공학박사학위논문

딥러닝 기반 고장 진단을 위한 정보 활용 극대화 기법 개발

Maximal Information Use for
Deep Learning Based Fault Diagnosis Techniques

2021 년 8 월

서울대학교 대학원

기계항공공학부

김 명 연

딥러닝 기반 고장 진단을 위한 정보 활용 극대화 기법 개발

Maximal Information Use for Deep Learning Based Fault Diagnosis Techniques

지도교수 윤 병 동

이 논문을 공학박사 학위논문으로 제출함

2021 년 4 월

서울대학교 대학원

기계항공공학부

김 명 연

김명연의 공학박사 학위논문을 인준함

2021 년 6 월

위 원 장 : 김 윤 영

부위원장 : 윤 병 동

위 원 : 김 도 년

위 원 : 조 규 진

위 원 : 김 흥 수

Abstract

Maximal Information Use for Deep Learning Based Fault Diagnosis Techniques

Myungyon Kim
Department of Mechanical and Aerospace Engineering
The Graduate School
Seoul National University

Unexpected failures of mechanical systems can lead to substantial social and financial losses in many industries. In order to detect and prevent sudden failures and to enhance the reliability of mechanical systems, significant research efforts have been made to develop data-driven fault diagnosis techniques. The purpose of fault diagnosis techniques is to detect and identify the occurrence of abnormal behaviors in the target mechanical systems as early as possible. Recently, deep learning (DL) based fault diagnosis approaches, including the convolutional neural network (CNN) method, have shown remarkable fault diagnosis performance, thanks to their autonomous feature learning ability.

Still, there are several issues that remain to be solved in the development of robust

and industry-applicable deep learning-based fault diagnosis techniques. First, by stacking the neural network architectures deeper, enriched hierarchical features can be learned, and therefore, improved performance can be achieved. However, due to inefficiency in the gradient information flow and overfitting problems, deeper models cannot be trained comprehensively. Next, to develop a fault diagnosis model with high performance, it is necessary to obtain sufficient labeled data. However, for mechanical systems that operate in real-world environments, it is not easy to obtain sufficient data and label information. Consequently, novel methods that address these issues should be developed to improve the performance of deep learning based fault diagnosis techniques.

This dissertation research investigated three research thrusts aimed toward maximizing the use of information to improve the performance of deep learning based fault diagnosis techniques, specifically: 1) study of the deep learning structure to enhance the gradient information flow within the architecture, 2) study of a robust and discriminative feature learning method under insufficient and noisy data conditions based on parameter transfer and triplet loss, and 3) investigation of a domain adaptation based fault diagnosis method that propagates the label information across different domains.

The first research thrust suggests an advanced CNN-based architecture to improve the gradient information flow within the deep learning model. By directly connecting the feature maps of different layers, the diagnosis model can be trained efficiently thanks to enhanced information flow. In addition, the dimension reduction module also can increase the training efficiency by significantly reducing the number of trainable parameters.

The second research thrust suggests a parameter transfer and metric learning based fault diagnosis method. The proposed approach facilitates robust and discriminative feature learning to enhance fault diagnosis performance under insufficient and noisy data conditions. The pre-trained model trained using abundant source domain data is transferred and used to develop a robust fault diagnosis method. Moreover, a semi-hard triplet loss function is adopted to learn the features with high separability, according to the class labels.

Finally, the last research thrust proposes a label information propagation strategy to increase the fault diagnosis performance in the unlabeled target domain. The label information obtained from the source domain is transferred and utilized for developing fault diagnosis methods in the target domain. Simultaneously, the newly devised semantic clustering loss is applied at multiple feature levels to learn discriminative, domain-invariant features. As a result, features that are not only semantically well-clustered but also domain-invariant can be effectively learned.

Keywords: Fault diagnosis
Deep learning (DL)
Convolutional neural network (CNN)
Transfer learning
Unsupervised domain adaptation (UDA)
Maximal information use

Student Number: 2015-22708

Table of Contents

List of Tables

List of Figures

Nomenclatures

DL	deep learning
B_t	mini-batch with cardinality $\|B_t\|$
ϵ	learning rate
CNN	convolutional neural network
UDA	unsupervised domain adaptation
PHM	prognostics and health management
SVM	support vector machine
LR	logistic regression
RF	random forest
KNN	k-nearest neighbors
ReLU	rectified linear unit
D_S	source domain
D_T	target domain
T_S	source task
T_T	target task
K_{strong}	strong stiffness axis
K_{weak}	weak stiffness axis
1D-CNN	one-dimensional convolutional neural network
2D-CNN	two-dimensional convolutional neural network
BN	batch normalization
NB	naïve Bayes
ELM	extreme learning machine
MLP	multi-layer perceptron
t-SNE	t-distributed Stochastic Neighbor Embedding

SNR	signal-to-noise ratio
dB	decibel
a	anchor
p	positive sample
n	negative sample
λ	balancing parameter
CORAL	correlation alignment
MMD	maximum mean discrepancy
DCTLN	deep convolutional transfer learning network
DANN	domain-adversarial neural network
L_C	classification loss
RKHS	reproducing kernel Hilbert space
L_D	domain-related loss
L_{SC}	semantic clustering loss
CWRU	case western reserve university
IMS	intelligent maintenance systems
XJTU-SY	xi'an jiaotong university- Changxing Sumyoung Technology Co., Ltd.
N	normal
B	ball fault
IR	inner raceway fault
OR	outer raceway fault
SCI	semantic clustering index

Chapter 1

Introduction

1.1 Motivation

Many types of mechanical components and systems are used in various industrial fields, such as the power generation, manufacturing, and transportation industries. Unexpected failures of mechanical systems can lead to substantial social and financial losses in many industries. In order to detect and prevent sudden failures and to enhance the reliability of mechanical systems, significant research efforts have been made to develop robust fault diagnosis techniques. The purpose of fault diagnosis techniques is to detect and identify the occurrence of abnormal behaviors of target mechanical systems as early as possible. As recent technological advances have made it possible to obtain a large amount of data from mechanical systems, data-driven fault diagnosis techniques have been attracting the interest of many researchers. However, conventional data-driven fault diagnosis methods have a big disadvantage, specifically, domain knowledge dependency problems. These drawbacks make it very difficult and inefficient to develop fault diagnosis methods using conventional approaches.

In order to overcome the aforementioned problems of conventional fault

diagnosis techniques, deep learning (DL) based fault diagnosis methods have recently drawn the interest of many researchers. Deep learning is a special type of machine learning technique that can autonomously learn optimal features from data. By stacking several neural network layers and training them by back-propagation methods, the proper feature representations for given tasks can be obtained. As a result, by using DL based diagnosis methods, expertise and domain knowledge dependency issues can be mitigated. Therefore, the time and cost required to establish fault diagnosis techniques can be reduced.

However, several issues remain to be solved to develop deep learning based fault diagnosis techniques for real-world applications. First, in the case of deep learning based approaches, the depth of representations is of central importance for many tasks. By making the neural network models deeper, enriched hierarchical features can be obtained through high-level abstractions of input data, and better performance can be achieved. Nevertheless, it becomes more difficult to comprehensively train deep learning models as the models become deeper. This is caused by the inefficient flow of gradient information and the overfitting problem due to a large number of learning parameters.

Next, there are data-related issues. In order to develop high-performance fault diagnosis methods using data-driven strategies, including deep learning, sufficient labeled data is required. Otherwise, it is impossible to obtain accurate and robust diagnosis methods. However, for mechanical systems that are operating in the real world, it is not easy to obtain sufficient datasets. In this case, it is difficult to learn an optimal diagnosis model, and the model may overfit to the insufficient data. Also, it is difficult to get label information for all systems and health states of interest.

Therefore, it is hard to train robust fault diagnosis models with high generalization performance through the use of conventional supervised learning schemes. Consequently, the three issues outlined above should be properly addressed to improve the performance of deep learning based fault diagnosis techniques.

1.2 Research Scope and Overview

The goal of this doctoral dissertation research is to develop methods that maximize the use of information in order to improve the performance of deep learning based fault diagnosis techniques. Three research thrusts were pursued toward this aim. First, a direct connection based convolutional neural network that enhances the gradient information flow was developed. Next, a robust and discriminative feature learning method for fault diagnosis was studied by transferring the pre-trained model and making the features better separated by their classes. Finally, a domain adaptation method based on the semantic clustering loss was proposed, to allow learning of more discriminative domain-invariant features. The three thrusts are briefly described below.

Research Thrust 1: Direct Connection Based Convolutional Neural Network (DC-CNN) for Fault Diagnosis

Research Thrust 1 proposes a direct connection based convolutional neural network (DC-CNN) to significantly improve the training efficiency and diagnosis performance of deep learning based techniques developed for fault diagnosis of

mechanical systems. By directly connecting the feature maps of different layers within the CNN architecture, we can maximize the gradient information flow and, therefore, train the deep neural networks efficiently. Simultaneously, to deal with the problems that can be caused by the increased number of parameters that arise due to the direct connections, dimension reductions not only in width and height but also in the depth-wise direction are conducted. The performance of the proposed DC-CNN-based fault diagnosis method was validated using experimental data from a rotor testbed. Comparison studies with other machine learning and deep learning based methods support that the proposed method can substantially improve the diagnosis performance in terms of accuracy and efficiency. In addition, the effectiveness of the proposed method was verified by conducting ablation studies and visualization analyses. The proposed method showed more stable and robust diagnosis performance under conditions of insufficient or noisy data, as compared with other existing methods.

Research Thrust 2: Robust and Discriminative Feature Learning for Fault Diagnosis Under Insufficient and Noisy Data Conditions

Research Thrust 2 proposes a robust and discriminative feature learning method to enhance fault diagnosis performance under insufficient and noisy data conditions. First, by transferring and adopting the pre-trained neural network model learned from the source domain with abundant data, reliable and robust feature learning for the target domain is possible. Then, discriminative features can be obtained based on the metric learning concept. In this research, a CNN model trained with a sufficient

image dataset is employed as the pre-trained model and semi-hard triplet loss is used to learn semantically well-separated features. The superior performance of the proposed method was verified using experimental data obtained from a rotor testbed. Based on comparisons with other deep learning based fault diagnosis methods, outstanding diagnosis performance of the proposed method, under both insufficient data conditions and noisy data conditions, was confirmed. Further, the visualization results of the learned features demonstrate the robust and discriminative characteristics of the features learned by the proposed method. These results show that the proposed approach can enhance fault diagnosis performance by transferring the model parameters learned from the source domain with abundant data and taking into account the class-wise distances between samples.

Research Thrust 3: A Domain Adaptation with Semantic Clustering (DASC) Method for Fault Diagnosis

Research Thrust 3 proposes a novel domain adaptation based fault diagnosis technique to improve diagnosis performance for the unlabeled target domain by using the label information obtained from the source domain. The DASC method aims to learn discriminative and domain-invariant features that both minimize domain discrepancy between the source and target domains and also make the samples from each class semantically well-clustered. This is achieved by proposing the semantic clustering loss, which brings samples that have the same class label closer and causes differently labeled samples to separate. Furthermore, by applying this loss at multiple feature levels, more robust features with desired properties are

obtained. The effectiveness of the DASC approach was validated via various analyses that examined experimental data from three bearing systems. The results indicate that the DASC approach significantly increases generalized diagnosis performance for mechanical systems, as compared with existing approaches. Further, the results of visualization of the learned feature distributions show that the DASC method can obtain discriminative features with better clustering characteristics. The proposed method's efficacy was also confirmed to a further degree through ablation studies. In addition, by defining an index that evaluates how well the target features are clustered semantically, DASC's ability to make target domain features well-clustered class-wise was verified. These results confirm that the proposed DASC approach is able to greatly increase the fault diagnosis performance for target domain systems via transferring and using the label information obtained from the source domain.

1.3 Dissertation Layout

This doctoral dissertation is organized as follows. Chapter 2 provides a technical background and literature review of fault diagnosis techniques, transfer learning, and metric learning concepts. Chapter 3 suggests a direct connection based convolutional neural network (DC-CNN) for fault diagnosis. In Chapter 4, robust and discriminative feature learning for fault diagnosis under insufficient and noisy data conditions, based on the parameter transfer and metric learning, is described. Then, Chapter 5 introduces a domain adaptation with semantic clustering (DASC) method for fault diagnosis. Finally, Chapter 6 concludes the doctoral dissertation by

summarizing the contributions of this work and suggesting future research.

Chapter 2

Technical Background and Literature Review

In this doctoral research, fault diagnosis techniques using deep learning methods were studied. In particular, various studies were conducted to improve the performance and efficiency of deep learning based fault diagnosis techniques. Before explaining the details of the research, this chapter provides the technical background and literature review necessary to understand the research described in this dissertation. Specifically, in Chapter 2.1, the concept of fault diagnosis techniques for mechanical systems is described. In Chapter 2.2, the concept and types of transfer learning are explained. Then, an overview of the metric learning concept is provided in Chapter 2.3. Finally, a summary and discussion are outlined in Chapter 2.4.

2.1 Fault Diagnosis Techniques for Mechanical Systems

Many types of mechanical components and systems are used in various industrial fields, including the power generation, manufacturing, and transportation industries. As mechanical systems are employed in more and more areas, the number of failures can increase. In addition, as these kinds of mechanical systems are used over long periods of time, the systems deteriorate and, therefore, the probability of

failure may increase. Unexpected failures of mechanical systems can happen for countless reasons and can lead to significant social and economic losses. In order to prevent unanticipated failures and to improve the reliability of mechanical systems, significant research efforts have been made to develop prognostics and health management (PHM) techniques. PHM is a framework that provides comprehensive maintenance and health risk management strategies that facilitate industrial asset management by enhancing the quality, safety, availability, and productivity of mechanical components and systems. An overview of PHM is presented in Figure 2-1. PHM is composed of fault diagnosis, prognostics, and health management. By focusing on improving these elements of PHM, this doctoral dissertation aims to develop fault diagnosis techniques that identify the health states of mechanical systems correctly and efficiently. This section is devoted to providing a review of fault diagnosis techniques, from conventional to recent approaches.

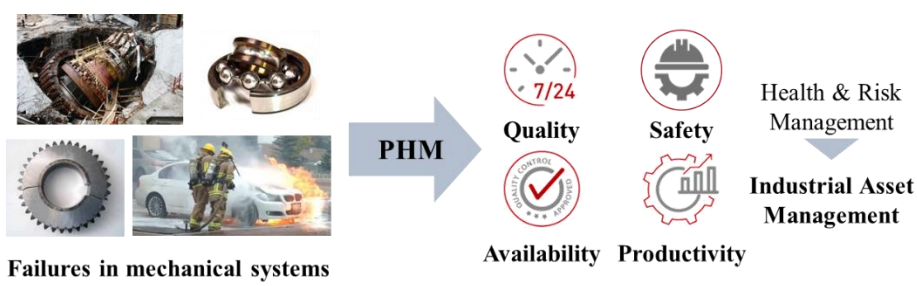


Figure 2-1 Prognostics and health management (PHM).

2.1.1 Fault Diagnosis Techniques

Mechanical systems can be vulnerable in the presence of uncertainties and health degradation; these factors can cause unexpected failures with huge associated economic losses. In order to prevent failures, fault diagnosis techniques for mechanical systems are becoming increasingly important. The goal of fault diagnosis techniques is to detect and recognize deterioration in the health states of the target mechanical systems. In general, fault diagnosis techniques can be categorized into three groups of methods; model-based, data-driven, and hybrid [1], [2]. In model-based schemes, fault diagnosis is performed based on physical or mathematical models that can appropriately represent the health states of the mechanical system of interest [3], [4]. In contrast, data-driven schemes rely on sensor signals measured from the target mechanical systems. Hybrid methods use both schemes simultaneously to leverage the advantages of each method [5]. Since it is almost impossible to develop accurate analytical models for complex mechanical systems, data-driven methods have recently drawn attention from many researchers. In addition, in the era of the fourth industrial revolution, the importance of data-driven diagnosis schemes is growing rapidly due to developments in data-related technologies, such as sensor technology, big-data computing technology, and algorithms to deal with large amounts of sensor data.

As shown in Figure 2-2, conventional data-driven fault diagnosis methods include a rule-based method and a health feature based method. First, the rule-based method is an approach that diagnoses the health states of mechanical systems based on human-created or curated rule sets. For this method, it is essential to devise proper rules that well-distinguish the health states of the target systems. Second, as shown

in Figure 2-3, the health feature based fault diagnosis method consists of four basic steps: (1) data acquisition, (2) data preprocessing, (3) feature engineering, and (4) health classification [6]. For fault diagnosis of mechanical systems, vibration signals are dominantly used because they contain a great deal of information about mechanical systems' physical behaviors [7]–[9]. To extract critical information that can well represent the health states of target systems, from raw input data, data preprocessing and feature engineering steps are carried out. Lastly, based on the extracted features, diagnosis models are developed using several statistical learning methods, as can be seen in Figure 2-4. Then, through these developed models, fault diagnosis can be conducted for datasets obtained from the target mechanical systems.

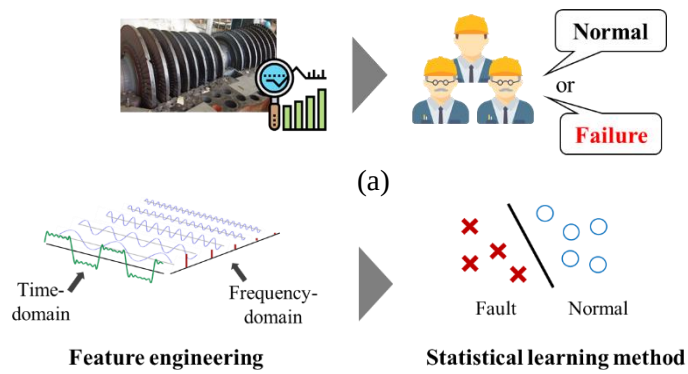


Figure 2-2 Conventional fault diagnosis methods: (a) rule-based method; (b) health feature based method.

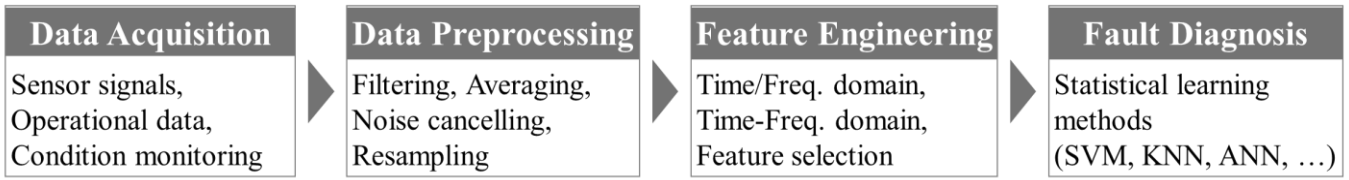


Figure 2-3 General procedure of the conventional health feature based fault diagnosis method.

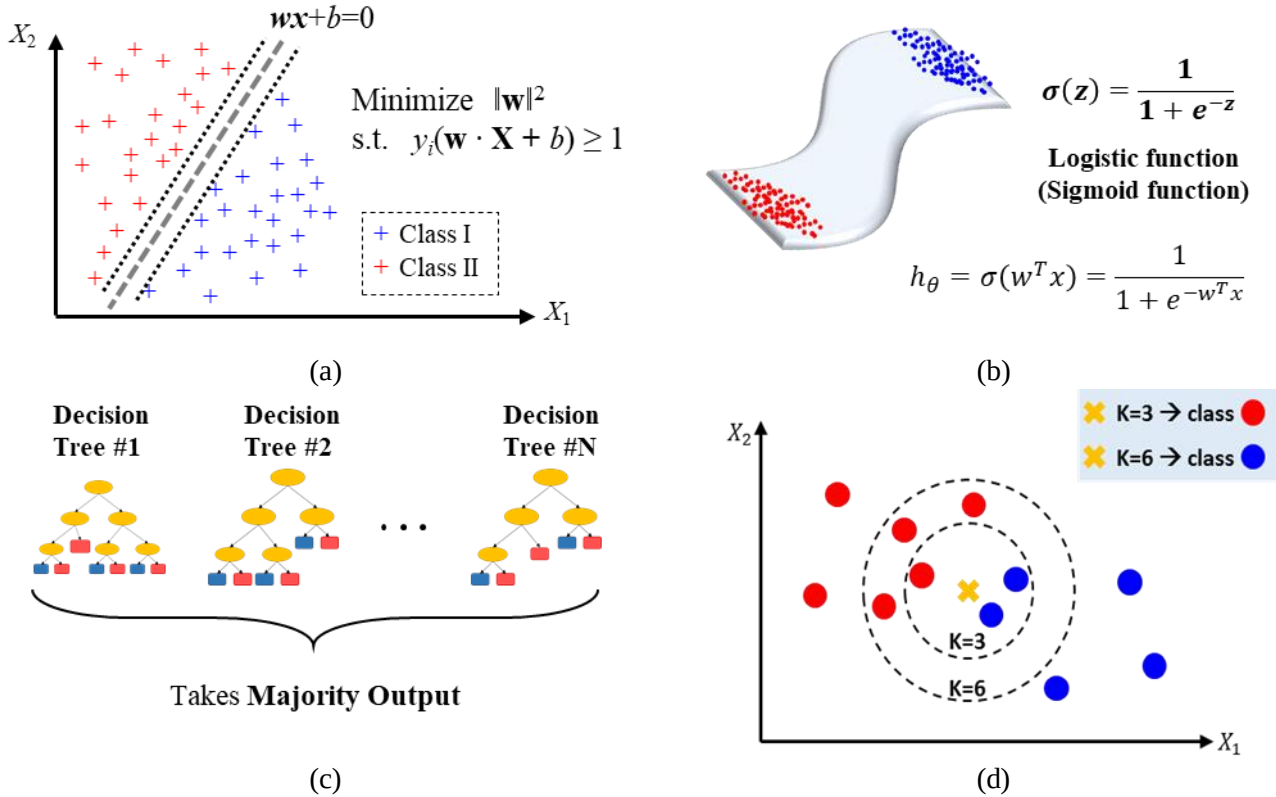


Figure 2-4 Examples of statistical learning methods for fault diagnosis: (a) Support vector machine (SVM); (b) Logistic regression (LR); (c) Random forest (RF); and (d) K-nearest neighbors (KNN).

