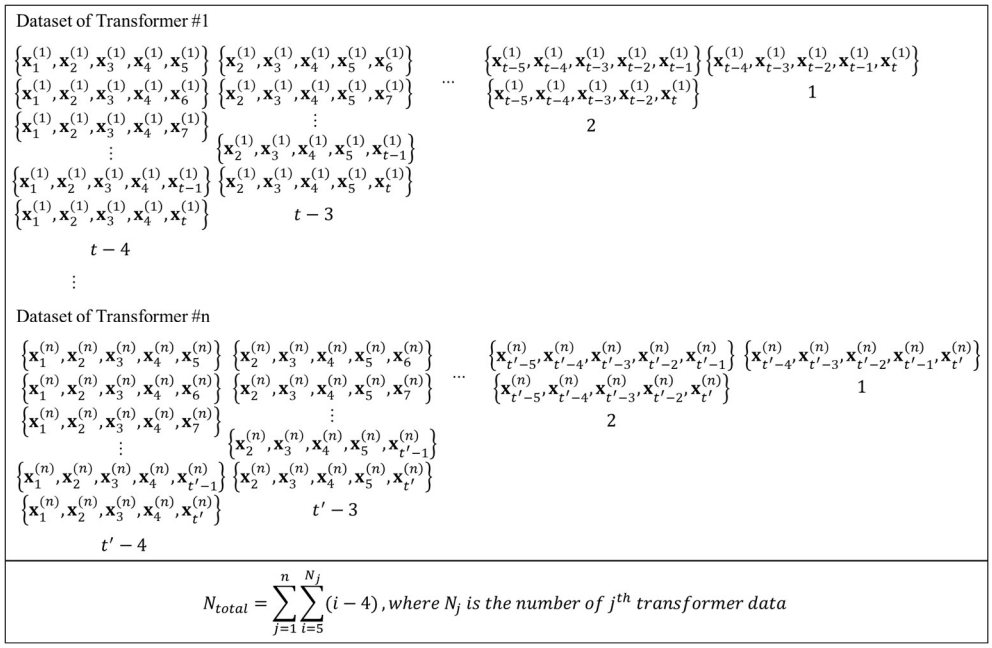


Figure 5-6 Dataset configuration for XGBoost regression

5.3.2 Prognosis Accuracy of XGBoost for Power Transformers

In general, the maintenance work is planned differently according to the health status of the transformer at the industrial site. Therefore, it is important to predict the health status of the transformer several months in advance. In this work, we evaluate the performance by focusing on how well the proposed XGBoost regression model predicts the actual health status. Here, the health status is classified into four levels based on the health index as shown in Table 5-2. The thresholds of health index by levels are derived based on the cluster of data in HFS obtained in Chapter 4. Back to the point, the accuracy calculation metric is defined as a ratio of the number of correctly predicted samples and the total number of samples, as shown in Eq. (28). The average accuracy is calculated through cross validation of seven folds.

$$Accuarcy = \frac{The\ number\ of\ correct\ prediction}{The\ total\ number\ of\ samples} \quad (28)$$

Table 5-2 Range of health index according to health status

Health status	Range of HI
Normal	$0.5 < HI \leq 1$
Warning	$0.2 < HI \leq 0.5$
Critical	$0 < HI \leq 0.2$
Fault	$HI = 0$

(1) Case Study 1: Comparison of Prognosis Accuracy with Other Algorithms

This case study compares the prognosis accuracy of the proposed XGBoost with other algorithms such as light GBM, random forest, and DNN. Light GBM and random forest are ensemble learning methods based on decision tree structure, like XGBoost. A typical gradient boosting method adopts a level-wise method that binary division is performed for pruning to achieve balance. On the other hand, light GBM adopts the leaf-wise method that pruning is carried out around less trained branches. This saves both time and memory for learning compared to XGBoost. Unlike XGBoost, random forest learns the model by adopting a bagging method. In the Boosting method, the next tree is learned to minimize the error of the previous tree, whereas in the bagging method, each tree is independent and has the same weight. The performance is generally lower than that of XGBoost, but the learning process is relatively simple. So, it is one of the most used algorithms. DNN is a neural network-based learning method that consists of several hidden layers between input and output layers. This is effective in modeling the nonlinear relationship between features based on various activation functions. However, due to the chronic problem of overfitting and high time complexity, care must be taken to find optimal activation

functions and learning parameters.

The highest accuracy for each algorithm is compared through a parameter study, and the results are summarized in Table 5-3. As can be seen in Table 5-3, the accuracy of XGBoost is the highest at about 87.1%. The degradation pattern varies greatly depending on the manufacturer, operating conditions, manufacturing year, and data sampling interval of a transformer, so I think that the performance of XGBoost, which is strong against overfitting problem, is the best.

Table 5-3 Comparison of prognosis accuracy (%) by training model and parameters

Model	Learning rate (η)	The number of estimators (trees)				
		1000	2000	3000	5000	10000
XGBoost	0.05	81.957	83.118	83.912	84.999	86.406
	0.1	83.068	84.502	85.384	86.283	**87.090**
	0.2	84.420	85.815	86.449	86.944	87.088
	0.3	84.967	86.122	86.534	86.853	86.854
Light GBM	0.05	81.207	82.304	83.130	84.272	85.799
	0.1	82.247	83.747	84.634	85.709	86.767
	0.2	83.616	85.015	85.721	86.491	**86.991**
	0.3	84.149	85.473	86.047	86.589	86.880
Random forest	-	77.708	77.745	77.748	**77.759**	77.752
DNN	**77.328** (N_{layers}: input layer (1), hidden layers (3), output layer (1) Layer depth: input layer (8), hidden layers (8), output layer (1) Activation function: elu)					

Light GBM is ranked second by a slight difference. To compare the performance of XGBoost and light GBM, case studies were additionally performed by varying the number of datasets. Table 5-4 summarizes the results of comparing the performance of XGBoost and light GBM according to the number of datasets. The number of estimator and learning rate were set to 10000 and 0.1, respectively. Table 5-4 shows that the smaller the number of datasets, the better the performance of XGBoost than light GBM. This is because the leaf-wise tree growth learning method of light GBM increases the likelihood of overfitting as the tree depth increases when the number of data is small. As expected, random forest is much less accurate than XGBoost. This is due to the limitations of the bagging method which is relatively weak for model optimization. DNN also has a low performance due to an overfitting problem. But also, it may be because the structure of DNN is simple in this study.

Table 5-4 Comparison of prognosis accuracy between XGBoost and light GBM according to the number of datasets

		The number of datasets				
		100	200	300	400	500
Prognosis accuracy (%)	XGBoost	70.00	72.50	74.00	76.25	80.20
	Light GBM	53.00	63.50	70.30	75.25	79.60

(2) Case Study 2: Performance Comparison According to Prediction Period

According to KEPCO's transformer maintenance manual, a normal inspection is conducted every three years and a detailed inspection is conducted every six years. It is important to predict and prevent failures that occur during the inspection period in advance. Therefore, the prognosis accuracies of XGBoost are compared as the prediction period increases from one to five. The number of datasets used for each

period is as follows: 1) 0~1 year: 3,855 sets, 2) 1~2 years: 4,694 sets, 3) 2~3 years: 4,024 sets, 4) 3~4 years: 3,398 sets, 5) 4~5 years: 3,015 sets.

The prognostic performances are described in Table 5-5. Prediction accuracy, precision, and recall values are summarized for each prediction period, and the overall average accuracy is emphasized in bold. As the predict period increases, the performance gradually decreases. Comparing the performance between 0~1 year and 4~5 years, a performance decrease about 4.5%. This is a logical result, because the longer the prediction period, the greater the uncertainty that affects the status of the transformer. In addition, 80% of accuracy of 4~5 years is considered sufficient for planning maintenance work in real industrial field.

In more detail, comparing the prognosis accuracy according to the health status, the accuracy of the normal grade is the highest and the warning and critical grades are relatively low. The range of health index in normal is the largest, and the range of health index in the other grades become gradually smaller. Therefore, the performance decreases further because the amount of learning data for warning or critical grades is smaller.