For conventional data-driven diagnosis techniques, however, there are several problems that make it difficult to develop accurate diagnosis models and that hinder the improvement of diagnosis performance. The biggest problem in conventional fault diagnosis techniques is that a lot of domain knowledge and expertise are required. In the case of rule-based methods, a great deal of physical and mechanical knowledge is necessary to design optimal rule sets capable of distinguishing all states of interest. For health feature based diagnosis methods, a large amount of domain knowledge is required, especially in the signal processing and feature engineering steps. In these steps, experts' knowledge and experiences play a very important role to devise suitable features that can properly indicate the current health state of the mechanical system [10], [11]. Among various candidates, proper techniques should be selected based on physical knowledge related to target systems and obtained signals. Unless the proper features are designed and extracted, the performance of the fault diagnosis method cannot be guaranteed. Further, even if appropriate features are designed for a specific target system or health state, those features may not be suitable for different tasks. This means that for different target systems and states, different feature engineering approaches must be designed [12]. Consequently, it is time-consuming and expensive to develop conventional rule-based or health feature based fault diagnosis methods. In addition, conventional fault diagnosis methods have a disadvantage in that it is difficult for these techniques to be universally employed for various target systems. These drawbacks make it very hard and inefficient to develop fault diagnosis methods in conventional ways.

2.1.2 Deep Learning Based Fault Diagnosis Techniques

Recently, to overcome the aforementioned problems of conventional fault diagnosis techniques, deep learning (DL) based fault diagnosis methods have drawn the attention of many researchers [13], [14]. Deep learning is a special type of machine learning method that consists of multiple layers of neural networks, as shown in Figure 2-5; these networks are mathematical models inspired by the connectivity patterns of biological neural networks. The classification results from the DL-based classification model can be expressed as follows:

$$\begin{aligned} &P(y = c|x, w, b) \\ &= f_l(f_{l-1}(\cdots f_2(f_1(x; w_1, b_1); w_2, b_2)\cdots; w_{l-1}, b_{l-1})\cdots; w_l, b_l) \end{aligned} \tag{2.1}$$

where x denotes input data; w_l and b_l denote weight and bias at the l^{th} layer to be trained; y denotes output value; c denotes class; f_l denotes nonlinear functions at the l^{th} layer; and $P(y = c|x, w, b)$ denotes the probability that input data x belongs to class c, based on the model with trainable parameters *w* and *b*. DL models are trained to learn optimal weight and bias values to discover the proper features that work well on target tasks. Through the use of back-propagation methods, the error gradient with respect to each parameter is obtained, which implies the influence of each parameter on the final loss. Then, using a mini-batch gradient descent algorithm, parameters can be optimized in order to minimize the target loss function. The mini-batch gradient descent algorithm can be expressed as follows:

$$\theta_{t+1} \leftarrow \theta_t - \epsilon_t |B_t|^{-1} \sum_{i \in B_t} \nabla_\theta L(x_i; \theta_t) \tag{2.2}$$

where x_i denotes the i[th] input data; θ_t denotes the parameters, including weights and biases at step t; B_t denotes the mini-batch with cardinality $|B_t|$; $\nabla_\theta L$

denotes the gradient of cost function L with respect to θ; and ϵ_t denotes the learning rate at step t. Based on the back-propagation and gradient descent algorithms, DL can automatically learn optimal features for given tasks from the input dataset. Thanks to this autonomous feature learning capability, several advantages can be gained by adopting DL-based fault diagnosis methods, as presented in Figure 2-6. First, domain knowledge dependency problems existing in conventional diagnosis methods can be alleviated by using DL. As a result, the time and cost required to develop the fault diagnosis technique can be reduced. In addition, DL also enables realization of an end-to-end learning strategy to learn the feature extractor part and classifier part at the same time.

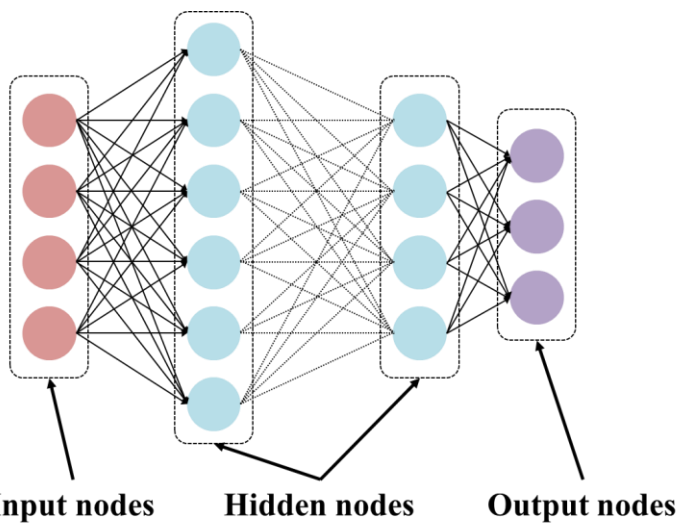


Figure 2-5 Schematic diagram of a deep neural network.

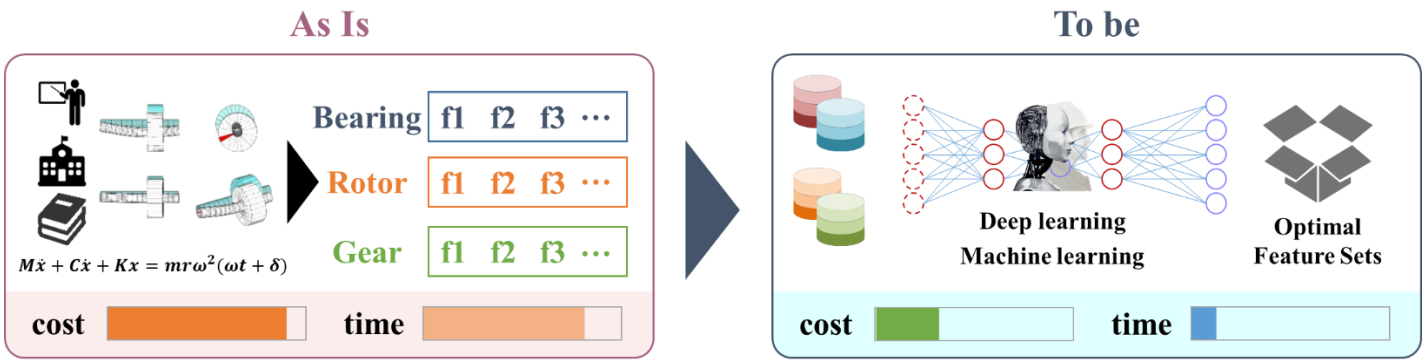


Figure 2-6 Advantages of deep learning based fault diagnosis techniques.

Among many DL techniques, the convolutional neural network (CNN) approach has been widely used in many fields, thanks to its advantages, which include local connectivity, parameter sharing, and the ability to consider high-dimensional information within the input data [15]–[17]. Figure 2-7 shows the structure of the CNN method. A CNN consists of combinations of multiple layers, including convolutional layers and pooling layers. In contrast to conventional fully connected neural networks, a CNN uses weight matrices with dimensions smaller than those of input data, which are called kernels or filters. Therefore, a CNN has a local connectivity characteristic that allows it to learn local patterns within small regions of input data [15]. Furthermore, a CNN uses multiple dimensional kernels to extract features, and therefore, it can consider high-dimensional information within the input data. In the convolutional layers, by sliding kernels within the input data and conducting convolution operations, output layers called feature maps, whose depth is the same as the number of kernels, are produced. Based on those convolution operations, features are extracted from the input data. In addition, a CNN uses kernels with the same weight values for these convolution operations within the entire input data; this is referred to as parameter sharing [16]. As a result, a CNN has the ability to significantly reduce the number of parameters to be trained and increase computational efficiency. Depending on the dimension of the input data, several types of convolutional layers can be used. In general, one-dimensional CNN (1D-CNN) using 1D kernels and two-dimensional CNN (2D-CNN) using 2D kernels are used. The pooling layers pool out specific values, (i.e., maximum or average values) from the sub-regions; thus, feature maps from previous layers are downsized. Moreover, like other conventional neural networks, nonlinear activation functions – for example, rectified linear unit (ReLU) – are used in order to learn nonlinear

features. Lastly, a fully connected layer acting as a classifier follows. By stacking several combinations of those components, various CNN architectures can be achieved [18]–[21]. Due to its advantages explained above, the CNN approach has been extensively used in various research areas, including face recognition, disease diagnosis, and natural language processing [22]–[24].

Recently, due to the merits of the CNN approach, CNNs have also been actively used in the field of machine fault detection and diagnosis. In order to deal with one-dimensional raw sensor signals, such as time-series vibration signals, 1D-CNN has been employed. Wu et al. [34] used 1D-CNN for the diagnosis of gearbox systems and achieved better diagnosis accuracy, as compared to other methods that are based on traditional signal demodulation methods. Jiang et al. [35] proposed a 1D-CNN-based fault diagnosis method for a wind turbine gearbox that can consider the multiscale characteristics inherent in vibration signals. Kim et al. [17] proposed a 1D-CNN-based parameter repurposing method for fault diagnosis of rolling element bearings with small datasets.

2D-CNN can be used to consider multiple dimensional information and correlations within 2D input data. Wen et al. [25] proposed a CNN-based, data-driven fault diagnosis method for bearing and pump systems. By using the CNN approach, along with the signal-to-image conversion method, these researchers achieved significant improvements in diagnosis performance. Liu et al. [26] used a 2D-CNN-based model for fault diagnosis of motor bearings and centrifugal pumps; the approach extracts fault features from images constructed by continuous wavelet transform. Maraaba et al. [27] developed a fault diagnosis method for permanent magnet synchronous motors based on 2D-CNN. In the Maraaba et al. work, 2D

matrices, which are composed of three-phase, steady-state motor currents, are adopted as the input data. The results show that high diagnosis accuracy can be obtained without the need for a manual feature extraction phase. In addition to the aforementioned papers, many studies have been conducted to develop fault diagnosis techniques using CNN [28]–[30].

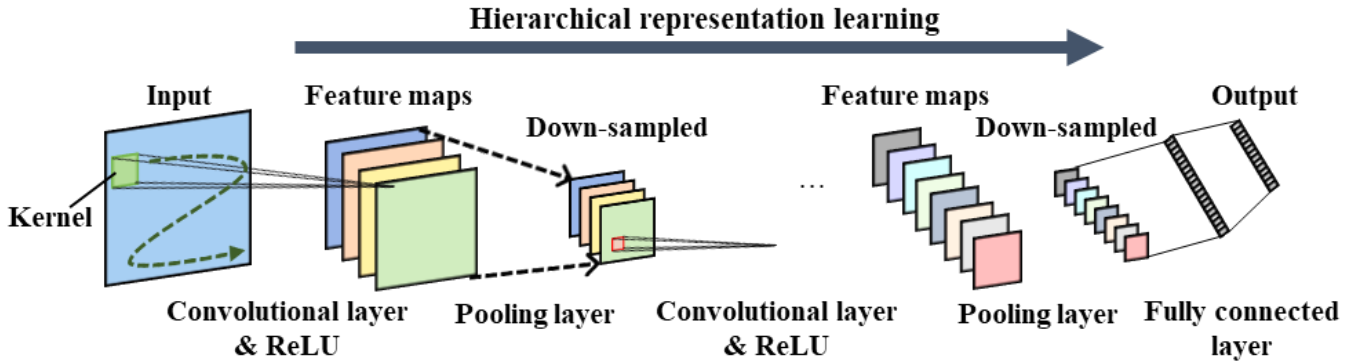


Figure 2-7 Schematic diagram of a convolutional neural network.

2.2 Transfer Learning

In order to develop robust fault diagnosis methods using data-driven strategies, including deep learning, sufficient labeled datasets must be obtained from the target system for every health condition of interest [31], [32]. Depending on the quantity and quality of data used when learning the diagnosis models, there may be a large difference in the generalization performance of the developed diagnosis models. In addition, for health states with extremely little data, it may be impossible to learn an accurate diagnosis model that can correctly diagnose those health states. Further, it is useful to point out that data-driven methods offer satisfactory performance when both the training and test data can be assumed to share a distribution of the same type [33], [34]. However, for complex mechanical systems that are operating in the real world, it is challenging to secure sufficient labeled datasets. Moreover, in many cases, the training and test data exhibit different types of distribution for a variety of reasons, including environmental influences or changing operating conditions. As a result, in the real world, it is hard to train robust fault diagnosis models with high generalization performance through the use of conventional supervised learning schemes.

On the other hand, transfer learning is a learning strategy that can be used when there is insufficient data or when the distributions of the training data and test data are different. Transfer learning is a technique that utilizes the knowledge and information obtained from one domain or task to solve problems in other domains or tasks [35]. In other words, it refers to a learning strategy that uses knowledge obtained from the source domain (D_S) to construct a diagnosis model $P(Y_T|X_T)$ that determines the health state of the mechanical system in the target domain (D_T). As

shown in Figure 2-8, through the transfer learning scheme, knowledge learned from the source domain, where it is relatively easy to obtain sufficient data, can be used to construct a fault diagnosis model in the target domain, where it is difficult to obtain data. As a result, it is possible to efficiently learn a robust fault diagnosis model by overcoming the problems that we encounter when using conventional supervised learning strategies under insufficient data conditions. As can be seen in Figure 2-9, the benefits of using transfer learning are as follows. First, because the information or knowledge learned from the source domain is used, learning starts with an already improved performance. Next, learning can proceed faster, and ultimately, better target performance can be obtained. In summary, by using a transfer learning strategy, we can save the time required to develop the diagnosis techniques, and we can get better diagnosis models with improved performance.

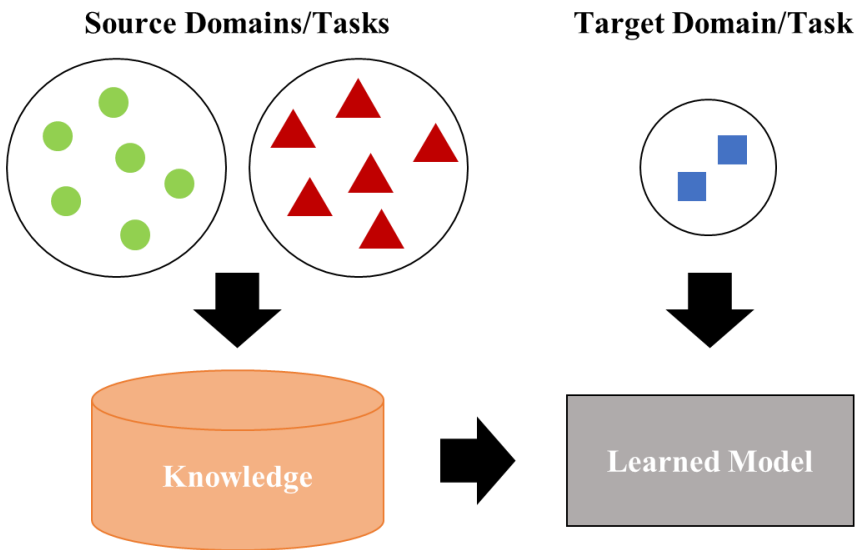


Figure 2-8 Learning process of transfer learning.

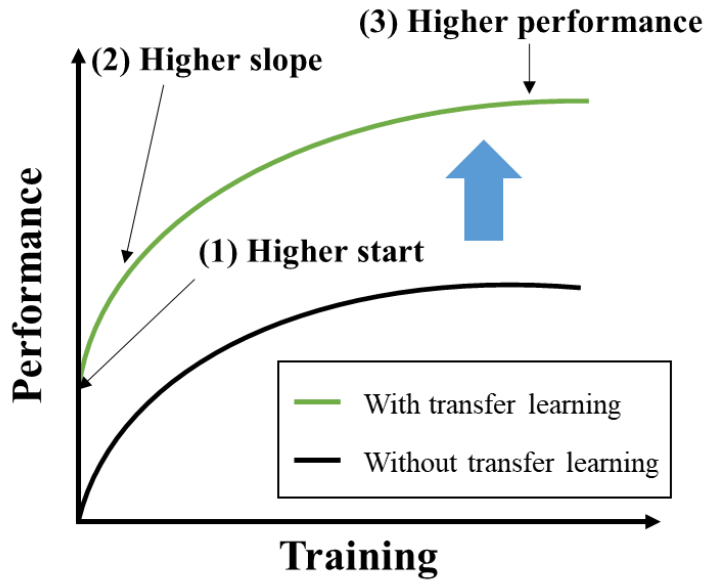


Figure 2-9 Three possible benefits obtained by using transfer learning during the learning process.

There are various learning settings, based on the relationship or similarity between the domains and tasks in a given problem situation, as shown in Table 2-1 [35]. First, for conventional machine learning problem settings, which are the types of settings we generally encountered, the source and target have the same domains and tasks. On the other hand, inductive transfer learning means a setting where the source task and the target task are different; Transductive transfer learning means a problem setting with the same source and target tasks, however, different source and target domains. Lastly, a setting where both domains and tasks are different is called unsupervised transfer learning.

Table 2-1 Various types of learning settings.

Learning setting	Source and target domains ($\boldsymbol{D_S}$ & $\boldsymbol{D_T}$)	Source and target tasks ($\boldsymbol{T_S}$ & $\boldsymbol{T_T}$)
Traditional machine learning	Same	Same
Inductive transfer learning	Same	Different, but related
Transductive transfer learning	Different, but related	Same
Unsupervised transfer learning	Different, but related	Different, but related

Transfer learning can be classified into several categories of approaches, in accordance with which part of the knowledge or information is transferred across domains or tasks, as shown in Figure 2-10. The instance transfer learning approach refers to a method of reusing some part of the source domain data in the target domain. Next, the parameter transfer approach is a method of transferring information by discovering and transferring parameters that can be shared in two domains. Finally, the feature representation transfer approach is a learning strategy in which information can be transferred by learning and sharing common feature representations that can be shared in two domains.

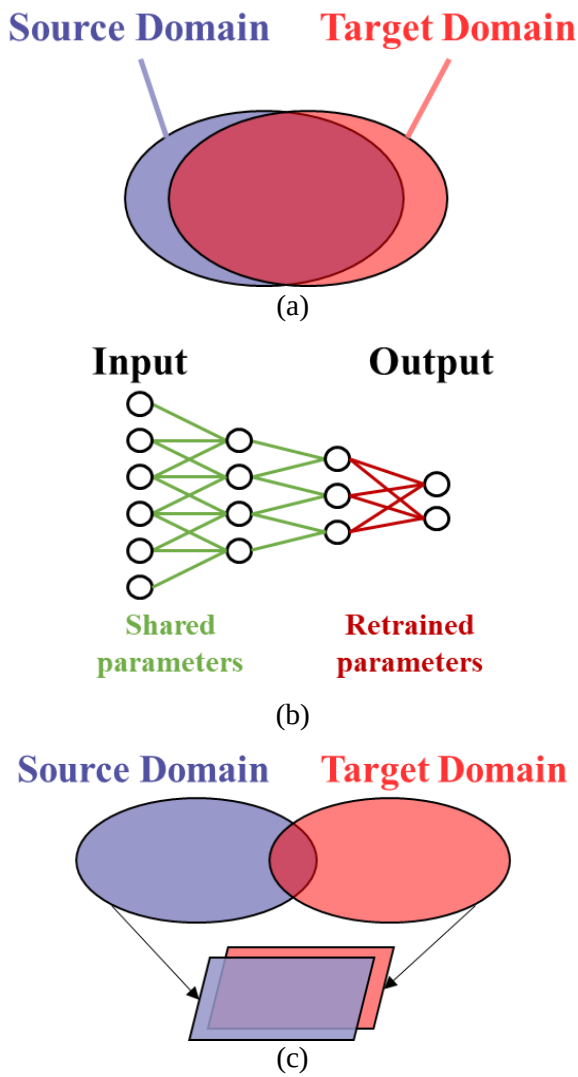


Figure 2-10 Different transfer learning approaches: (a) instance transfer; (b) parameter transfer; and (c) feature representation transfer.