Table 5-5 Prognosis accuracy according to prediction period

Period	Data length	Performances
0~1 year	3,855	Normal / accuracy: 96.8131 / precision: 0.9681 / recall: 0.9644 Warning / accuracy: 82.1429 / precision: 0.8214 / recall: 0.7965 Critical / accuracy: 77.4141 / precision: 0.7741 / recall: 0.7870 Fault / accuracy: 82.4351 / precision: 0.8244 / recall: 0.8568 **total accuracy: 84.7013**
1~2 years	4,024	Normal / accuracy: 96.9590 / precision: 0.9696 / recall: 0.9536 Warning / accuracy: 77.1505 / precision: 0.7715 / recall: 0.7746 Critical / accuracy: 76.6212 / precision: 0.7662 / recall: 0.7728 Fault / accuracy: 84.0000 / precision: 0.8400 / recall: 0.9038 **total accuracy: 83.0249**
2~3 years	3,398	Normal / accuracy: 96.5031 / precision: 0.9650 / recall: 0.9448 Warning / accuracy: 76.7188 / precision: 0.7672 / recall: 0.7496 Critical / accuracy: 69.7674 / precision: 0.6977 / recall: 0.7544 Fault / accuracy: 81.9249 / precision: 0.8192 / recall: 0.8410 **total accuracy: 83.6827**
3~4 years	3,015	Normal / accuracy: 95.0863 / precision: 0.9509 / recall: 0.9366 Warning / accuracy: 72.9021 / precision: 0.7290 / recall: 0.7303 Critical / accuracy: 72.6562 / precision: 0.7266 / recall: 0.7223 Fault / accuracy: 80.2353 / precision: 0.8024 / recall: 0.8525 **total accuracy: 81.2285**
4~5 years	4,694	Normal / accuracy: 96.2420 / precision: 0.9624 / recall: 0.9595 Warning / accuracy: 78.0000 / precision: 0.7800 / recall: 0.7577 Critical / accuracy: 74.9216 / precision: 0.7492 / recall: 0.7399 Fault / accuracy: 82.9358 / precision: 0.8294 / recall: 0.8968 **total accuracy: 80.2200**

5.4 Summary and Discussion

Many studies have been conducted to diagnose the state of the transformer, but few studies have been conducted for prediction. Therefore, we proposed a new methodology that prognoses the health status of power transformers using DGA data in this study. The proposed methodology proceeds via two main steps: 1) calculation

of health index through orthogonal projection in the health feature space, and 2) learning the XGBoost regression model using time series health index trend. The model is validated through two case studies that examined a large amount of real industrial data.

The first study compares the prognostic performance with other machine learning algorithms, light GBM, random forest, and DNN. The predicted accuracy of the proposed XGBoost was the best at about 87.1%. It is considered that the proposed method is powerful for data overfitting compared to other algorithms. The second study examines the performance variations over prediction period. The performance gradually decreases as the prediction period increases. It is a reasonable result because the longer the prediction period, the more factors affecting the status of the transformer increase.

The proposed XGBoost-based prediction model has distinctive advantages as follows: 1) It is robust to irregular sampling intervals, and 2) it can avoid data overfitting by searching the optimal ensemble model from several random tree models. The prognosis model guarantees at least 80% accuracy over a period of up to 5 years. Therefore, we assured that the sudden safety accidents of transformers can be prevented, and it helps the workers plan the maintenance work substantially.

This study has an originality in that it is the first study to succeed in prognosing the status and RUL of the transformer by using a machine learning technique. Two previous studies have been attempted, but the performance was very low with an accuracy of 50% or less. However, the method developed in this study can predict the status of the transformer with a high accuracy of 85% or more. The machine

learning technique used in this study is XGBoost, which is generally used mainly for classification or regression problems. Given the excellent performance of the regression model of XGBoost, it was applied for the first time to prognose the transformer status. There is also a contribution that can be universally applied to prediction problems for time series data with irregular sampling rate.

However, there are limitations to further improvement of prognosis accuracy due to outliers such as sudden increase in gas concentration. This can be explained as two limitations of the data used in this study. One is a very low sampling rate, and the other is a very high variance. To overcome these limitations, further research is proposed as below.

A low sampling rate causes information losses about gradual degradation. Therefore, it is necessary to increase the sampling rate through online data acquisition rather than offline data acquisition. The sampling interval of offline data acquisition is about 3 months to 1 year, while the online data acquisition method has a much shorter sampling interval of 6 hours on average. However, the online method has a different data value from the offline method. This is because the online method has different gas concentration measurement mechanisms and requires additional calibration. Therefore, more research on the correlation analysis between online and offline measurements or online DGA-based transformer diagnosis is required.

If the variance of the data is high, the training model is likely to be overfit. Therefore, in this study, the XGBoost, the most appropriate technique for overfitting, was used, and the optimal hyperparameters were obtained through parameter studies. However, the fundamental solution is to reduce the variance of the data. The sudden

increase in gas concentration is closely related to the partial discharge of the transformer. In the gradual degradation process, the gas concentration gradually increases, but when a sudden event such as a partial discharge occurs, the gas concentration suddenly increases due to high electric energy. Therefore, Therefore, a study on the health prognosis of the transformer using the partial discharge signal will be conducted in future research. By adding a feature about the partial discharge to model training, it is possible to reduce the variance of the training data. This is because the cases in which gas concentration increases rapidly can be trained separately.

Chapter 6

Conclusion

6.1 Contributions and Significance

This doctoral dissertation provides a series of schemes for diagnosing and prognosing the health status of industrial power transformers based on deep learning techniques. Three major researches have been conducted to devise the new framework for preventive diagnosis of power transformers using dissolved gas analysis data: (1) iterative denoising autoencoder (IDAE) for data imputation, (2) semi-supervised autoencoder (SSAE) for health feature extraction, and (3) XGBoost regression for health prognosis of transformer. It is expected that the proposed research offers the following potential contributions.

Contribution 1: Improvement of data reliability through the restoration of DGA data

The gas concentration is often missing in the process of DGA due to worker's inexperience, which negatively affects transformer diagnostic accuracy. The first research proposes a iterative denoising autoencoder (IDAE) to impute the missing

values of dissolved gas analysis (DGA) data. Through the comparative studies, it was confirmed that the proposed IDEA could effectively estimate the original value for randomly generated missing gas concentrations. By recovering the DGA data through IDAE, the reliability of the data can be secured, which enables more accurate transformer diagnosis.

Contribution 2: Accurate and intuitive diagnosis of power transformer by health features with monotonous behavior

The second research proposes a semi-supervised autoencoder (SSAE) to extract representative health features for fault diagnosis of power transformers. A large amount of DGA data measured in real industrial site was used for this study. The proposed SSAE extracts two characteristic health features which have a highly linear correlation with the health status of power transformer. Through a comparative study, it was confirmed that the proposed SSAE could diagnose the transformers more accurate than conventional diagnostic methods. In addition, it is possible to construct the health feature space for intuitive diagnosis. Due to these advantages, the proposed SSAE is expected to be very useful for maintenance of the transformer in real industrial sites.

Contribution 3: Health prognosis of power transformer robust to irregular data sampling interval

Since DGA data is acquired based on the health status of the transformer, the sampling interval is irregular. Therefore, it is difficult to prognose the health status of the transformer with the existing prediction model based on sequential data. In order to overcome this, the third research proposes an XGBoost-based prediction model that can learn the measurement intervals and health indices simultaneously. Through the comparative studies, it was confirmed that the proposed method could accurately prognose the health status regardless of the sampling interval. The performance of the proposed method is also assured up to a prediction period of 5 years. This research is the first attempt for fault prognosis of transformer using a vast amount of industrial data, and it is meaningful in that it can be applied directly to the real industries.