2.3 Metric Learning

Metric learning is a type of learning strategy that learns more useful representations based on comparisons of the distance or similarity between samples [36]. First, the proper distance metric should be specified, as shown in Figure 2-11 (a). Then, based on this distance metric, superior representations can be learned by mapping similar samples closer to each other and mapping dissimilar samples to stay apart from each other, as shown in Figure 2-11 (b). For a supervised learning approach whose label information is available, this can be achieved by reducing the distance between samples with the same class label and increasing the distance between samples from different class labels.

These metric learning methods have been widely used in many research areas, such as image retrieval [37] and face identification and verification [38], [39]. For the purpose of calculating the similarity distance between samples, various types of network structures and loss terms have been developed. In many previous works, Siamese-network-based and triplet-network-based structures have been widely used to calculate the similarity metrics [38], [40]. Many studies have been conducted and various types of loss functions have been proposed to calculate similarity distances between samples. Sun et al. [39] used a contrastive loss and Wen et al. [22] used a center loss to obtain feature representations with large inter-class differences and small intra-class variations to solve face recognition tasks. In addition, various types of loss functions, including triplet loss and quadruple loss, have been proposed and used to learn better features for many tasks [40], [41]. Recently, studies based on metric learning have been actively conducted for many other tasks, including fault diagnosis [42]–[45]. Since metric learning can provide better feature representations

that are more discriminative and well-separated as depicted in Figure 2-12, more accurate and reliable fault diagnosis models with high generalization performance can be achieved.

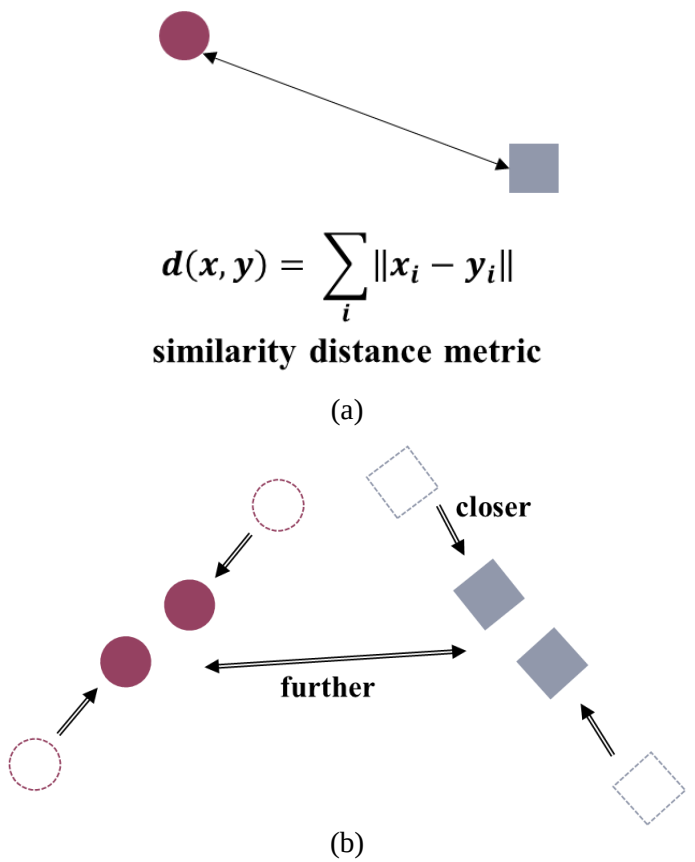


Figure 2-11 Schematic diagram of metric learning: (a) similarity distance metric; (b) learning strategy of metric learning.

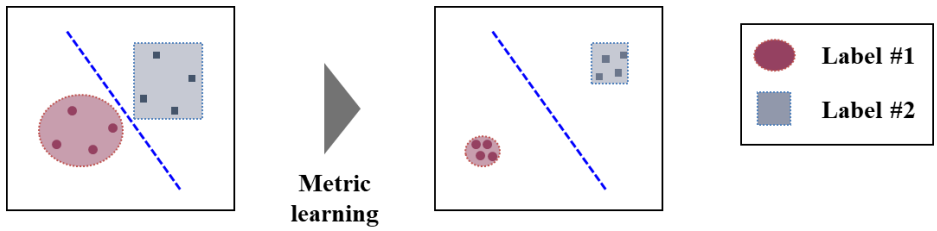


Figure 2-12 More discriminative and well-separated feature distributions obtained by adopting metric learning.

2.4 Summary and Discussion

The goal of this doctoral dissertation is to propose new methods for learning robust and high-performance diagnosis models, by addressing some difficulties that exist in the development of deep learning based fault diagnosis techniques. To this end, as described in Section 1.2, new methods are proposed to train deep learning models well, by maximizing the use of additional information. By developing improved methods to maximize the use of information for deep learning based fault diagnosis techniques, it is possible to obtain superior diagnosis models that can be used in the real industrial fields.

Chapter 3

Direct Connection Based Convolutional Neural Network (DC-CNN) for Fault Diagnosis

By stacking the neural network architectures deeper, DL can learn enriched hierarchical features through high-level abstractions of input data [15], [46]. Those abundant features are of central importance to solve given target tasks. Due to this capability, DL has yielded state-of-the-art performance in many areas [18], [47], [48]. Although the depth of the architecture is a significant factor, the number of layers cannot be increased indefinitely. The main reason is that, for deep architectures, it is extremely difficult to train whole models comprehensively due to the problems in gradient information flow that arise while training using back-propagation algorithms [49], [50]. In addition, deeper networks can easily experience overfitting problems due to a large number of trainable parameters.

In this chapter, to deal with this problem, we propose the direct connection based convolutional neural network (DC-CNN) for fault diagnosis of mechanical systems. It aims to improve the information flow within deep network architectures by directly connecting the feature maps of different layers within CNN. As a result, we can train the deep CNN architectures efficiently and learn enriched features for high diagnosis performances well by maximizing the gradient information flow.

Simultaneously, to deal with the problems that can be caused by the increased number of parameters due to direct connections, dimension reductions not only in width and height but also in the depth-wise direction are adopted.

For validation of the proposed method, experimental data obtained from the rotor testbed was employed. In order to consider the intrinsic anisotropic characteristics of rotor systems, vibration images containing not only temporal information but also spatial information along the circumference of the rotating shaft are generated and used as the input data. Compared with other machine learning and deep learning based methods, the proposed method attains the best diagnosis performance accuracy. Also, the effectiveness of the proposed method was verified in detail by conducting ablation studies and by analyzing visualization results of the learned features. Lastly, the proposed method shows stable and robust diagnosis performance under conditions of insufficient or noisy data.

The rest of this chapter is organized as follows. In Section 3.1, the concept of the directly connected convolutional module is explained. In Section 3.2, the dimension reduction module is described. Then, Section 3.3 introduces the way to generate input vibration images. In Section 3.4, the entire structure and flowchart of the proposed method are discussed in detail. Next, in Section 3.5, validations of the proposed method with experimental results and analysis are presented. Lastly, conclusions of this work are outlined in Section 3.6

3.1 Directly Connected Convolutional Module

One of the main parts of DC-CNN is the directly connected convolution module, which is based on the direct connections between different layers in the CNN architecture. The major idea of this module is to enhance the gradient information flow by improving the connectivity between various layers within CNN. In this module, to increase the training efficiency of deep CNN architectures, not only the conventional feed-forward path but also short paths that connect the feature maps of different layers are employed. The various feature maps, which are the outputs of different convolutional layers, were directly connected by concatenating them. For example, as shown in Figure 3-1, by connecting the outputs from different convolutional layers, the final output of the convolution module with two convolutional layers can be expressed as follows:

$$y_0 = [x_0, x_1, x_2] = [x_0, conv(x_0), conv([x_0, conv(x_0)])] \tag{3.1}$$

where x_0 denotes the input data; x_1 and x_2 denote the outputs of the first and second convolutional layers; y_0 denotes the final output of the convolution module; $conv$ denotes the convolutional operation; and $[a, b]$ denotes the concatenation of a and b. By adopting those direct connections between different layers, the gradient information can flow efficiently through multiple paths, rather than through a single path during the back-propagation steps. As a result, based on those directly connected convolution modules, it is possible to train deeper CNN architectures well, without falling into the problems caused by poor gradient information flows.

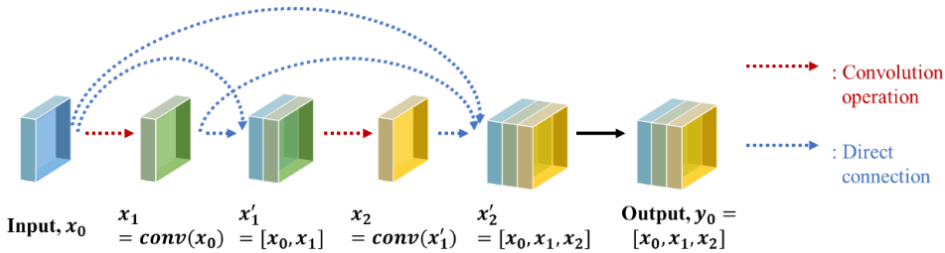


Figure 3-1 Schematic diagram of a directly connected convolution module with two convolutional layers.

3.2 Dimension Reduction Module

Although the gradient information flow can be enhanced by adopting direct connections between different feature maps, this increases the depth of the input for the subsequent layer. Consequently, the number of parameters for convolution operation for that input increases in the depth-wise direction. An increased number of trainable parameters may hinder efficient training and make the networks more prone to overfitting problems [51], [52].

In this research, to cope with those problems, a dimension reduction module was adopted. Different from conventional CNNs, which only use spatial pooling operations to decrease the dimensions of input in width and height, depth-wise pooling operations were additionally utilized in this method. For the purpose of reducing the dimension in the depth-wise direction, 1x1 convolutions were used. As shown in Figure 3-2, 1x1 convolutions produce an output with depth 1 by performing convolutional operations using 1x1xD dimension kernels. Therefore, 1x1 convolutions can reduce the number of parameters needed in the next convolutional

layers and, consequently, they can enhance the training efficiency and prevent overfitting problems.

In our dimension reduction modules, both spatial pooling and depth-wise pooling were adopted to reduce the dimensions along the H, W, and D directions of the feature maps. By using these kinds of dimension reduction modules, the number of trainable parameters can be reduced significantly. Thereby, we can train much deeper neural network architectures efficiently without the aforementioned problems that might otherwise be caused by the increased number of parameters.

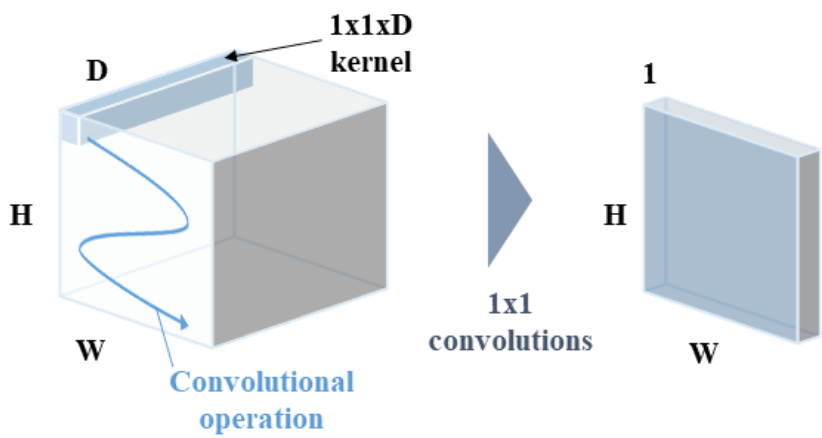


Figure 3-2 Depth-wise dimension reduction based on 1x1 convolutions.

3.3 Input Vibration Image Generation

To fully understand the health states of complex machinery, it is advantageous to obtain as much data as possible. Particularly, for rotor systems, information along the circumferential direction of the rotating shaft is important. This information is important because of the asymmetric and anisotropic characteristics of the rotor systems that arise due to uneven mass distributions, shape asymmetries, and anisotropic stiffness of the systems. As shown in Figure 3-3, anisotropic stiffness is caused by the stiffness of the components, such as the casing, bearing support, foundation, and other attachments, which typically have different values in the horizontal and vertical directions. Unequal fluid film stiffness in the radial and tangential directions can be another reason [9]. Anisotropic stiffness is common in rotor systems and has strong effects on rotor system responses. Another reason for the importance of information along the circumferential direction of the rotating shaft is the directional characteristics of some abnormal health conditions. For those health conditions with directional characteristics, such as rubbing and misalignment, the responses of rotor systems are affected by the direction of applied external forces [53]. Therefore, vibration signals obtained via sensors installed at fixed positions show different waveforms depending upon the relative positions of the faults.

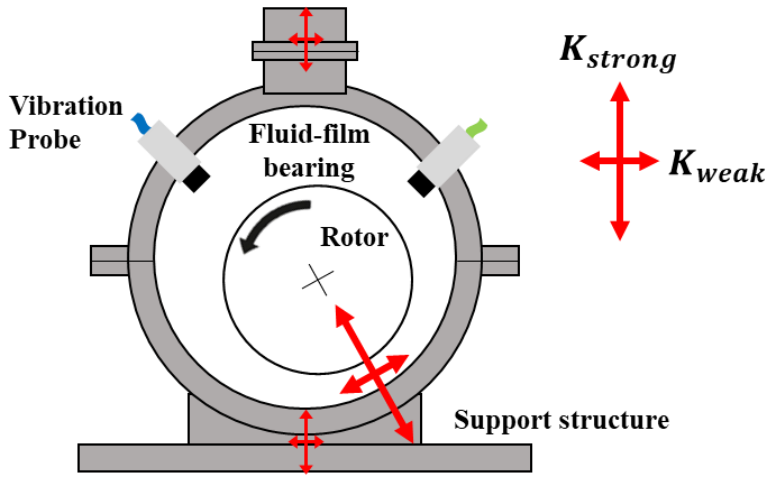


Figure 3-3 Sources of anisotropic stiffness in rotor systems. Stiffness distribution of rotor systems can be resolved into a strong stiffness axis (K_{strong}) and a weak stiffness axis (K_{weak}).

To consider those issues, the virtual vibration signal generation technique was adopted. Based on the Omnidirectional Regeneration technique proposed in [53], virtual vibration signals can be generated as follows:

$$\begin{bmatrix} \mathbf{x}_v \\ \mathbf{y}_v \end{bmatrix} = \begin{pmatrix} \cos\Delta\theta_i & \sin\Delta\theta_i \\ -\sin\Delta\theta_i & \cos\Delta\theta_i \end{pmatrix} \cdot \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix} \tag{3.2}$$

where x_0 and y_0 denote signals obtained from perpendicularly installed sensors; x_v and y_v denote a pair of generated virtual signals; and $\Delta\theta_i$ denotes the relative angular difference between $x_0, (y_0)$, and x_v (y_v). By adjusting $\Delta\theta_i$, the signals at various circumferential positions can be obtained. Generated virtual vibration signals contain information along the circumference of the rotating shaft. As a result, by using those signals, we can improve the robustness and performance of the fault

diagnosis for rotor systems under the aforementioned anisotropic characteristics. As shown in Figure 3-4, by stacking those signals, vibration images containing both spatial and temporal information can be generated.

In this paper, virtual signals generated along the full circumference of the rotating shaft were used to convey sufficient information. In addition, thanks to the shift-invariant characteristics and autonomous feature learning capability of CNN, phase synchronization between generated images is not required. Consequently, unnecessary data losses that occur during phase synchronization by removing the out-of-phase parts, can be eliminated [54]. As a result, the computational efficiency is enhanced and the number of generated images is increased. This is beneficial for developing DL-based diagnosis techniques. This can be also advantageous for online diagnosis schemes since, in the case of real-time diagnosis, the acquired signals typically have unsynchronized, random phases.

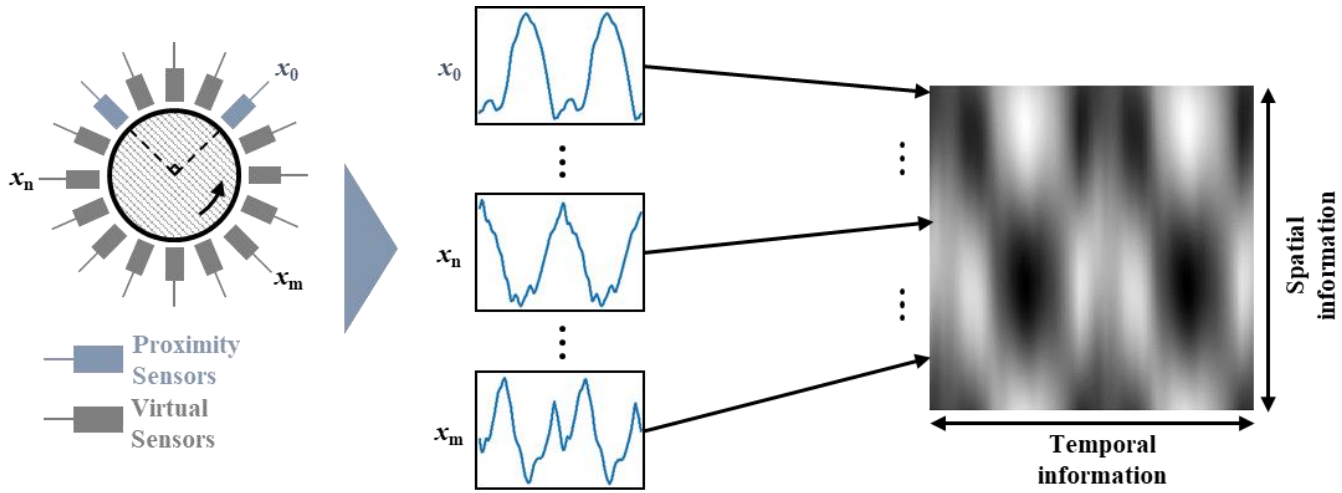


Figure 3-4 Overall procedure for generating the input vibration images.

3.4 DC-CNN-Based Fault Diagnosis Method

Based on the components explained in previous sections, this doctoral dissertation proposes an advanced DL-based fault diagnosis technique for rotor systems using DC-CNN [30]. The major building blocks and the overall architecture of the proposed method are shown in Figure 3-5. Figure 3-5 (a) shows the directly connected convolution module introduced in Section III-B. In contrast to conventional CNN, the feature maps of different layers are directly connected. This enhances the training performance by making gradient information flow more efficiently through additional connections. In the research outlined in this paper, basically, two 2-dimensional convolutional layers were used for each convolution module. For every convolutional operation, 16 kernels with 3x3 dimension and stride 1 were used and zero-padding was adopted. Between each convolutional operation, batch normalization (BN) was used to accelerate and stabilize the training process and ReLU was used as an activation function [55]. Figure 3-5 (b) shows the dimension reduction module introduced in Section III-C. By using this module, we can reduce the total number of trainable parameters within networks; therefore, we can reduce the computational budget and prevent overfitting problems even with much deeper CNN architectures. For every 1x1 convolution operation, 16 kernels were used. Then, a max-pooling layer that takes the maximum values from each sub-region with a 2x2 dimension was adopted. Also, BN and ReLU were used before 1x1 convolutions.