6.2 Suggestions for Future Research

Although this doctoral dissertation achieves various technical advances in fault diagnosis and prognosis of power transformers, there are still several research topics for improvement. Specific suggestions for future research are described as follows:

Suggestion 1: Predictive diagnosis of power transformers using real-time online DGA sensors

In this study, the fault diagnosis and prognosis techniques of power transformer are proposed based on offline DGA data. Offline DGA data has some limitations for

fault diagnosis, such as missing data and irregular sampling intervals. In recently, to overcome these problems, online DGA sensors are being introduced at industrial sites. Although the measurement accuracy of online sensor is lower than the gas chromatography sensor in the laboratory, it is gradually improving. Therefore, it is necessary to prepare an online DGA-based transformer diagnosis study in advance, while leading to a study on the correlation analysis between online and offline DGA data.

Suggestion 2: Comprehensive fault diagnosis by multi-sensor fusion system

This study was conducted to diagnose transformers using only DGA data. However, DGA data only contains indirect information on health conditions, such as blood tests, in human terms. In fact, in the industrial field, data are acquired through various types of tests such as doble test, furan test, partial discharge test, and thermal image test as well as DGA. Therefore, for a more accurate and detailed diagnosis of the power transformer, it is necessary to utilize the various types of test data to identify the correlation between the data and comprehensively analyze the results. However, in the real industry, it is difficult to obtain these data because a sensor fusion system that can integrate the sensory data is not currently installed. Recently, as research on sensor fusion with digital transformation has become active, it is expected that sensor fusion technology will be introduced to substations and that the comprehensive diagnosis research on transformers using all sensor information will be activated.

References

[1] *IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers*, IEEE C57.104-2019, 2019.

[2] *Mineral Oil-Filled Electrical Equipment in Service. Guidance on the Interpretation of Dissolved and Free Gases Analysis [Electronic Resource] B.EN*, document 60599: 2016, 2016.

[3] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," *IEEE electrical Insulation magazine,* vol. 18, no. 3, pp. 8-17, 2002.

[4] E. Dornenburg and W. Strittmatter, "Monitoring oil-cooled transformers by gas-analysis," *Brown Boveri Review,* vol. 61, no. 5, pp. 238-247, 1974.

[5] R. Rogers, "IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis," *IEEE transactions on electrical insulation,* no. 5, pp. 349-354, 1978.

[6] S. A. Khan, M. D. Equbal, and T. Islam, "A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 22, no. 1, pp. 590-596, 2015.

[7] N. Poonnoy, C. Suwanasri, and T. Suwanasri, "Fuzzy Logic Approach to Dissolved Gas Analysis for Power Transformer Failure Index and Fault Identification," *Energies,* vol. 14, no. 1, p. 36, 2021.

[8] M. Žarković and Z. Stojković, "Analysis of artificial intelligence expert

systems for power transformer condition monitoring and diagnostics," *Electric Power Systems Research,* vol. 149, pp. 125-136, 2017.

[9] H. Zheng and R. Shioya, "A Comparison between Artificial Intelligence Method and Standard Diagnosis Methods for Power Transformer Dissolved Gas Analysis Using Two Public Databases," *IEEJ Transactions on Electrical and Electronic Engineering,* vol. 15, no. 9, pp. 1305-1311, 2020.

[10] L. Cheng and T. Yu, "Dissolved gas analysis principle-based intelligent approaches to fault diagnosis and decision making for large oil-immersed power transformers: A survey," *Energies,* vol. 11, no. 4, p. 913, 2018.

[11] S. S. Ghoneim, I. B. Taha, and N. I. Elkalashy, "Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 23, no. 3, pp. 1838-1845, 2016.

[12] A. Muthi, S. Sumarto, and W. Saputra, "Power transformer interruption analysis based on dissolved gas analysis (DGA) using artificial neural network," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 384, no. 1: IOP Publishing, p. 012073.

[13] M. D. Equbal, S. A. Khan, and T. Islam, "Transformer incipient fault diagnosis on the basis of energy-weighted DGA using an arti cial neural network," *Turkish Journal of Electrical Engineering & Computer Sciences,* vol. 26, no. 1, pp. 77-88, 2018.

[14] Y. Benmahamed, M. Teguar, and A. Boubakeur, "Diagnosis of power transformer oil using PSO-SVM and KNN classifiers," in *2018 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, 2018: IEEE, pp. 1-4.

[15] A. Dhini, I. Surjandari, A. Faqih, and B. Kusumoputro, "Intelligent fault diagnosis for power transformer based on DGA data using support vector machine (SVM)," in *2018 3rd International Conference on System Reliability and Safety (ICSRS)*, 2018: IEEE, pp. 294-298.

[16] H. Shang, J. Xu, Z. Zheng, B. Qi, and L. Zhang, "A novel fault diagnosis method for power transformer based on dissolved gas analysis using hypersphere multiclass support vector machine and improved D–S evidence theory," *Energies,* vol. 12, no. 20, p. 4017, 2019.

[17] B. Zeng, J. Guo, W. Zhu, Z. Xiao, F. Yuan, and S. Huang, "A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM," *Energies,* vol. 12, no. 21, p. 4170, 2019.

[18] J. Fang, H. Zheng, J. Liu, J. Zhao, Y. Zhang, and K. Wang, "A transformer fault diagnosis model using an optimal hybrid dissolved gas analysis features subset with improved social group optimization-support vector machine classifier," *Energies,* vol. 11, no. 8, p. 1922, 2018.

[19] K. Bacha, S. Souahlia, and M. Gossa, "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine," *Electric power systems research,* vol. 83, no. 1, pp. 73-79, 2012.

[20] S. Kim, J. Park, W. Kim, S.-H. Jo, and B. D. Youn, "Learning from even a weak teacher: Bridging rule-based Duval method and a deep neural network for power transformer fault diagnosis," *International Journal of Electrical Power & Energy Systems,* vol. 136, p. 107619, 2022.

[21] S. Kim *et al.*, "A Semi-Supervised Autoencoder With an Auxiliary Task (SAAT) for Power Transformer Fault Diagnosis Using Dissolved Gas Analysis," *IEEE Access,* vol. 8, pp. 178295-178310, 2020.

[22] L. Luo, X. Pei, and C. Shuai, "Convolutional Bi-directional Long Short Term Memory Network based Dynamic Fault Diagnosis for Transformer DGA," in *Journal of Physics: Conference Series*, 2021, vol. 1914, no. 1: IOP Publishing, p. 012045.

[23] D. Yang, J. Qin, Y. Pang, and T. Huang, "A Novel Double-Stacked Autoencoder for Power Transformers DGA Signals with Imbalanced Data Structure," *IEEE Transactions on Industrial Electronics,* 2021.

[24] Y.-J. L. Sun-Min Cho, Young-Sung Kim, Jae-Chul Kim, and Dong-Jin Kweon, "A study on Cause of Errors of Dissolved Gases Analysis in Transformer," in *Power Engineering Society Fall Conference of Korean Institute of Electrical Engineers*, 2006, pp. 141-143.

[25] J. J. Dukarm, "Data Issues in Transformer DGA," Delta-X Research Inc., Victoria BC Canada, 2017.

[26] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical systems and signal processing,* vol. 25, no. 5, pp. 1803-1836, 2011.

[27] K. Le Son, M. Fouladirad, A. Barros, E. Levrat, and B. Iung, "Remaining useful life estimation based on stochastic deterioration models: A comparative study," *Reliability Engineering & System Safety,* vol. 112, pp. 165-175, 2013.

[28] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation–a review on the statistical data driven approaches," *European journal of operational research,* vol. 213, no. 1, pp. 1-14, 2011.

[29] T. Wang, J. Yu, D. Siegel, and J. Lee, "A similarity-based prognostics approach for remaining useful life estimation of engineered systems," in

2008 international conference on prognostics and health management, 2008: IEEE, pp. 1-6.

[30] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering,* vol. 137, p. 106024, 2019.

[31] F. C. Sica, F. G. Guimarães, R. de Oliveira Duarte, and A. J. Reis, "A cognitive system for fault prognosis in power transformers," *Electric Power Systems Research,* vol. 127, pp. 109-117, 2015.

[32] J. Zarei, M. Shasadeghi, and A. Ramezani, "Fault prognosis in power transformers using adaptive-network-based fuzzy inference system," *Journal of intelligent & fuzzy systems,* vol. 26, no. 5, pp. 2577-2590, 2014.