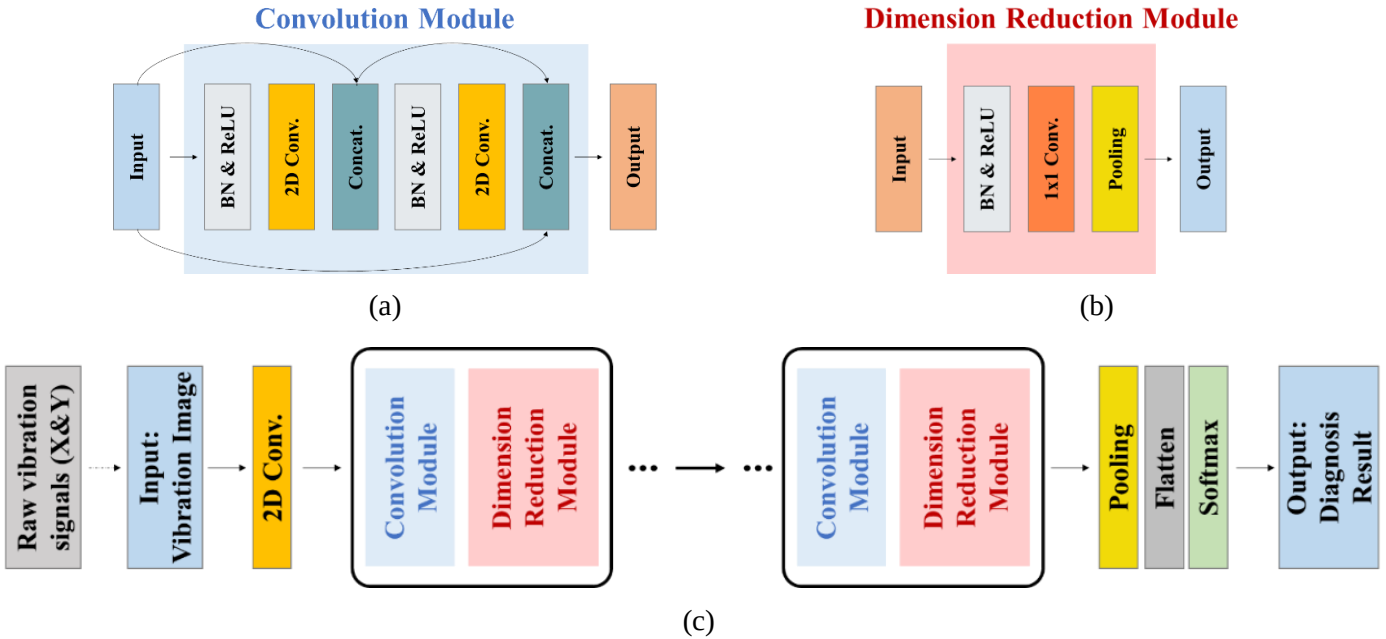


Figure 3-5 Proposed direct connection based CNN (DC-CNN) architecture: (a) directly connected convolution module; (b) dimension reduction module; (c) the overall structure of DC-CNN.

Figure 3-5 (c) shows the overall structure of the proposed architecture, which is composed of several sets of convolution modules and dimension reduction modules. First, based on the raw vibration signals obtained from two sensors in perpendicular positions, vibration images containing both spatial and temporal information of rotor systems are generated, as described in Section 3.3. In this research, vibration images containing temporal information of two cycles of vibration signals were generated to consider sub-harmonic characteristics of rotor systems. As spatial information, virtual signals along the full circumference were utilized to consider sufficient information along the shaft. After that, those vibration images were normalized to have pixel values in a range from 0 to 1 and resized into 150x150 dimensions; then, provided as the input to the 2-dimensional convolutional layer. Next, deep CNN architectures were constructed by stacking multiple sets of convolution modules and dimension reduction modules for the purpose of learning high-level representations to accurately diagnose the health states of rotor systems. In this research, network architectures constructed with 1 to 6 sets of those modules were tested. By using 2-dimensional kernels in convolutional layers, both spatial and temporal information within the input vibration images can be analyzed simultaneously. After the last dimension reduction module, a max-pooling layer was used to extract the most significant information. Then, a flatten layer was used to make the output feature maps a one-dimensional vector. Finally, the softmax function was utilized to output the vector of probability values p_i^j, which estimates the probability that the input x_i belongs to class j, to make the final decision of the fault diagnosis. The number of nodes for the softmax function is the same as the number of health conditions, C. Eventually, the final output class of the target rotor system can be obtained as the index j that has the maximum value of p_i^j, which

can be expressed as $\underset{j}{\operatorname{argmax}}\, p_i^j$. Along with the softmax function, the cross-entropy loss was used as the cost function for training our model, which can be expressed as:

$$L = \sum_{i=1}^{|B_t|} \left(-\sum_{k=1}^{C} y_i^k \log(p_i^k) \right) \tag{3.3}$$

where y_i^k denotes the value of the k^{th} node for target vector $\boldsymbol{y_i}$. Then, based on the back-propagation algorithm and the mini-batch gradient descent method, we can train our DC-CNN-based fault diagnosis models to automatically learn proper features from the given training data. In this paper, the batch size was empirically selected as 128 from {32, 64, 128, 256}, and the number of training epochs was set to 20. In addition, an Adam optimizer was utilized with a learning rate of 0.001, chosen from {0.00001, 0.0001, 0.001, 0.01, 0.1}.

Figure 3-6 shows the flowchart of the DC-CNN-based fault diagnosis method for rotor systems; the flowchart summarizes the entire procedure of our proposed method. First, as explained previously, it is possible to understand the health states of the rotor systems better based on vibration images, which contain both spatial and temporal information. Based on the training vibration image data, we train the DC-CNN-based model using a mini-batch gradient descent method. By using the proposed DC-CNN method, a fault diagnosis model with deep network architectures can be efficiently trained thanks to the enhanced information flow and the reduced number of parameters within the networks. In addition, effective and enriched hierarchical features can be obtained by using deeper network architectures. As a result, diagnosis models with higher and more stable diagnosis performance can be achieved.

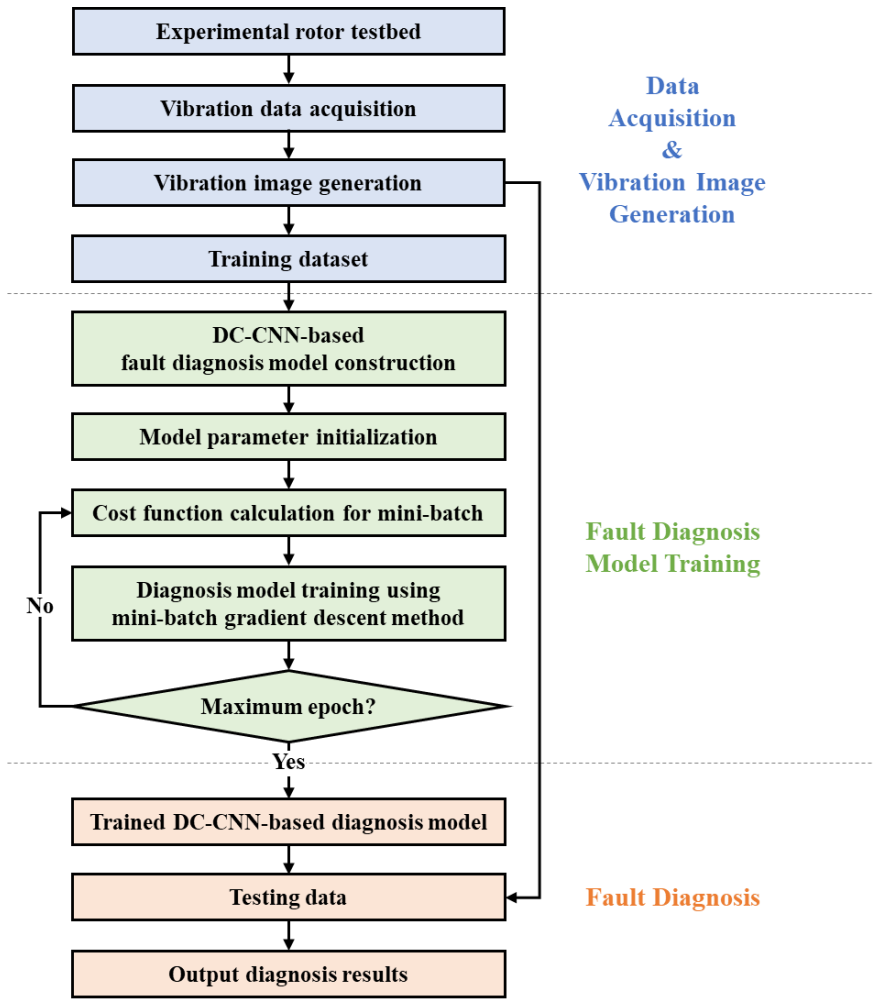


Figure 3-6 Flowchart of the proposed DC-CNN-based fault diagnosis method.

3.5 Experimental Studies and Results

3.5.1 Experiment and Data Description

In this research, the vibration signals obtained from the RK4 rotor testbed shown in Figure 3-7 were used to validate the proposed method. This testbed, supported by fluid-film bearings, was produced by GE Bently Nevada. The main components of this test apparatus are a dc motor, a coupling, rotating shafts, fluid-film bearings, a tachometer, and proximity sensors. The key-phasor signals were obtained via a tachometer to measure the rotating speed. The vibration signals were obtained via two proximity sensors installed in perpendicular positions. The vibration signals used in this work were obtained in four types of health conditions: normal, rubbing, misalignment, and oil-whirl. The experiments were conducted under steady-state conditions with a rotating speed of 3,600 RPM. For each experiment for each class, 60-second-long signals were obtained with a sampling rate of 8,500Hz. Figure 3-8 shows examples of generated vibration images for each of the four types of conditions that were used as the input data.

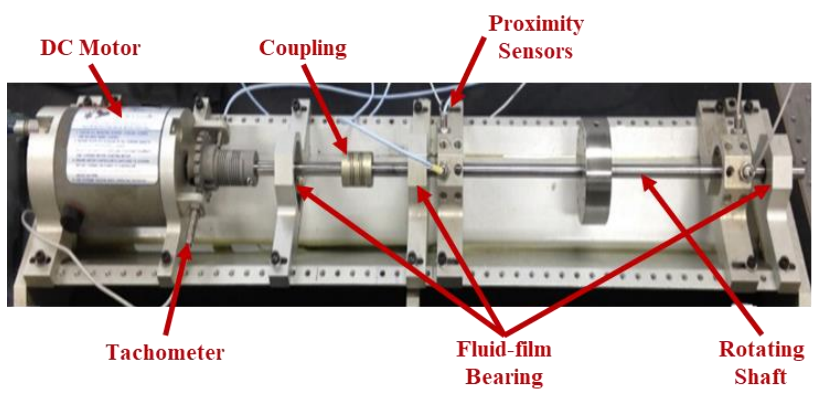
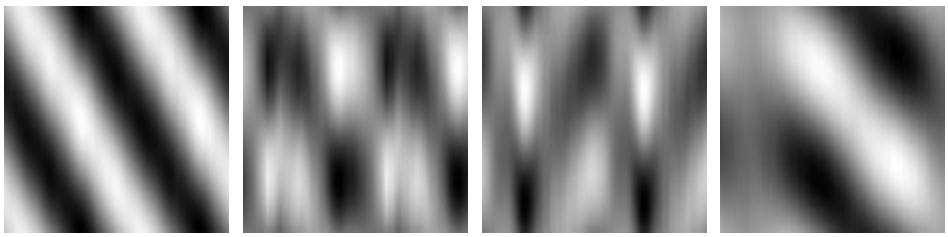


Figure 3-7 Rotor testbed: RK4 rotor kit produced by GE Bently Nevada.



(a) (b) (c) (d)

Figure 3-8 Examples of generated vibration images: (a) normal; (b) rubbing; (c) misalignment; (d) oil-whirl.

In total, five datasets were obtained from five repetitions of the experiments. For each dataset, 1000 vibration images were generated for each health condition. In addition, to take the influences of experimental conditions and randomness into account and to validate the generalization performance of the proposed methods, we made ten combinations of training and test datasets, as shown in Table 3-1. Based on those ten combinations of data, we evaluated the diagnosis performance of our proposed method and other compared methods. The implementations of all methods were carried out on a desktop computer equipped with an Intel Core i7-8700 CPU, 16 gigabytes of RAM, and NVIDIA GeForce RTX 2080 Ti.

Table 3-1 Combinations of training and testing sets based on five experimental datasets.

Dataset Combination	#1	#2	…	#10
Training Set	1, 2, 3	1, 2, 5	…	3, 4, 5
Testing Set	4, 5	3, 5	…	1, 2

3.5.2 Compared Methods

In order to validate the effectiveness of our proposed method, the diagnosis performances of other popular methods were compared. As examples of conventional machine learning methods, naïve Bayes (NB), random forest (RF), logistic regression (LR), k-nearest neighbors (KNN), and support vector machine (SVM) were used [56]–[60]. For these ML-based methods, the feature engineering steps are essential prerequisites. In this research, 8 time-domain features and 11 frequency-domain features, as shown in Table 3-2, were utilized [53], [61]. In addition, an extreme learning machine (ELM), a single hidden layer feedforward neural network whose hidden-to-output-layer weights are estimated based on the Moore-Penrose generalized inverse matrix theory, was used for comparisons [62]. As a basic DL method, multi-layer perceptron (MLP), a multi-layer fully connected artificial neural network whose weights are learned by the gradient descent algorithm, was also used. In this comparison study, we tested the MLPs with 1 to 6 hidden layers; the details of their structures are shown in Table 3-3. At the end of the MLP structure, the output layer with a softmax function, whose number of nodes is the same as the number of health classes, is followed. Since ELM and MLP require 1-dimensional inputs, the vibration image data flattened into a 1-dimensional vector was used, for fair comparisons. For the conventional CNN methods, CNNs with 1 to 6 sets of convolutional layers and pooling layers were utilized. For fair comparisons, 32 kernels with 3x3 dimensions were used for each convolutional layer, since two convolutional layers with 16 kernels were employed for each convolution module in DC-CNN. The same max-pooling layers and the same input vibration images with DC-CNN were utilized. Finally, the flatten layer and softmax function followed.

Table 3-2 Time- and frequency-domain features used for ML-based methods.

Time-domain	Frequency-domain
Max	Frequency Center
Mean	RMS Frequency
RMS	Root Variance Frequency
Skewness	0.5x / 1x
Kurtosis	2x /1x
Crest Factor	(3x~5x) / 1x
Shape Factor	(3x,5x,7x,9x) / 1x
Impulse Factor	(1x~10x) / 1x
	(0~0.39x) / 1x
	(0.4x~0.49x) / 1x
	(0.51x~0.99x) / 1x

Table 3-3 The number of hidden layers and nodes for MLP-based fault diagnosis methods.

Number of hidden layers	Number of nodes in hidden layers
1	4000
2	4000-2000
3	4000-2000-1000
4	4000-2000-1000-500
5	4000-2000-1000-500-250
6	4000-2000-1000-500-250-100

Lastly, a comparison with the existing method which tries to improve gradient information flow using short connections was performed. As shown in Figure 3-9, a model using residual learning based on identity mapping was employed for comparison [21]. This approach combines features through summation of the results of nonlinear mapping and identity mapping. The schematic diagram of this skip connection based CNN model is shown in Figure 3-10. For fair comparisons, the same number of convolutional kernels with 3x3 dimensions were used for each convolutional layer: two convolutional layers with 16 kernels were employed for each skip connection module. Also, the same structure was used for all other parts.

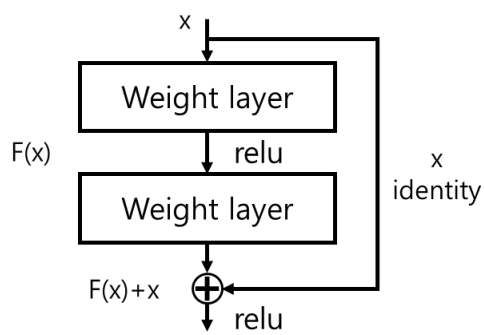


Figure 3-9 Conceptual diagram of residual learning.

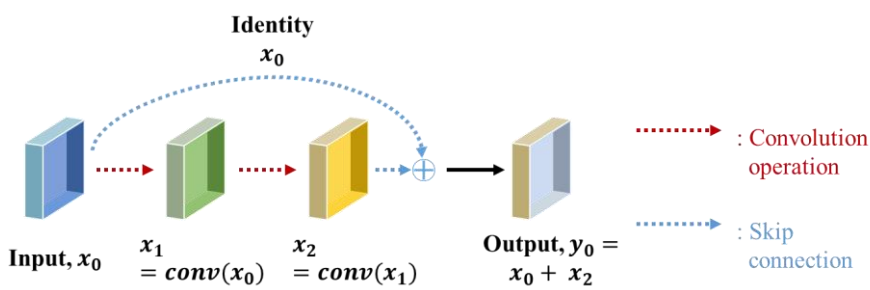


Figure 3-10 Schematic diagram of skip connection based CNN model.

3.5.3 Diagnosis Performance Results

The diagnosis accuracy of the proposed and compared methods for the RK4 rotor testbed are shown in Figure 3-11. Diagnosis accuracy in this figure shows the mean accuracy values for the ten combinations of datasets mentioned in Table 3-1; the error bars show their standard error values. The number 1 to 6 in the name of the DL-based method represents the number of building blocks for each method.

Figure 3-11 shows that DC-CNN-based methods achieved better performance than any other compared methods. ML-based methods show relatively poor performance, although they used domain knowledge based features. By looking at the results from neural network-based methods, similar diagnosis performance with ML-based methods can be achieved even without feature engineering steps. Therefore, we can say that domain knowledge dependency problems can be alleviated by using DL. CNN-based methods show better diagnosis performance than MLP-based methods that employed the vibration image data flattened into a 1-dimensional vector as input data. This result shows that the fault diagnosis performances can be improved by adopting the 2D vibration images containing both temporal and spatial information as the input data, along with CNN, which can consider multiple dimensional information within input data.

By stacking the networks deeper, we can somewhat improve the diagnosis performance of DL-based models. However, from the results of MLP and CNN, we can see that the performance does not increase monotonically as the number of layers is increased. As explained in previous sections, poor gradient information flow and overfitting can be the main reasons for those problems. In contrast, using DC-CNN, we can achieve better diagnosis performance by making network architectures

deeper. For DC-CNN, the model with the deepest network architecture (DC-CNN6) achieved the best overall performance with the lowest standard error value. In conclusion, we can say that DC-CNN-based fault diagnosis methods lead to the best diagnosis performance among all of the compared methods, by making the efficient training of deep CNN models possible.

Moreover, in order to verify the effectiveness of the proposed DC-CNN-based fault diagnosis method, ablation studies were conducted. The average diagnosis accuracy and standard error values for four types of models were investigated; the four types of models are as follows: (1) the CNN model without either directly connected convolutional modules or dimension reduction modules, (2) the proposed model without dimension reduction modules, (3) the proposed model without directly connected convolutional modules, and (4) the proposed DC-CNN-based model, with both types of modules. The results are shown in Table 3-4.

By comparing the conventional CNN model, which does not adopt either type of module, and the methods in which the proposed modules were used, the effectiveness of the directly connected convolutional module and dimension reduction module can be confirmed. In addition, our proposed method, which adopted both modules, maximizes the diagnosis accuracy with the smallest standard error value. In conclusion, from these ablation studies, we can validate the effectiveness and robustness of our proposed DC-CNN-based fault diagnosis method.

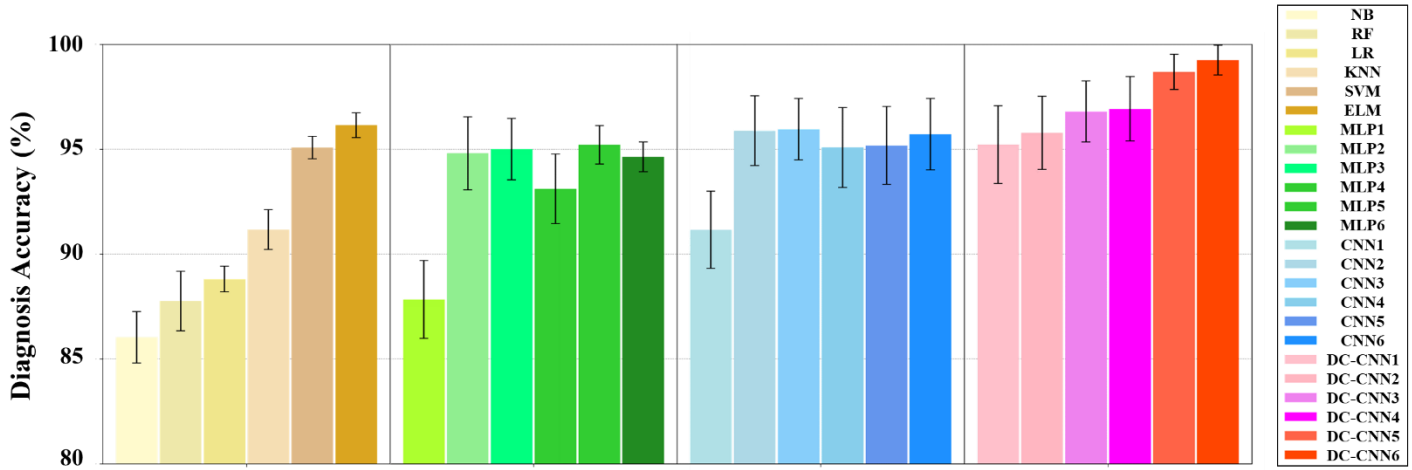


Figure 3-11 The diagnosis performance results of the proposed and compared methods.

Table 3-4 Average diagnosis accuracy and standard error results for ablation studies.

Model types	Diagnosis accuracy (%)	Standard error
w/o both modules	95.71	0.0169
w/o dimension reduction modules	98.33	0.0120
w/o directly connected convolutional modules	98.59	0.0117
Proposed	**99.25**	**0.0071**

Next, the results of comparing the training time of each DL-based method are shown in Table 3-5. In the case of DC-CNN, it can be seen that the training time required per epoch is higher than other methods. However, as shown in Figure 3-12, DC-CNN takes fewer training epochs to complete model training. Therefore, eventually, taking a longer time per epoch for DC-CNN is not a big problem in the learning process.

Table 3-5 Diagnosis accuracy and training time for DL-based methods.

Model type	MLP	CNN	**DC-CNN**
Diagnosis accuracy (%)	95.21	95.96	**99.25**
Training time (seconds)	6	6	25

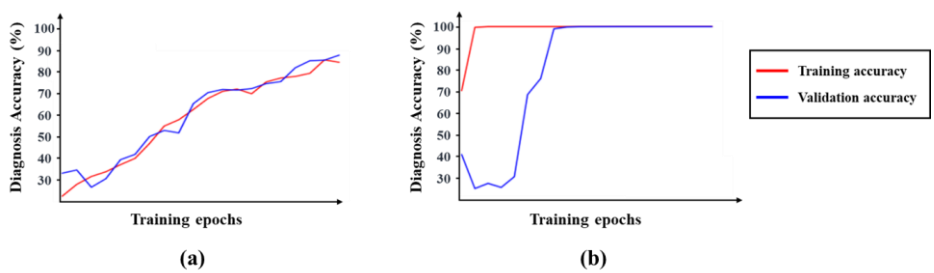


Figure 3-12 Learning curves of (a) CNN model; and (b) DC-CNN model.

Lastly, the results of comparing the diagnosis accuracy and training time of the proposed method and the skip connection based method are shown in Table 3-6. As can be seen in this result, the proposed DC-CNN method requires a shorter learning

time per training epoch and shows higher final diagnosis performance compared to the skip connection based method. Also, in the case of the proposed method, by concatenating several features through direct connections, it can maintain all feature information from several layers and therefore, has the advantage of being able to learn diverse features.

Table 3-6 Diagnosis accuracy and training time for the skip connection based method and the proposed methods.

Model type	Skip Connection	**DC-CNN**
Diagnosis accuracy (%)	97.98	**99.25**
Training time (seconds)	60	**25**

3.5.4 The Number of Trainable Parameters

In order to reduce the number of parameters for constructing efficient DL models, dimension reduction modules were employed in DC-CNN. In the research outlined in this section, the number of trainable parameters for DL-based methods was investigated to verify the effects of the proposed DC-CNN-based fault diagnosis method.

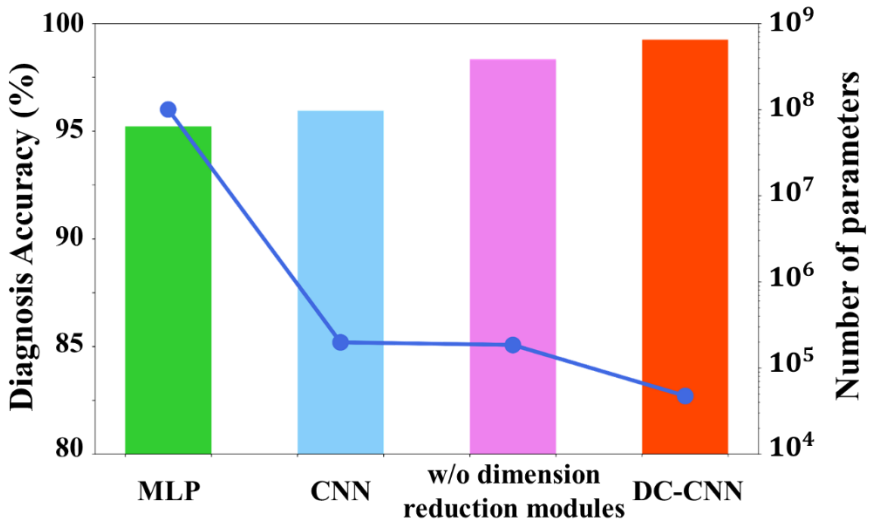


Figure 3-13 Diagnosis accuracy and the number of trainable parameters for DL-based methods. (Bar: diagnosis accuracy, Dot: the number of parameters)

Figure 3-13 shows the number of trainable parameters and diagnosis accuracies of the various DL-based methods. The values for MLP, CNN, and DC-CNN in this figure are from the models with the highest diagnosis accuracy among the six different models of each DL-based method described in Section 3.5.3. In addition, the result of the proposed model without dimension reduction modules is compared to validate the effects of those modules on the number of parameters within the model. As can be seen in Figure 3-13, our proposed method results in the best diagnosis performance with the smallest number of parameters. It is important to note that for DC-CNN, the best model was the deepest model, and – even for this deepest model – the number of parameters was smaller than in other DL-based models. This advantage is achieved through the dimension reduction modules of the

proposed method; this can be confirmed by comparing the number of parameters according to whether or not the dimension reduction modules are used. By adopting the dimension reduction modules, the number of parameters could be greatly reduced. As a result, the advantages of DC-CNN-based methods make it possible to build diagnosis models with deep architectures efficiently; this, in turn, leads to better diagnosis performance.

3.5.5 Visualization of the Learned Features

In order to examine and understand the DL-based models qualitatively, visualization of the learned features can be conducted. In this section, two kinds of visualization results are provided to show the effectiveness of the proposed DC-CNN-based method. First, the distributions of the learned features were visualized based on the t-distributed Stochastic Neighbor Embedding (t-SNE) method. t-SNE is a dimension reduction method that maps the high-dimensional data into the low-dimensional embedded space [63]. Using t-SNE, we can map the features extracted by DL-based methods into two-dimensional spaces. By visualizing them, we can verify how efficient the learned features are for the diagnosis of target rotor systems.

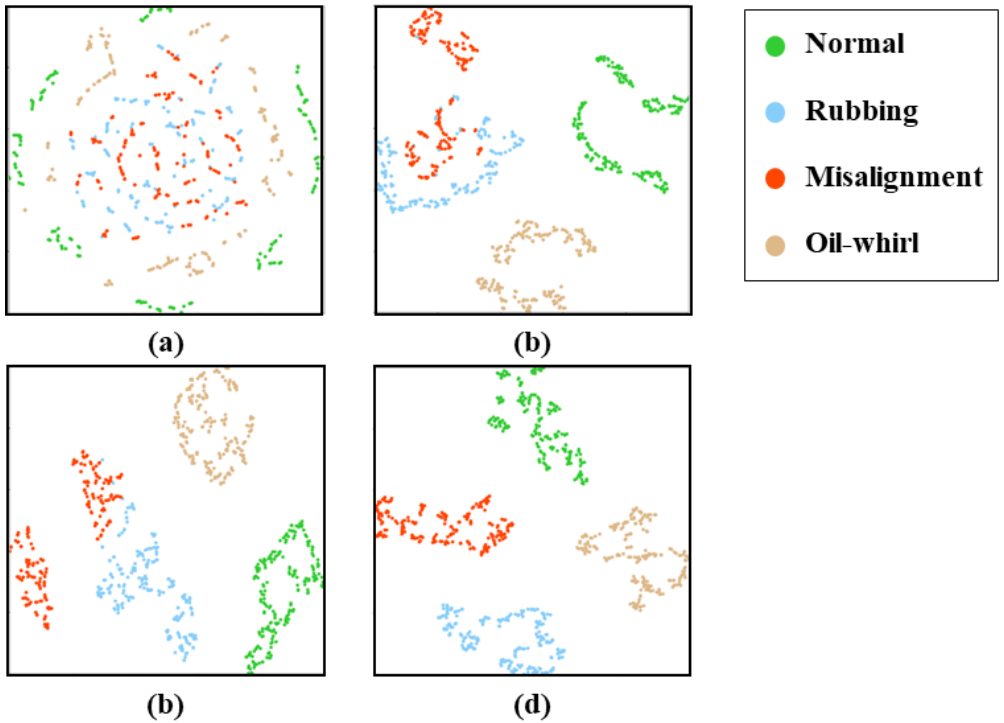
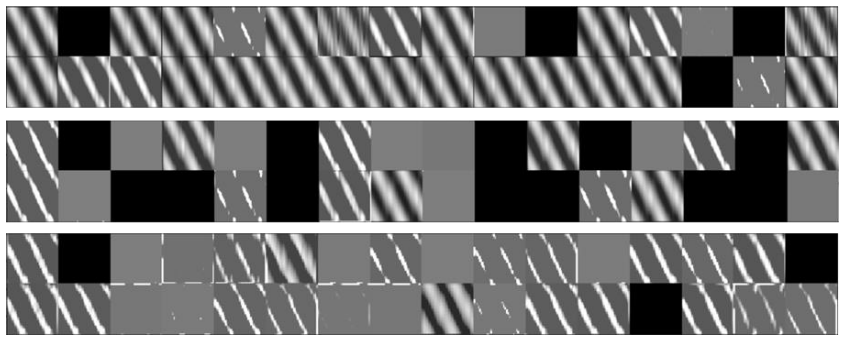


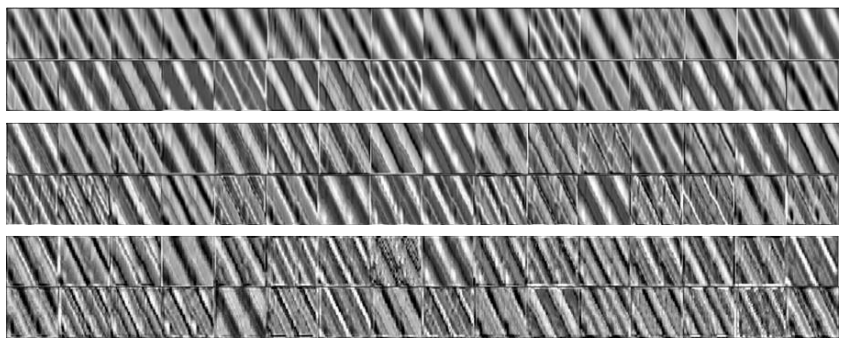
Figure 3-14 Feature visualization using t-SNE: (a) input raw data; learned features of (b) MLP; (c) CNN; (d) DC-CNN.

Figure 3-14 shows the visualization of the raw input data and learned features based on the DL methods. Based on those t-SNE-based plots, the effectiveness of the learned features in the diagnosis of the fault classes can be investigated. The learned features from the proposed DC-CNN-based methods result in the best clustering characteristics. This means that data in the same health condition are well-grouped in the learned feature space and also well-separated from the data from other conditions. Therefore, based on the proposed method, high diagnosis accuracy can be obtained by learning the features that make data well-clustered for each health condition. On the other hand, for other methods, there are some overlaps in the data from different classes and some data from the same class is separated. Those features that are not well-clustered for each health condition adversely affect the diagnosis accuracy.

Second, for CNN-based methods, the visualization of the feature maps can provide meaningful insights [46], [64]. Therefore, the feature maps of CNN and DC-CNN were visualized as shown in Figure 3-15 for the normal condition. In Figure 3-15 (a), for the CNN-based model, there are lots of inactivated, noise-like feature maps, and similar feature maps that may provide redundant information. In contrast, in Figure 3-15 (b), the DC-CNN-based model has absolutely no inactivated feature map and there are diverse kinds of learned feature maps. Those observations imply that DC-CNN-based methods can learn enriched features for fault diagnosis efficiently and properly.



(a)



(b)

Figure 3-15 Visualization of the feature maps of (a) CNN-based; (b) DC-CNN-based models for a normal condition.

In conclusion, from the above two visualization results of learned features, it can be seen that DC-CNN can learn better features more effectively than other methods. This is one of the major reasons why DC-CNN-based approaches yield better diagnosis performance.

3.5.6 Robustness of Diagnosis Performance

The ultimate goal of developing fault diagnosis techniques is to enhance the reliability of the machinery operating in real industrial fields. Therefore, when developing fault diagnosis techniques, it is necessary to consider issues that may occur in real-world settings and to develop algorithms that are robust when faced with those issues. The first issue we may encounter in the field is the amount of data available. In order to train diagnosis models completely, a sufficient amount of data is required. However, in real operating situations, it is not easy to collect enough training data. An insufficient amount of training data may worsen the training efficiency and lower the performance of diagnosis models. As a result, it is very important to develop the diagnosis methods to be robust to the amount of training data. In this section, to examine the robustness of DC-CNN methods to the amount of training data, we investigated its diagnosis performance for situations with various amounts of training data.

Figure 3-16 presents the diagnosis accuracies of DL-based models trained with various numbers of training samples. The number of training data in the x-axis expressed as a percent represents the ratio of the number of data used for training to the total number of training data. In this figure, it can be seen that as the number of training data decreases, so does the diagnosis performance. In addition, we can see that our DC-CNN-based method shows the best performance in all cases. For other DL-based methods, diagnosis performance decreases more steeply as the number of training data decreases. On the other hand, in the case of DC-CNN-based methods, the diagnosis performance is relatively high, even when the number of data is very small.

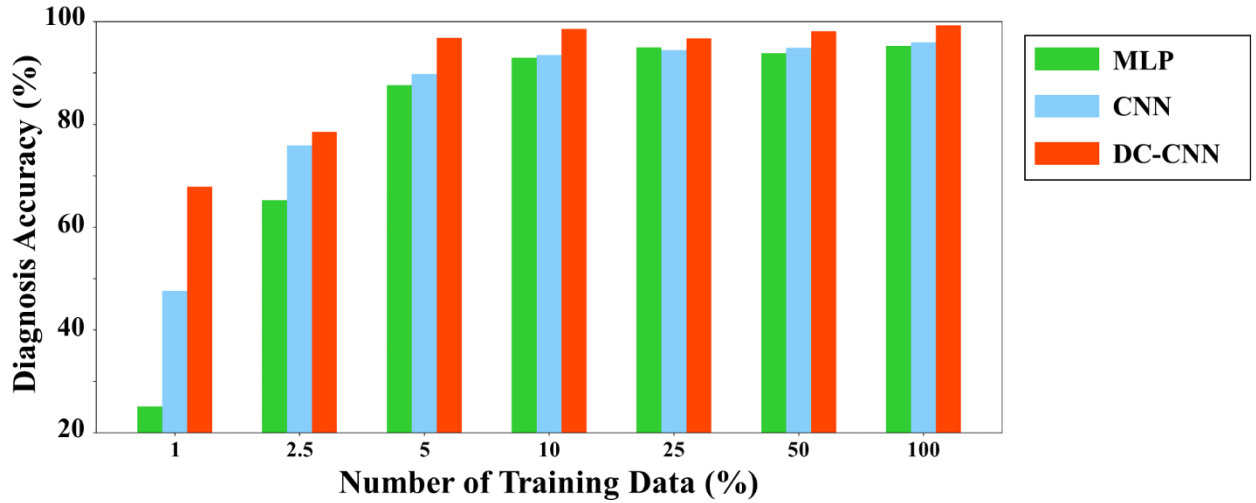


Figure 3-16 The diagnosis performance results of DL-based methods with different numbers of training data.

These results show that our proposed method can learn the diagnosis models very efficiently and provide high diagnosis performance even with a small amount of data. These stable and robust characteristics might be the results of the increased training efficiency that is due to the improvement of gradient information flow. In contrast, for MLP- and CNN-based methods, the diagnosis performance is highly dependent on the amount of data; therefore, these methods may be inadequate for use in real settings where there is a lack of data.

The next issue that may exist in real industrial sites is the issue of noisy data, which can be caused by environmental and operational randomness. In the real world, various noises inevitably exist and it is impossible to train the diagnosis models to consider all possible noises in advance. Thus, the noise that wasn't present during the training phase may be present in the test data; this can drastically degrade the effectiveness of the trained diagnosis models. Therefore, it is necessary to develop diagnosis models that are robust for various noisy data. For the purpose of verifying the robustness of the proposed methods against noisy data, additive white Gaussian noises with different signal-to-noise ratios (SNR) were added to test datasets. The SNR in decibels (dB) is defined as follows:

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) = P_{signal,dB} - P_{noise,dB} \tag{3.4}$$

where P_{signal} and P_{noise} denote the power of the signal and the additive white Gaussian noise, respectively.

The diagnosis results for the test data with additive noise are shown in Figure 3-17. It can be observed that for all methods, the higher the ratio of noise, the lower

the diagnosis performance, since the additional noise causes some variations in the characteristics and distributions of the test data. It can also be seen that the proposed DC-CNN-based method outperforms the other DL-based methods for all SNR values. In addition, the proposed method shows high diagnosis performance even if some noise is added; whereas, the other methods show very rapid degradation of performance with additive noises.

In conclusion, the proposed method yields more robust and stable generalization performance against additive noise, as compared to other methods. This may be due to the following advantages of DC-CNN. First, enriched features for diagnosis can be obtained through efficient training thanks to improved connectivity within network architectures. At the same time, by reducing the number of parameters, the overfitting problem can be prevented. This results in improved generalization performance.

In summary, based on the above two experimental results, it was confirmed that the proposed DC-CNN-based fault diagnosis method can work well under several issues that may occur in real-world operating environments, such as in cases of insufficient training data or inevitable noise conditions.

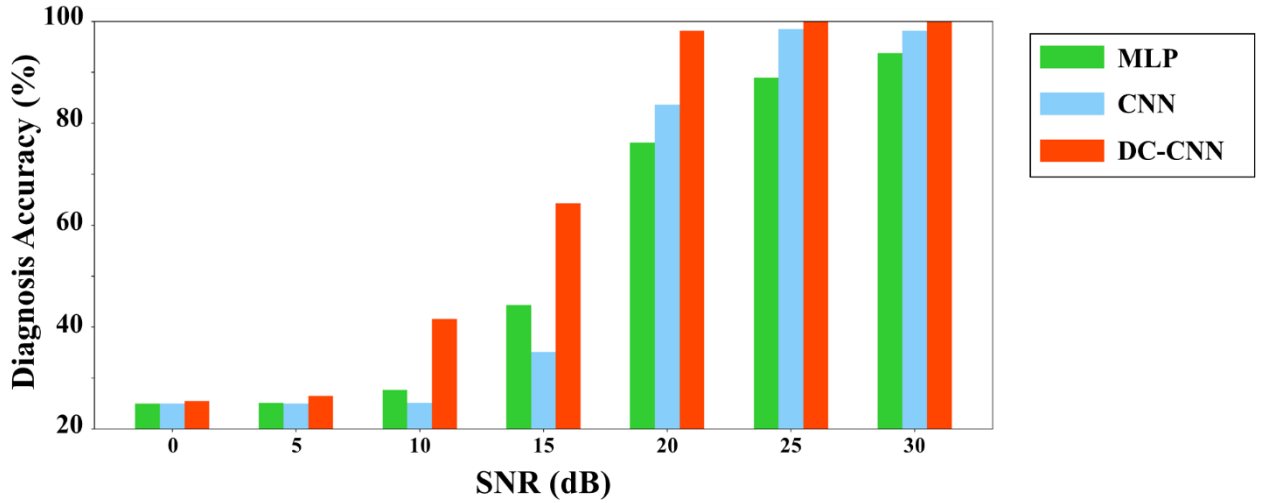


Figure 3-17 The diagnosis performance results of DL-based methods under different levels of additive noise .

3.6 Summary and Discussion

In this research, a novel DC-CNN method, composed of directly connected convolution modules and dimension reduction modules, was proposed for fault diagnosis of mechanical systems. The goal of DC-CNN is to enhance the training efficiency and diagnosis performance by improving the gradient information flow and reducing the number of parameters. In addition, in this study, the vibration image generation technique was adopted to consider the intrinsic anisotropic characteristics of the rotor system, which was used for validation. By using DC-CNN, not only can outstanding diagnosis performance (over 99% diagnosis accuracy) be obtained, efficient training can also be conducted. The proposed method outperformed other DL-based methods, even with the smallest number of parameters. The ablation studies demonstrated that the proposed method maximizes the diagnosis performance with the smallest standard error value, thanks to the proposed modules. The visualizations of the learned features showed that the proposed method can learn effective and enriched features. In addition, the proposed method showed stable and robust performance, even with a limited number of training data and additive noise conditions; this demonstrates the superiority of our method for real-world problems.

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

1) **Myungyon Kim,** Joon Ha Jung, Jin Uk Ko, Hyeon Bae Kong, Jinwook Lee**,** and Byeng D. Youn, "Direct Connection Based Convolutional Neural Network (DC-CNN) for Fault Diagnosis of Rotor Systems," *IEEE Access*, Vol 8, pp 172043-172056, 2020.