[33] S. Chakravorti, D. Dey, and B. Chatterjee, "Recent trends in the condition monitoring of transformers," *Power Systems Springer-Verlag: London, UK,* 2013.

[34] O. Grechko and N. Kalacheva, "Current trends in the development of in-service monitoring and diagnostic systems for 110-750 kV power transformers (A survey)," *APPLIED ENERGY-NEW YORK-C/C OF IZVESTIIA-ROSSIISKAIA AKADEMIIA NAUK ENERGETIKA,* vol. 34, pp. 84-97, 1996.

[35] V. Sokolov, Z. Berler, and V. Rashkes, "Effective methods of assessment of insulation system conditions in power transformers: a view based on practical experience," in *Proceedings: Electrical Insulation Conference and Electrical Manufacturing and Coil Winding Conference (Cat. No. 99CH37035)*, 1999: IEEE, pp. 659-667.

[36] F. Jakob, P. Noble, and J. J. Dukarm, "A Thermodynamic Approach to Evaluation of the Severity of Transformer Faults," *IEEE Transactions on Power Delivery,* vol. 27, no. 2, pp. 554-559, 2012, doi: 10.1109/tpwrd.2011.2175950.

[37] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1096-1103.

[38] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in neural information processing systems*, 2012, pp. 341-349.

[39] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th international conference on data mining workshops (ICDMW)*, 2016: IEEE, pp. 241-246.

[40] K. Draszawka and J. Szymański, "Analysis of denoising autoencoder properties through misspelling correction task," in *International Conference on Computational Collective Intelligence*, 2017: Springer, pp. 438-447.

[41] G. Jiang, P. Xie, H. He, and J. Yan, "Wind turbine fault detection using a denoising autoencoder with temporal information," *IEEE/Asme transactions on mechatronics,* vol. 23, no. 1, pp. 89-100, 2017.

[42] M. B. Lazreg, M. Goodwin, and O.-C. Granmo, "Vector representation of non-standard spellings using dynamic time warping and a denoising autoencoder," in *2017 IEEE Congress on Evolutionary Computation (CEC)*, 2017: IEEE, pp. 1444-1450.

[43] Y. Jung, Y. Kim, Y. Choi, and H. Kim, "Joint Learning Using Denoising Variational Autoencoders for Voice Activity Detection," in *Interspeech*,

2018, pp. 1210-1214.

[44] Z. Chai, W. Song, H. Wang, and F. Liu, "A semi-supervised auto-encoder using label and sparse regularizations for classification," *Applied Soft Computing,* vol. 77, pp. 205-217, 2019.

[45] Y. Bengio, L. Yao, G. Alain, and P. Vincent, "Generalized denoising auto-encoders as generative models," *arXiv preprint arXiv:1305.6663,* 2013.

[46] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician,* vol. 46, no. 3, pp. 175-185, 1992.

[47] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.

[48] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems,* vol. 30, pp. 3146-3154, 2017.

[49] F. Tang and H. Ishwaran, "Random forest missing data algorithms," *Statistical Analysis and Data Mining: The ASA Data Science Journal,* vol. 10, no. 6, pp. 363-377, 2017.

[50] Q. Meng, D. Catchpoole, D. Skillicom, and P. J. Kennedy, "Relational autoencoder for feature extraction," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017: IEEE, pp. 364-371.

[51] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing,* vol. 95, pp. 187-204, 2017.

[52] S. Bustamante, M. Manana, A. Arroyo, R. Martinez, and A. Laso, "A

Methodology for the Calculation of Typical Gas Concentration Values and Sampling Intervals in the Power Transformers of a Distribution System Operator," *Energies,* vol. 13, no. 22, p. 5891, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/22/5891.