Chapter 4

Robust and Discriminative Feature Learning for Fault Diagnosis Under Insufficient and Noisy Data Conditions

For data-driven strategies including deep learning, sufficient data is required to train high-performance fault diagnosis methods. Otherwise, it is very hard to secure accurate and robust diagnosis performances as can be seen in Figure 4-1. However, in real industrial fields, it is not easy to obtain as much data as we need from systems of interest to train diagnosis models well. Especially, it is hard to obtain a sufficient amount of fault data, since machines usually work in a normal condition and faults are rare. In addition, in the case of newly operated systems, the amount of acquired data is small. Moreover, data-driven methods work well under the assumption that the training and test data share the same distributions. However, for the mechanical systems operating in real industrial sites, there can be lots of noises caused by environmental and operational randomness. Noises can cause variations in the distribution of training and test data, and as a result, it may become hard to acquire high diagnosis performances for test data. This can be seen in Figure 4-2.

In this chapter, we adopt the transfer learning and metric learning concepts to train superior fault diagnosis models even under insufficient data and noisy data

conditions. First, robust features can be learned by transferring the feature and parameter information obtained from a different domain that has an abundant amount of data. In addition, semantically better aligned and more discriminative features can be learned by adopting semi-hard triplet loss for training the fault diagnosis models. As a result, better fault diagnosis models can be obtained under insufficient and noisy data conditions.

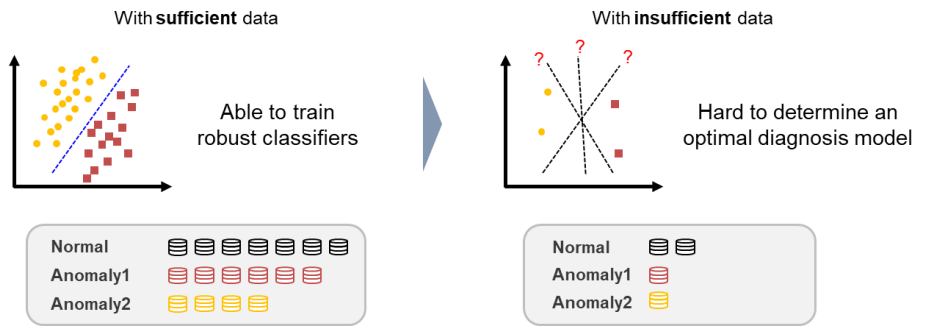


Figure 4-1 The learning results of the fault diagnosis models according to the amount of data.

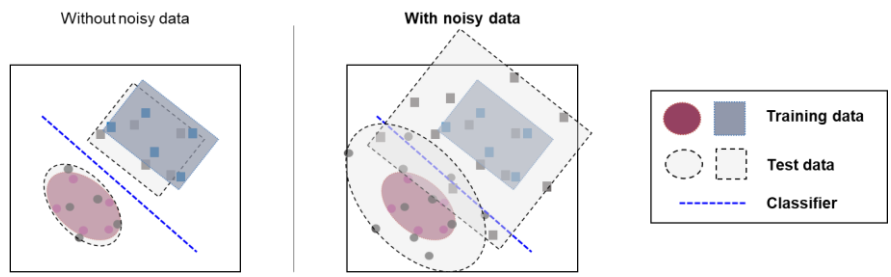


Figure 4-2 The learning results of the fault diagnosis models according to the amount of noisy data.

The effectiveness of the proposed method was verified using the data obtained from the rotor testbed. The fault diagnosis performances were comprehensively compared for the case where there was insufficient data and the case where the data contained noise. In addition, the learning curve was analyzed to compare how well the model learned, and the visualization results of the feature distribution were compared to confirm the superiority of the learned features.

The remainder of Chapter 4 is organized as follows. Section 4.1 provides the concept of parameter transfer learning. Then, robust feature learning based on the pre-trained model is presented in Section 4.2. In Section 4.3, discriminative feature learning by adopting the triplet loss term is described. Section 0 describes the overall learning procedure of the proposed method. Next, in Section 4.5, the experimental results and analyses are given. Finally, summary and discussion are outlined in Section 4.6.

4.1 Parameter transfer learning

As mentioned in Chapter 2.2, transfer learning is a learning method that transfers and uses information and knowledge acquired from different but related source domains. This allows the model to be trained more efficiently even in situations where data is insufficient. Due to this advantage, transfer learning is actively used when training the data-driven models. Among them, the parameter transfer approach is a method that conveys the parameters of the model learned from the source domain to the target domain. In other words, by discovering and sharing the common parameters, knowledge can be transferred across different domains and

tasks.

In practice, since data is often insufficient, it is not easy to obtain the model with high performance when training the deep learning model from scratch. So, instead of training a randomly initialized model from scratch, many people train the deep learning model by using a model already trained in a source domain that has sufficient data as an initial model. This kind of model that has already been trained in the source domain is called a pre-trained model. The schematic diagram of the parameter transfer learning scheme is presented in Figure 4-3. First, the model is trained using supervised learning based on the abundant source domain data. After that, some parts of this pre-trained model are transferred to the target domain. Then, the new model for the target domain, containing shared parameters is retrained using target domain data. This parameter transfer method based on the pre-trained models assumes that individual models for related domains and tasks may share some parameters. As mentioned in Section 2.2, and as shown in Figure 2-9, by utilizing the pre-trained model, training of the deep neural network models can become stable, efficient, and fast. Also, the final diagnosis performance can be improved.

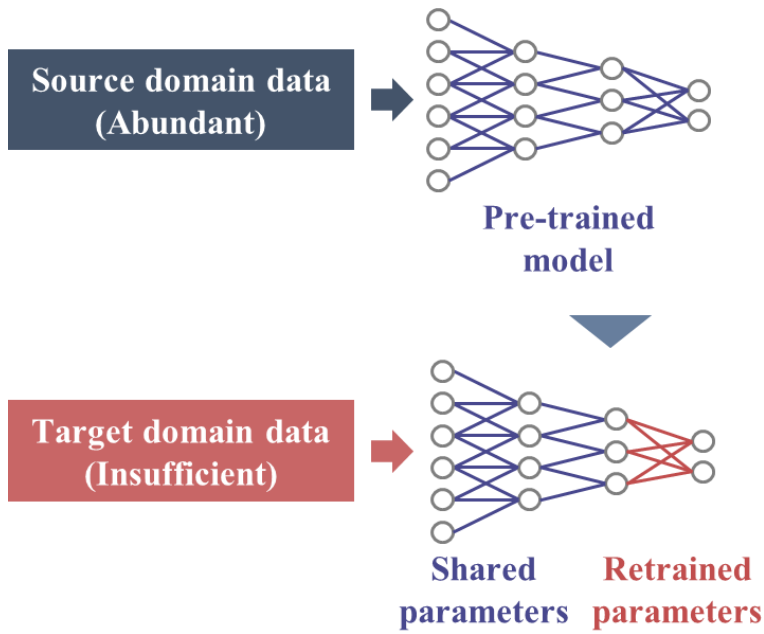


Figure 4-3 Schematic diagram of the parameter transfer learning scheme for neural network based model.

4.2 Robust Feature Learning Based on the Pre-trained model

In this study, in order to efficiently train deep learning based fault diagnosis techniques even under insufficient data conditions, the parameter transfer method is used. The diagnosis model with high generalization performances for the target system can be learned by transferring and employing the pre-trained model obtained from the source domain containing a large amount of data. This enables robust and reliable feature learning of the diagnosis model. In this research, the ImageNet challenge (ImageNet Large-Scale Visual Recognition Challenge, ILSVRC) dataset

was used as the source domain. This dataset is a very large-scale database designed for use in visual object recognition research. It consists of more than 14 million images that are annotated. Since the ImageNet dataset contains a large amount and various image data, it is possible to obtain general and superior features that are effective in representing images from this data. As a pre-trained model, the VGG19 model trained using this ImageNet data was adopted [18]. VGG19 is a model developed by researchers from the University of Oxford, which is a deep convolutional network model made by repeatedly stacking 3x3 convolutions deeply, as shown in Figure 4-4. This model is used in many areas due to its high performance and relatively simple network structure.

The strategy to increase the learning efficiency and diagnosis performance in the target domain by transferring the parameters of the pre-trained model is as follows. By transferring and reusing the feature extractor part learned through a large number of source domain images, robust and reliable feature learning can be achieved when training the diagnosis model in the target domain. Then, the classifier part is newly constructed according to the number of target domain labels, attached to the top of the feature extractor part, and trained with the target data. At this time, by fine-tuning the entire model, which consists of the transferred feature extractor part and the newly attached classifier part, with target data, a diagnosis model suitable for the target domain can be finally obtained. Figure 4-5 shows a schematic diagram of the training processes of the CNN-based fault diagnosis method. A learning strategy using a conventional supervised learning scheme is presented in Figure 4-5 (a). Figure 4-5 (b) demonstrates the diagnosis model training process based on the parameter transfer learning strategy. As the pre-trained model, the CNN

model trained using ImageNet data as the source domain was adopted. By using a transfer learning scheme like this, robust and stable feature learning is possible.

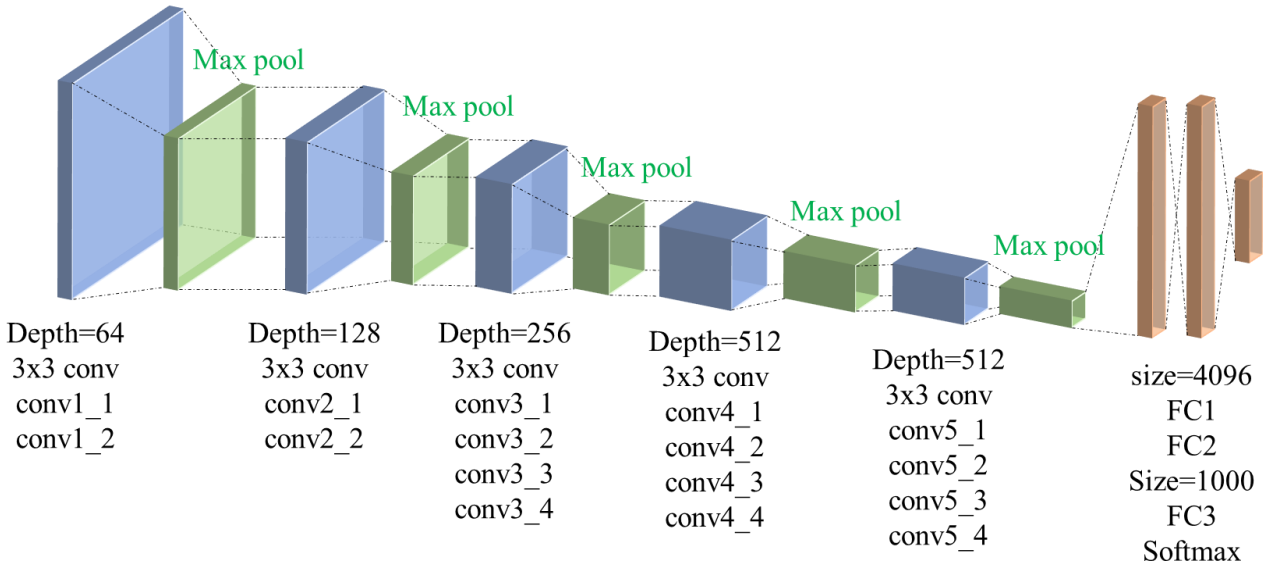


Figure 4-4 The network structure of the VGG 19 model.

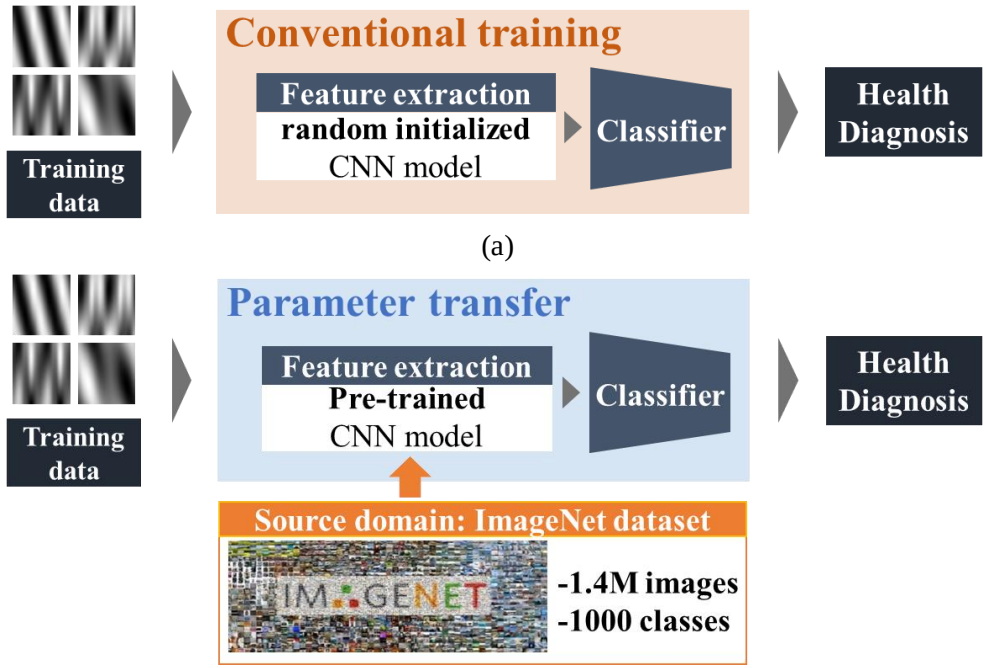


Figure 4-5 Schematic diagram of the training processes of CNN-based fault diagnosis method: (a) using conventional supervised learning strategy; (b) using transfer learning strategy.

4.3 Discriminative Feature Learning Based on the Triplet loss

In this research, we propose a learning approach that can obtain more discriminative features by training a deep learning based fault diagnosis model using an additional metric learning based loss term. As explained in Chapter 2.3, metric learning is a type of learning strategy that can acquire semantically well-aligned and well-separated features based on a distance metric. This can be achieved by placing similar samples with the same label closer together and dissimilar samples with different labels farther apart. In this study, triplet loss is used as an additional loss function to learn superior feature embedding. The concept of triplet loss can be confirmed in Figure 4-6. It aims to make the distance between a pair of samples with same label smaller than the distance between a pair with different labels. As shown in Figure 4-6, the anchor (a) is a baseline sample, and the positive sample (p) is a sample that has the same label with an anchor. Oppositely, the negative sample (n) is a sample whose label is different from the anchor sample. Based on the triplet loss, we want to ensure that the anchor is closer to all other samples with the same label than any sample with different labels, and this can be expressed as follows:

$$\|f(a) - f(p)\|^2 + \alpha < \|f(a) - f(n)\|^2 \tag{4.1}$$

where $f(x)$ denotes the feature embedding; and α denotes a margin that is enforced between positive and negative pairs. The triplet loss function to be minimized can be expressed as follows:

$$L_{triplet}(a, p, n) = \max(\|f(a) - f(p)\|^2 - \|f(a) - f(n)\|^2 + \alpha, 0) \tag{4.2}$$

Therefore, by minimizing this loss term, we push the distance between a and p to

0, and make the distance between a and n to be greater than the positive distance plus margin. As a result, we can make negative distances at least greater than the positive distance plus the margin.

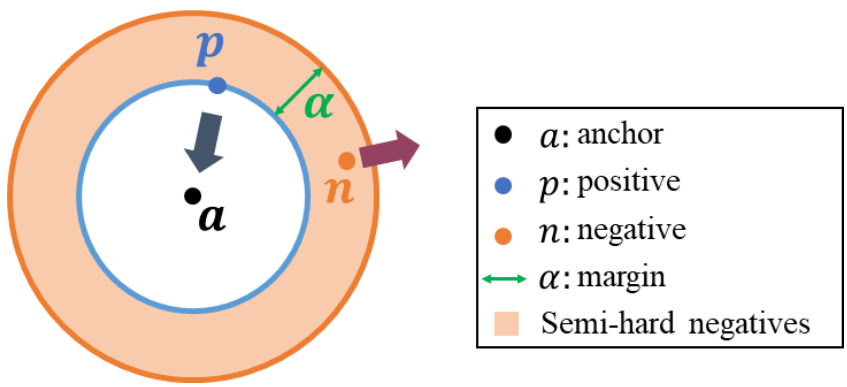


Figure 4-6 Schematic diagram of semi-hard triplet loss.

For effective feature learning and fast convergence, it is essential to select triplet samples that violate the constraint in Equation (4.1). That is, a hard positive, which is a positive sample as far as possible from the anchor, and a hard negative, which is a negative sample as close as possible to the anchor sample, should be selected. Such hard positive and hard negative can be expressed as follows:

$$argmax_{p_i} \|f(a) - f(p_i)\| \tag{4.3}$$

$$argminx_{n_i} \|f(a) - f(n_i)\| \tag{4.4}$$

where p_i and n_i denote possible candidates for positive and negative samples. In practice, however, when the most difficult negative sample was selected, training may not be performed efficiently in the early stage of learning. In other words, the

model can fall into bad local minima, and it may lead to a collapsed model whose feature embedding results in like this: $f(x) = 0$. To solve this problem, we use the semi-hard negative samples located in the orange region in Figure 4-6. These are the negative samples that are not closer to the anchor than the positive samples, but still difficult; which means that these negative samples make the loss function positive. In other words, those semi-hard negatives are further away from the anchor than the positive samples, and simultaneously, their distance to the anchor sample is close to the anchor-positive distance. So, they lie inside the margin α. In addition, for calculating the triplet loss function, the positive and negative samples are selected within a mini-batch, and then semi-hard triplet loss is computed based on the semi-hard negative and hard positive samples. This is called online triplet mining. Consequently, based on this additional triplet loss function, discriminative features with semantically well-separable characteristics can be learned. As a result, as can be seen in Figure 4-7, by learning separated features, it is possible to compensate for the performance degradation caused by the distribution variation due to noises and obtain a good diagnosis model.

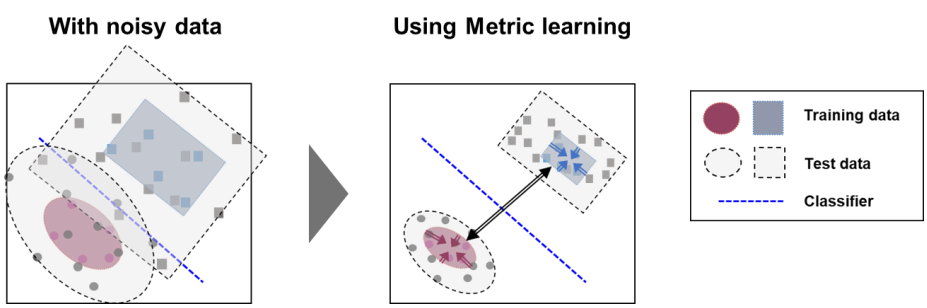


Figure 4-7 Schematic diagram showing the effect of metric learning under noise conditions.

4.4 Robust and Discriminative Feature Learning for Fault Diagnosis

The method proposed in this study uses the parameter transfer and metric learning concepts to obtain high target diagnosis performances even under insufficient data and noisy data conditions. First, as mentioned in Chapter 4.2, the feature extractor part of the VGG model, a deep CNN model learned using abundant ImageNet data as a source domain, is transferred as a pre-trained model to the target domain. Through this, robust feature learning for fault diagnosis of the mechanical systems under the issue of insufficient data is possible. Next, as mentioned in Chapter 4.3, discriminative feature learning is possible by using a semi-hard triplet loss as an additional loss function. Through this, it is possible to improve health state classification performance by learning well-separated features according to the class label. As a result, high diagnosis performance can be obtained even under lack of data issues or noise problems.

Figure 4-8 is a schematic diagram showing the overall learning procedure of the proposed method. First, the feature extractor part of the CNN model learned from the source domain having abundant data is transferred to the target domain. As the pre-trained model, VGG19 shown in Figure 4-4 is utilized. The feature extractor part consisting of a combination of several convolutional and pooling layers, which is the part before the fully connected layer, is transferred. Then, a new classifier for fault diagnosis of the target domain is constructed and attached to the end of the transferred feature extractor part. This classifier part is generally composed of a fully connected neural network, and the number of output nodes is C, the number of health class labels of the target system. In the output layer, the softmax function is

used to outcome the probability values that the input data belongs to each class label. The entire model composed of the transferred feature extractor part and target classifier part is trained based on the categorical cross-entropy loss, by which the conventional fault diagnosis models are trained. This categorical cross-entropy loss can be expressed as follows:

$$L_{cross-entropy} = -\sum_{i=1}^{C} Y^i \, log(p^i) \tag{4.5}$$

where Y^i denotes the value of the i^{th} node of true label vector Y; and p^i denotes the value of the i^{th} node for the output vector p, which is the result of the softmax function. At the same time, our proposed method maximizes the diagnosis performance by increasing the class-wise separability of the learned features based on the additional semi-hard triplet loss. In conclusion, the final loss function of the proposed fault diagnosis method is expressed as:

$$L = L_{cross-entropy} + \lambda L_{triplet} \tag{4.6}$$

where $L_{cross-entropy}$ and $L_{triplet}$ denote the categorical cross-entropy loss and semi-hard triplet loss that can be calculated through Equations (4.5) and (4.2); and λ denotes the balancing parameters for both loss functions. Then, based on the back-propagation and mini-batch gradient descent method, we can train the proposed fault diagnosis models to autonomously learn robust and discriminative features from the training data. In this research, the batch size was set to 32, and the number of training epochs was set to 50. Furthermore, the RMSprop optimizer was utilized with a learning rate of 2e-5. Also, the balancing parameter λ for loss functions was empirically selected as 1. In addition, for the classifier part, a fully connected neural

network that consists of two layers with dimensions of 512 and C, is utilized. Through the proposed fault diagnosis method, it was possible to obtain a fault diagnosis model for mechanical systems with high diagnosis performances under insufficient and noisy data conditions.

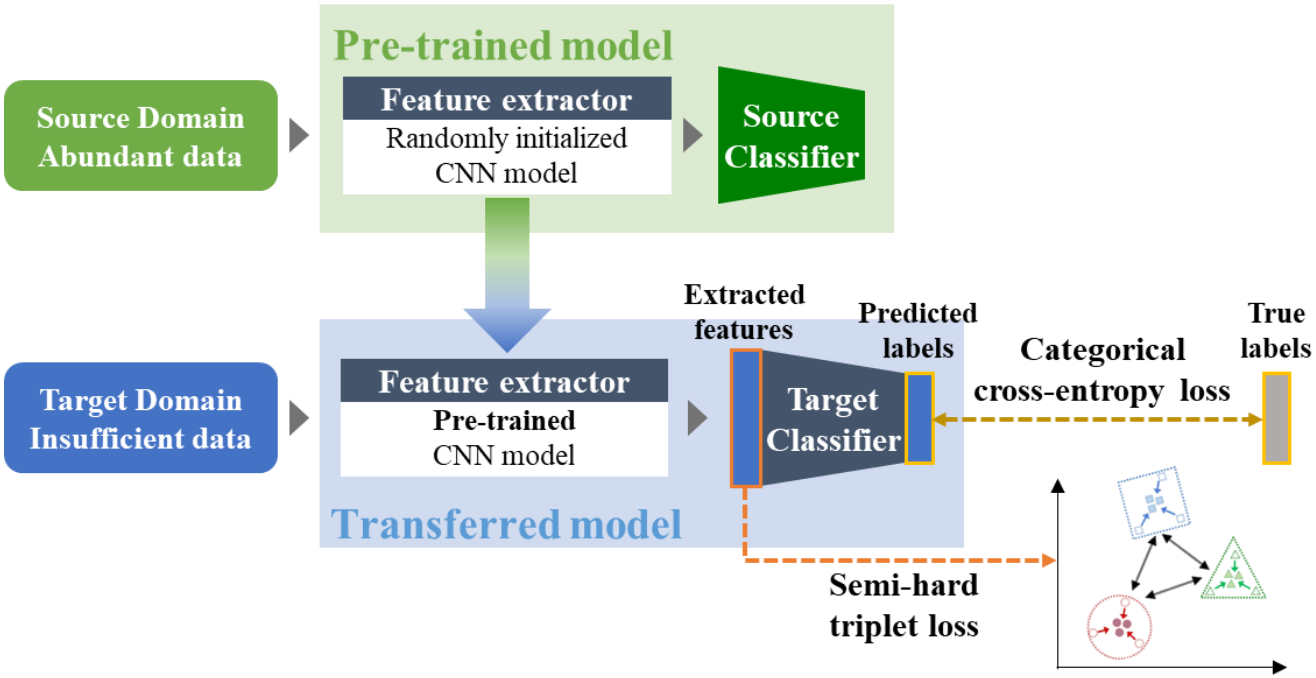


Figure 4-8 Schematic diagram of the proposed fault diagnosis method based on the parameter transfer and semi-hard triplet loss.

4.5 Experimental Studies and Results

4.5.1 Experiment and Data Description

In this research, the vibration signals obtained from the RK4 rotor testbed shown in Figure 3-7 were used to validate the proposed method. The detailed information for the testbed is mentioned in Chapter 3.5.1. The vibration signals used in this study were obtained for five types of health conditions: normal, unbalance, rubbing, misalignment, and oil-whirl. The vibration images containing both spatial and temporal information were generated based on the ODR techniques as explained in Chapter 3.3. In the direction of the time axis, a two-cycle signal consisting of 256 sample points was used. Using real and virtual sensor signals at 64 points along the entire circumferential direction, a 256x64 dimensional vibration image was finally generated and used.

The proposed method was validated using five datasets obtained through five repetitions of the experiments. Each dataset consisted of 400 data for each health state, a total of 2000 data. In addition, in order to effectively verify the performance of the proposed method in consideration of the influence due to environmental factors or randomness, the verification was performed based on all possible training and test data combinations, a total of 20 cases. These combinations of datasets are shown in Table 4-1. The implementations of the proposed method and compared methods were conducted on a desktop computer equipped with an Intel Core i7-8700 CPU, 16 GB of RAM, and NVIDIA GeForce RTX 2080 Ti.

Table 4-1 Combinations of training and testing sets based on five experimental datasets.

Dataset Combination	#1	#2	#3	#4	#4	…	#19	#20
Training Set	1	1	1	1	2	…	5	5
Testing Set	2	3	4	5	1	…	3	4

4.5.2 Compared Methods

To verify the effectiveness of our proposed method based on the parameter transfer and metric learning concepts, the diagnosis performances of other CNN-based methods were compared. First, to validate the effect of the parameter transfer learning scheme, the comparison was performed using the CNN model whose architecture is identical with the proposed method, but the weight parameters are randomly initialized. Moreover, a comparison using the shallower model was also performed. The shallow model consisting of four sets of convolutional and pooling layers was compared, and the detailed information of the used architecture is shown in Table 4-2. In those compared models, for the classifier part, the same structure explained in Chapter 0 was adopted. In addition, to confirm the effectiveness of the semi-hard triplet loss, the diagnosis results of the CNN-based models using $L_{cross-entropy}$ only and the CNN-based model using both loss functions were compared. In summary, five compared methods were analyzed, except for the proposed diagnosis method which adopted both deep pre-trained model and semi-hard triplet loss.

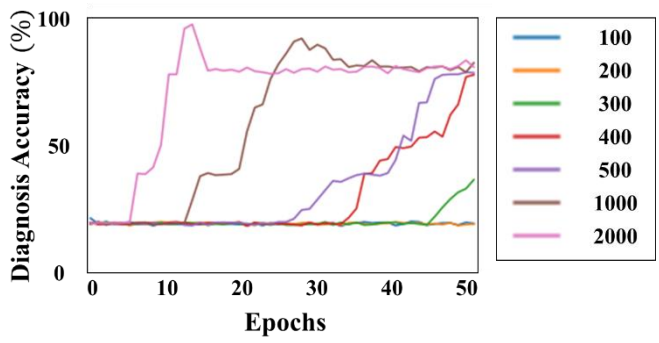
Table 4-2 The configuration of the shallower CNN-based fault diagnosis model.

Shallower CNN model configuration	
Feature extractor part	3x3 conv (32)
	max pool
	3x3 conv (64)
	max pool
	3x3 conv (128)
	max pool
	3x3 conv (128)
	max pool
	flatten
Classifier part	FC (512)
	FC (5)
	Softmax

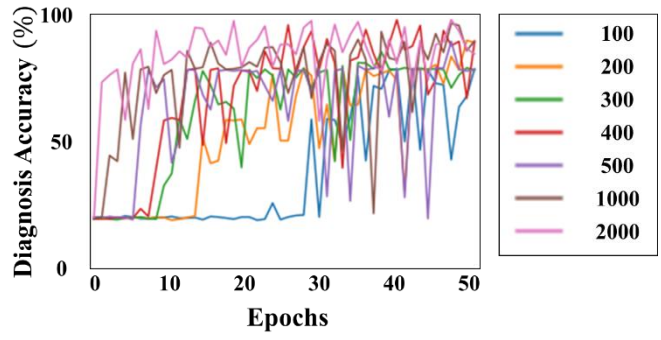
4.5.3 Experimental Results Under Insufficient Data Conditions

In this chapter, the experimental results under insufficient data conditions are presented. First, the learning curves of the proposed and compared methods are shown in Figure 4-9. In this comparison study, the effects of parameter transfer of pre-trained model can be confirmed. As can be seen in this figure, much more stable training can be performed through the proposed method. On the other hand, deep CNN model without parameter transfer and shallower model showed unstable, wiggly training curve. Consequently, the final diagnosis performances of the proposed method are higher than any other compared methods.

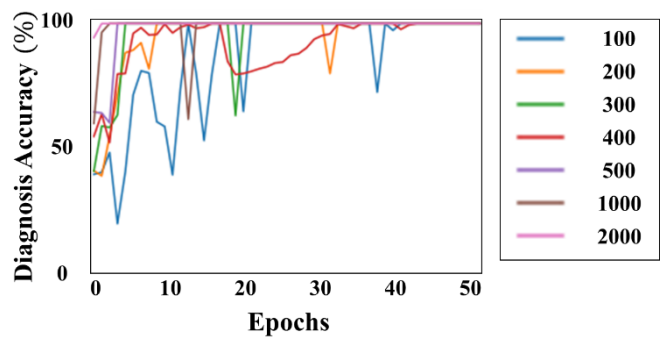
The diagnosis performance results with different numbers of training data are shown in Figure 4-10. The proposed method shows relatively slow performance degradation as the number of training data decreases, compared to other methods. Thanks to robust feature learning based on the transfer of abundant information learned from the source domain, superior diagnosis performances can be acquired by the proposed method. In addition, based on the proposed method which adopted both transfer learning and metric learning concepts, discriminative and semantically well-aligned features were obtained, as shown in Figure 4-11. As a result, we can confirm that accurate and robust fault diagnosis models can be learned based on the proposed method.



(a)



(b)



(c)

Figure 4-9 Learning curves of (a) shallow CNN model; (b) randomly initialized deep CNN model; and (c) pre-trained deep CNN model (proposed).

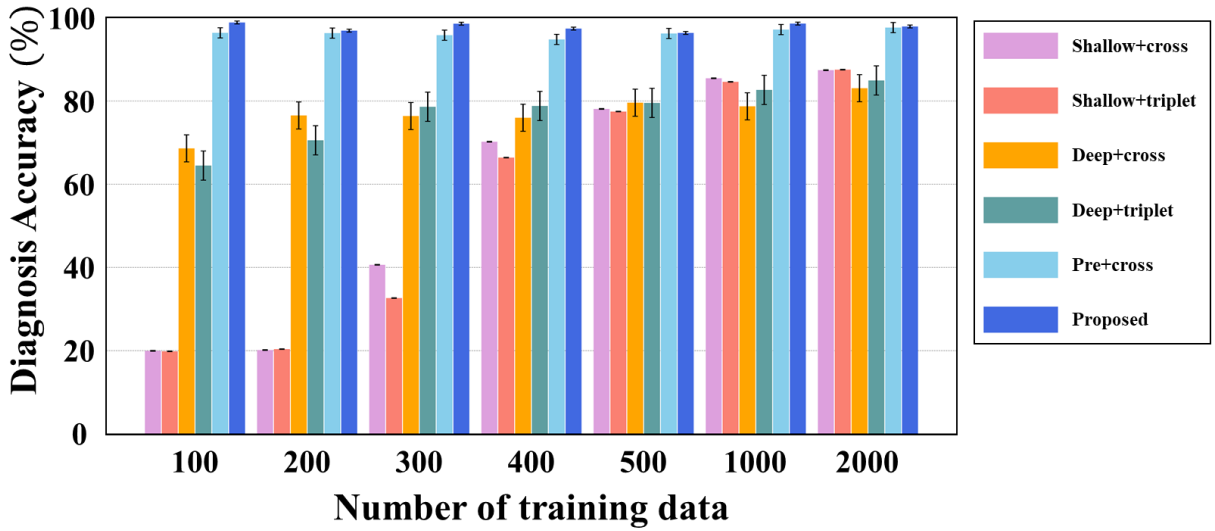


Figure 4-10 The diagnosis performance results of the proposed and compared methods with different numbers of training data.

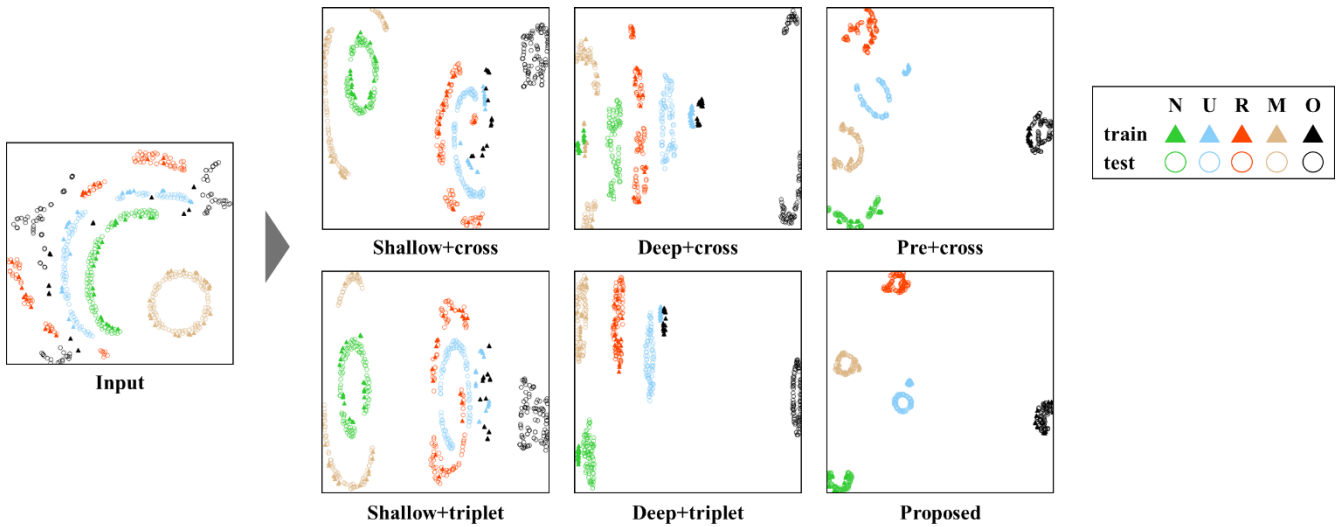


Figure 4-11 Feature visualization using t-SNE under insufficient data condition.

Lastly, the results of comparing the training time of each method are shown in Table 4-3. For deep CNN architectures, the training time required per epoch increases compared to shallower model. However, as shown in Figure 4-9, for the proposed method, the number of training epochs required for the model to converge is much smaller than other methods. Consequently, it can be seen that our proposed method can train the diagnosis model much faster than other methods, and also the final diagnosis performance is much higher.

Table 4-3 Diagnosis accuracy and training time for the proposed and compared methods.

Model type	Shallow CNN	Deep CNN	Pre-trained deep CNN
Diagnosis accuracy (%)	87.4	84.9	**97.9**
Training time (seconds)	2	2	8

4.5.4 Experimental Results Under Noisy Data Conditions

The diagnosis performance results with additive noise are shown in Figure 4-12. The proposed method shows relatively small performance degradation with increasing noises, compared to other methods. This is due to the robust and discriminative feature learning capability of the proposed method based on the parameter transfer and metric learning concepts. In addition, based on the proposed method, semantically well-separated features were obtained under noisy data conditions, as shown in Figure 4-13. Consequently, based on the proposed method, the high diagnosis performances for test data can be achieved under noisy data conditions.

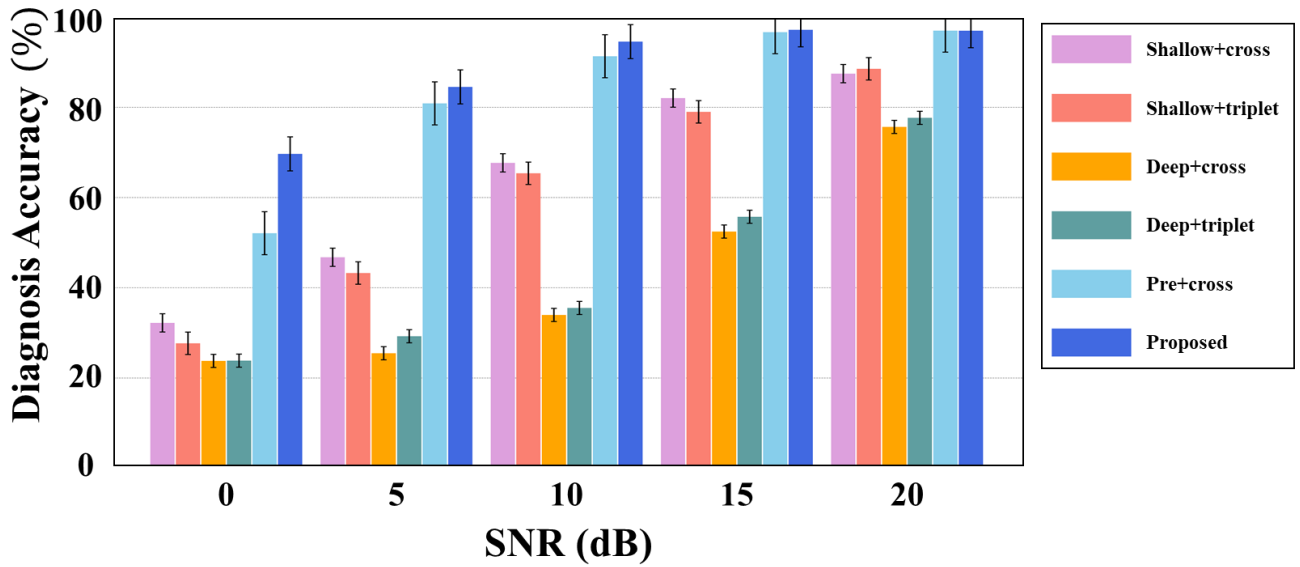


Figure 4-12 The diagnosis performance results of the proposed and compared methods under different levels of additive noise.

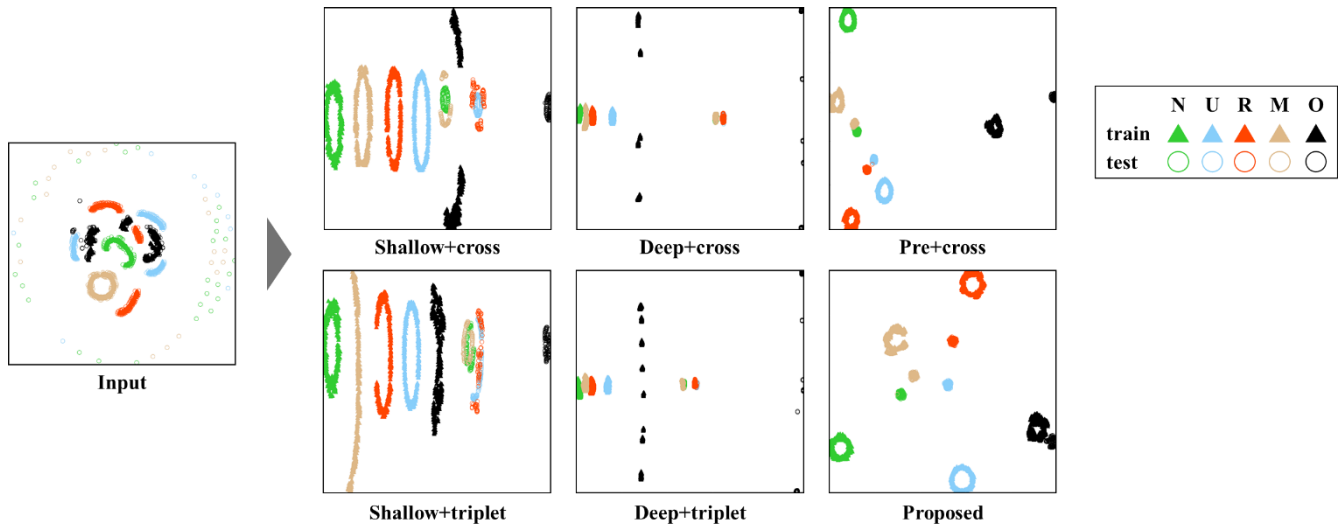


Figure 4-13 Feature visualization using t-SNE under noisy data condition.

4.6 Summary and Discussion

In this research, to obtain high target diagnosis performances even under insufficient data and noisy data conditions, a fault diagnosis method based on the parameter transfer and metric learning concept was proposed. First, the feature extractor part of the VGG model which was learned using abundant ImageNet data is transferred as a pre-trained model to the target domain. In addition, by using a semi-hard triplet loss as an additional loss function, discriminative features that are semantically well-aligned were obtained. As a result, high diagnosis performances can be obtained even under lack of data issues or noise problems. Comparative studies were performed using the various amount of training data and signal-to-noise ratio. For all cases, the proposed method showed the highest diagnosis performances. Also, as can be seen from the learning curve results, the proposed method was able to learn the diagnosis models most reliably and quickly. In addition, through the visualization results of the feature distributions, it was confirmed that the proposed method learned the superior and well-separated features.