국문 초록

오염된 유중가스분석 데이터에 대한 딥러닝 기반 유입식 변압기 상태예측 연구

서울대학교 대학원
기계항공공학부
서 보 성

스마트 그리드, 에너지 저장 시스템, 전기자동차 등 에너지 시장의 가속화와 함께 안전하고 지속적인 전력 공급을 위한 신뢰성 높은 전력 계통에 대한 수요가 증가하고 있다. 이를 충족시키기 위해, 송·변전 시스템의 핵심 설비에 대한 진단 기법 및 예방정비에 대한 연구들이 많이 수행되고 있다. 그 중에서 특히 주변압기는 사용자의 목적에 맞게 전압을 변화시켜주는 장치로 전력 시스템의 중추적인 역할을 하고 있다. 따라서 주변압기 진단을 위한 다양한 검사 방법들이 개발되었으며, 유중가스분석법 (DGA: Dissolved Gas Analysis)이 가장 대표적인 방법이다. 유중가스분석법은 변압기의 결함에 의해 내부 절연물이 분해되면서 발생하는 가스농도를 측정하는 방법이다. IEEE와 IEC 등 다양한 전기전자 국제기구에서 수십년간 연구와 산업 경험을 통해 DGA

데이터 기반의 변압기 진단 표준을 수립하였다. 하지만 이러한 방법은 전문가의 경험과 해석에 근거하기 때문에 오진단율이 높다. 따라서 본 연구에서는 실제 산업 현장에서 취득한 대용량의 DGA 데이터를 사용하여 데이터 기반의 성능이 우수한 변압기 예측진단 방법을 개발하고자 하였다.

변압기 예측진단 성능을 개선하기 위해서 다음과 같이 해결해야할 세 가지 주요 이슈들이 존재한다: 1) DGA 데이터 결측 이슈, 2) 저차원 데이터에 대한 건전성 특성인자 추출 이슈, 그리고 3) 불규칙한 샘플링 주기에 대한 상태예측 이슈. 소개된 세 가지 이슈들을 해결하기 위해서 본 학위논문은 다음 세 가지 연구를 제안하였다.

첫 번째 연구는 Iterative Denoising Autoencoder (IDAE)를 사용한 다중 결측치 보정 방법을 제안하였다. 제안하는 방법은 Denoising Autoencoder (DAE)를 반복적으로 수행함으로써 결측치의 원본값을 복원할 수 있다. 노이즈를 최소화하려는 DAE의 성질을 이용하여 결측치를 노이즈로 인식하게 함으로써 결측치의 원본값을 추정하는 것이다. 제안하는 방법은 DGA 데이터의 신뢰도를 높임으로써 더 정확한 변압기 진단이 가능하게 만든다.

두 번째 연구는 Semi-supervised Autoencoder (SSAE)를 통한 건전성 특성인자 추출 방법을 제안하였다. 제안하는 방법은 데이터의 차원 축소와 변압기의 상태 학습을 동시에 수행함으로써 단조로운 열화 거동을 가지는 두 개의 특성인자들을 추출할 수 있다. 방대한 양의 산업 데이터를 학습하여 가스 농도 간 상관관계를 모델링하였기 때문에 기존의 진단 방법보다 더 정확한 진단이 가능하다. 또한, 건전성 특성인자로 이루어진 건전성 평면을 시각화 함으로써 직관적으로 열화 추세를 이해할 수 있다.

마지막으로, 세 번째 연구는 XGBoost 회귀 분석법을 통한 변압기 상태예측 방법을 제안하였다. 제안하는 방법은 트리기반의 앙상블 학습방법을 사용하여 불규칙한 시계열 데이터를 학습함으로써 샘플링 주기에 강건한 상태예측 모델을 구할 수 있다. 제안하는 방법은 다양한 모델을 순차적으로 나열하여 오차가 최소화되게 학습하기 때문에 과적합을 방지하고 정확하게 상태를 예측할 수 있다. 최대 5년까지 우수한 성능을 보장하기 때문에 변압기의 예방정비에 큰 도움이 될 것으로 기대된다.

제안하는 세 가지 방법을 연속된 과정으로 수행함으로써 변압기의 상태예측 프레임워크를 구축하는데 사용될 수 있다. 또한, 실산업에서 취득된 방대한 양의 데이터를 사용함으로써 산업에 바로 적용가능한 범용 모델을 개발했다는 점에서 의의가 있다.

주요어: 예측진단
결측치 보정
주변압기
딥러닝
머신러닝
유중가스분석

학 번: 2013-20680