Chapter 5

A Domain Adaptation with Semantic Clustering (DASC) Method for Fault Diagnosis

Recently, deep learning based approaches have shown remarkable fault diagnosis performance, thanks to their autonomous feature learning ability which in turn alleviates domain-knowledge-dependent problems [30], [65], [66]. To develop robust fault diagnosis methods using deep learning, sufficient labeled datasets must be obtained from the target system for every health condition of interest as shown in Figure 5-1 [31], [32]. Further, it is useful to point out that data-driven methods, including deep learning, offer satisfactory performance when both the test and training data can be assumed to share a distribution of the same type as shown in Figure 5-2 [33], [34]. However, for real-world rotating machinery, it is challenging to secure sufficient labeled datasets. Moreover, in many cases, the training and test data exhibit the different types of distribution due to environmental noise or changing operating conditions [11], [67]. Furthermore, if training data and test data are obtained from distinct systems, these data sets are more likely to have dissimilar distributions [68], [69]. As a result, it is hard to use conventional supervised learning schemes, as shown in Figure 5-3 (a), to train robust fault diagnosis models with high generalization performance.

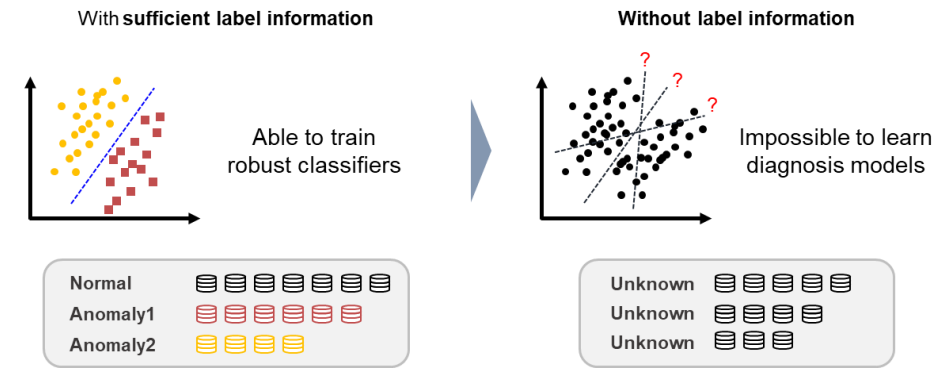


Figure 5-1 The learning results of the fault diagnosis models according to the amount of available label information.

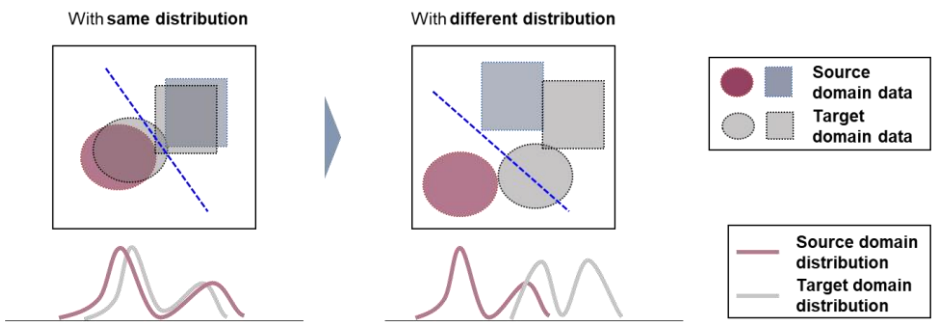


Figure 5-2 The learning results of the fault diagnosis models according to the similarity of the source and target domain distributions.

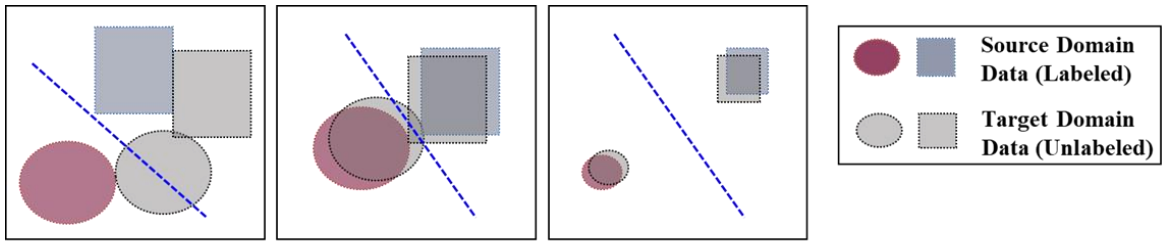


Figure 5-3 Conceptual diagram of distributions for source and target domain data with domain discrepancy in the feature space learned by using (a) the conventional approach, (b) the UDA approach, and (c) the proposed DASC approach. The dotted blue line represents the decision boundary of the learned classifier using labeled source domain data in each feature space. The circles and squares indicate the different classes; the colors indicate the label information. The unlabeled target domain is expressed as a gray color.

In order to overcome those problems and improve target diagnosis performance, unsupervised domain adaptation (UDA) based fault diagnosis methods, which have the ability to convey the knowledge or information gained from the labeled source data into the unlabeled target domain, have been widely studied [70], [71]. The main concept of UDA is to learn common features that can be shared in both domains; this is accomplished by decreasing the domain discrepancy across these two different but related domains. This concept is presented in Figure 5-3 (b). Based on these shared representations, the classifier that was trained using the source data with label information is possible to be adopted in diagnosing the unlabeled target domain. Several types of UDA approaches that are built using deep neural networks were previously studied for fault diagnosis, including discrepancy-based methods [68] and adversarial adaptation methods [72]. However, most of the previous UDA methods focus only on minimizing discrepancy across the marginal distributions of two different domains. In this case, even though the marginal distributions may align well, the diagnosis performance for each class in the target domain may not be satisfactorily generalized [42], [45]. This is because the target samples from each class may not be clustered well, or may not be aligned well with the source data with the same class labels. As a result, the samples located near the boundary of each class might be easily misclassified.

To address the aforementioned issues in UDA and enhance the diagnosis performances for target domain, the work outlined herein suggests a new domain adaptation with semantic clustering (DASC) method to diagnose faults of mechanical systems. By taking advantage of domain adaptation and metric learning concepts, the proposed method learns discriminative and domain-invariant features

that both minimize domain discrepancy and also make the samples from each class semantically well-clustered. To implement this strategy, an additional loss term called semantic clustering loss is proposed; this loss term brings samples that have the same class label closer and causes differently labeled samples to separate. This additional loss term is based on the pairwise distance metric between each labeled sample from the source domain; it is applied at multiple feature levels to obtain robust features with desired properties. In contrast with prior methods, our DASC approach can learn features with not only domain-invariant but also discriminative characteristics, which exhibit significant inter-class distances and minor intra-class distances, as depicted in Figure 5-3 (c). Thus, the proposed approach is able to achieve robust and well-generalized diagnosis models that can yield high diagnosis performances for the target domain, using label information from the source domain.

We validate the efficacy of the DASC approach via various analyses that examine experimental data from three bearing systems. The results indicate that the DASC approach significantly increases the generalized diagnosis performances for mechanical systems, as compared with prior approaches. Also, the results of visualization of the learned feature distributions confirm that the DASC method can obtain discriminative features with better clustering characteristics. The proposed method's efficacy was also confirmed to a further degree through ablation studies; these studies show that better diagnosis performance can be obtained by applying the semantic clustering loss term at multiple feature levels. In addition, by defining an index that evaluates how well the target features are clustered semantically, we verified DASC's ability to make target domain features well-clustered class-wise. These results confirm that the DASC approach can greatly increase the fault

diagnosis performance for target mechanical systems via transferring knowledge obtained from source mechanical systems and making learned features more discriminative.

The remainder of this chapter is constructed as follows. Section 5.1 introduces the concept of unsupervised domain adaptation (UDA). Next, in Section 5.2 CNN-based diagnosis model is described. Then, in Section 5.3 explain the learning scheme of domain-invariant features. The newly devised semantic clustering loss based learning is presented in Section 5.4. Section 5.5 outlines the proposed DASC based fault diagnosis method. Then, Section 5.6 describes the results of experiments and analysis of these results. Lastly, conclusions of this research are discussed in Section 5.7.

5.1 Unsupervised Domain Adaptation

For real-world mechanical systems, it can be expensive and not easy to secure a sufficient amount of labeled data. Also, in many cases, the test data distribution differs from that of the data used for training due to noise, changes in operating conditions, or other factors (e.g., when data is obtained from different systems). In order to develop fault diagnosis methods under these conditions, transfer learning can be utilized [35]. For a clear explanation of transfer learning, notation and definitions are presented here. First, a domain D is comprised of marginal probability $P(X)$ and the feature space $\mathcal{X}$, where $X \in \mathcal{X}$ denotes input datasets. Task $\mathcal{T}$ is composed of predictive function $P(Y|\mathrm{X})$ and label space $\mathcal{Y}$, where $Y \in$

$\mathcal{Y} = \{1, \dots, c\}$ signify the health condition and c represents the number of possible conditions. Transfer learning refers to the learning strategy that uses knowledge or information acquired from a different (but related) source domain D_S to result in a diagnosis model that is applicable to target domain D_T, which is $P(Y_T|\mathrm{X_T})$.

In the paper, we concentrate on the situations where labeled data acquired from D_S is used to improve target diagnosis performance in D_T, where label information cannot be obtained. In other words, the diagnosis models for the target domain, which has only unlabeled data $\{X_T\} = \{x_{T,i}\}_{i=1}^{n_T}$, will be developed by transferring information from the labeled data $\{\mathrm{X_S}, Y_T\} = \{x_{S,i}, y_{S,i}\}_{i=1}^{n_S}$ obtained from different (but related) source domains, where n_S signifies the number of source samples and n_T describes the number of target samples. Different domains can be thought of as data obtained from different systems or under different operating conditions; for those different domains, there exist domain discrepancies or domain shifts (i.e., $D_S \neq D_T$). In the problems dealt with in this section, both domains have the same label space $\mathcal{Y}$ $(= \mathcal{Y}_\mathrm{S} = \mathcal{Y}_\mathrm{T})$; this means that they have the same types of health conditions. This is reasonable because data from similar mechanical systems that has the same types of fault modes is transferred and utilized. These kinds of transfer learning problems are called unsupervised domain adaptation (UDA) problems [70], [71]. Due to domain discrepancies, we cannot apply the diagnosis models trained with labeled datasets obtained from D_S directly to D_T. To solve those problems, UDA algorithms can be utilized to minimize domain discrepancies and to maximize the generalization performance in D_T. The primary goal is to determine common representations of features that can be shared in both domains by minimizing the discrepancy between them. Then, based on those shared features, we can develop

models for fault diagnosis that work well in both domains, using the label information from D_S. This learning strategy can be confirmed by the following formula from [34]:

$$R_{D_T}(\eta) \leq R_{D_S}(\eta) + \hat{d}_{\mathcal{H}}(D_S, D_T) + Const. \tag{5.1}$$

where η denotes the trained classifier that belongs to the hypothesis class $\mathcal{H}$; $R_{D_T}(\eta)$ denotes the target risk with classifier η in D_T, which can be expressed as $\Pr_{X_T \sim D_T}(\eta(X_T) \neq Y_T)$; $R_{D_S}(\eta)$ denotes the source risk with η in D_S; $\hat{d}_{\mathcal{H}}(D_S, D_T)$ denotes the empirical $\mathcal{H}$-divergence across the samples derived from D_S and D_T, which implies the domain discrepancy; and $Const.$ denotes a constant term determined by model complexity and sample size. As can be seen from Equation (5.1), to determine diagnosis models that work well for the target domain, both source risk and domain discrepancy should be minimized at the same time.

There are several types of neural-network-based UDA methods that can be employed, depending on the learning strategy. First, discrepancy-based methods adopt specific metrics, including both Correlation Alignment (CORAL) and Maximum Mean Discrepancy (MMD) [73], [74]. These metrics quantify the discrepancy across the target and source domains. Then, by reducing those discrepancy metrics, domain-invariant features can be obtained. Yang et al. [68] used the MMD-based UDA method to acquire transferable features that could be employed for both bearings examined in the lab and bearings in locomotives. Second, adversarial adaptation approaches gain common features based on adversarial training schemes, in which the domain classifier and the feature extractor are in competition with each other [75]. Han et al. [72] proposed deep adversarial CNN,

which is a minimax-objective-based adversarial adaptation method, to diagnose gearbox and wind turbine faults. Furthermore, Guo et al. [69] obtained shared feature representations for bearing diagnosis by proposing a deep convolutional transfer learning network (DCTLN); This approach adopted adversarial and discrepancy-based strategies simultaneously.

5.2 CNN-based Diagnosis Model

The general goal of the diagnosis technique is to gain a good fault diagnosis model that correctly matches input data X to corresponding label vector Y. To develop a diagnosis model, we conduct supervised learning to minimize the classification loss, or error between predicted and true labels of the labeled training data. In our paper, CNN is adopted to acquire an intelligent method to diagnose faults in mechanical systems. As explained in Chapter 2.1.2, CNN consists of a feature extractor part, which is built upon sets of convolutional layers and pooling layers, and a classifier part, which is based on fully connected layers. Through the feature extractor and classifier, input data X is transformed into output vector $\hat{Y}$, as follows:

$$\hat{Y} = f_{\theta_c}\left[f_{\theta_f}(X)\right] \tag{5.2}$$

where θ_f and θ_c denote the parameters within the feature extractor and classifier, respectively; f_{θ_f} and f_{θ_c} denote the transformation function through the feature extractor and classifier. For multi-class diagnosis models, the dimension of $\hat{Y}$, which represents the number of nodes in the output layer, should be the same as the quantity of health states to be diagnosed. Then, a softmax function is adopted to output the

probability vector p, as follows:

$$p(\hat{Y}) = \begin{bmatrix} p^1 \\ \vdots \\ p^j \\ \vdots \\ p^C \end{bmatrix} = \frac{1}{\sum_{k=1}^{C} \exp(\hat{Y}^k)} \begin{bmatrix} \exp(\hat{Y}^1) \\ \vdots \\ \exp(\hat{Y}^j) \\ \vdots \\ \exp(\hat{Y}^C) \end{bmatrix} \tag{5.3}$$

where C denotes the number of health conditions of the target rotating machinery; and p^j and $\hat{Y}^j$ signify the values of the j^{th} node for vector p and $\hat{Y}$, respectively. The p^j value can be interpreted as the probability that the corresponding data X belongs to each j^{th} health condition among the C classes. The cross-entropy loss is adopted for estimating the classification loss (L_C) to learn the diagnosis model, and this can be represented as follows:

$$L_C = -\sum_{k=1}^{C} Y^k \, log(p^k) \tag{5.4}$$

where Y^k represents the value of the k^{th} node of true target label Y. By reducing the classification loss defined in this way, it is possible to develop a good diagnosis model that classifies the labeled source domain data well to the correct labels.

5.3 Learning of Domain-invariant Features

For solving UDA problems by transferring a classifier gained using source data to the target domain, domain-invariant representations are required. To learn the common features that can be shared across D_S and D_T, the discrepancy-based method is adopted. The strategy of this learning scheme is, first, to define a metric

that quantifies the difference between the two domains and then, to train the features in a way that minimizes that metric. Here, maximum mean discrepancy (MMD) is employed for estimating a discrepancy across distributions of different domains. This nonparametric measure can be determined as being the greatest difference in expectation values over functions in the universal reproducing kernel Hilbert space (RKHS) [76]. We can express this as:

$$MMD(p_S, p_T, H) \coloneqq \sup_{f \in H} \left(\underset{x_S \sim p_S}{E}[f(x_S)] - \underset{x_T \sim p_T}{E}[f(x_T)] \right) \tag{5.5}$$

where H denotes the universal RKHS; f denotes the functions within the function class, which is the unit ball in H; and x_S and x_T denote the random variables from the probability distributions, p_S and p_T. Substituting the population expectations with empirical expectations calculated for the samples, we can find an empirical estimate of MMD:

$$\widehat{MMD}(X_S, X_T, H) \coloneqq \sup_{f \in H} \left(\frac{1}{n_S} \sum_{i=1}^{n_S} f(x_{S,i}) - \frac{1}{n_T} \sum_{i=1}^{n_T} f(x_{T,i}) \right) \tag{5.6}$$

where $X_S = (x_{S,1}, \dots, x_{S,n_S})$ and $X_T = (x_{T,1}, \dots, x_{T,n_T})$ denote the samples drawn from distributions p_S and p_T; n_S and n_T denote the number of samples X_S and X_T. Based on the kernel mean embedding in RKHS, empirical estimation of squared MMD can be obtained in terms of the kernel functions, as follows [76]:

$$\widehat{MMD}^2(X_S, X_T, H)$$

$$= \frac{1}{{n_S}^2} \sum_{i,j=1}^{n_S} k(x_{S,i}, x_{S,j}) - \frac{2}{n_S n_T} \sum_{i,j=1}^{n_S, n_T} k(x_{S,i}, x_{T,j}) \quad (5.7)$$

$$+ \frac{1}{{n_T}^2} \sum_{i,j=1}^{n_T} k(x_{T,i}, x_{T,j})$$

where $k(x, x') \coloneqq < \phi(x), \phi(x') >_H$ denotes the kernel function, which is defined as the inner product between feature mappings, ϕ, from $\mathcal{X}$ to H. In this research, the Gaussian kernels $k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$, which are proven to be universal kernels [77], are used to calculate empirical MMD through Equation (5.7). MMD depends on the value of σ^2, as does the UDA result. In this research, we employed multi-kernel MMD, which leverages different kernels with different σ^2 values, to enhance the robustness and performance of the MMD-based UDA methods [78]. Finally, the domain-related loss (L_D) between the two domains (source and target) can be obtained using this MMD metric. L_D can be calculated as the MMD between outputs of the feature extractor part for data from both domains, correspondingly $f_{\theta_f}(X_S)$ and $f_{\theta_f}(X_T)$. This domain-related loss is essential for learning domain-invariant features to conduct UDA tasks.

5.4 Domain Adaptation with Semantic Clustering

Conventional UDA methods aim to discover features that are domain-invariant by reducing both classification loss and domain-related loss. However, these

conventional UDA methods only consider alignment across the marginal distributions for different domain data; they do not consider the separability of features according to their classes. Thus, discriminative features with significant inter-class differences and minor intra-class variations cannot be learned consistently; this results in low target domain generalization performance. To cope with this, we propose a metric-learning based approach to learn more discriminative representations that make samples from each class semantically well-clustered. This can be achieved by mapping the data from the same class closely together, while making the differently classed data sufficiently separate. For this purpose, we propose an additional loss term, called semantic clustering loss (L_{SC}), which can be expressed as follows:

$$L_{SC} = \sum_{k=1}^{n_f} \lambda_k \left[\sum_{i,j=1}^{n_s} \left\{ \left\| g_i^k - g_j^k \right\|_2^2 \cdot I_{ij} + max\left(0,\, d - \left\| g_i^k - g_j^k \right\|_2\right)^2 \cdot \left(1 - I_{ij}\right) \right\} \right] \tag{5.8}$$

g_i^k represents the feature vector of the i^{th} source sample, $x_{S,i}$, at the k^{th} feature level; I_{ij} denotes the value of the matrix I, which is 1 if the i^{th} and j^{th} samples are in the same class, and 0 for those that are in a different class; n_f denotes the number of feature layers to which L_{SC} is applied; λ_k denotes the balancing parameter for L_{SC} at each k^{th} feature level, which is determined experimentally; n_s represents the quantity of source samples; and d signifies the value that controls the minimum distance between two samples from dissimilar classes. Equation (5.8) shows that, L_{SC} is calculated based on the pairwise similarity distances between all

source samples to obtain robust and semantically well-clustered features. By reducing the L_{SC} during the training process, for sample pairs from the same class ($I_{ij} = 1$), the pairwise distances will be reduced; for sample pairs from different classes ($I_{ij} = 0$), the distances will be increased. This similarity-metric-based learning is performed using the source data because, under the UDA problem, only source-domain label information is available. Since the two domains are related, this learning strategy using semantic clustering loss for the source data will make target data also clustered well by each class. Furthermore, in our proposed method, the L_{SC} is applied at multiple feature levels. As a result, more discriminative feature representations can be obtained, which make samples semantically clustered better according to their health conditions. Thanks to this semantic clustering characteristic, the misclassification rate for the target samples located near the boundary of each class is decreased. In conclusion, the diagnosis performance in D_T can be improved through the use of our method.

5.5 Proposed DASC-based Fault Diagnosis Method

Through our proposed DASC-based fault diagnosis method for mechanical systems, we can acquire an accurate and robust diagnosis model for the target domain without any label information [79]. The data flow and architecture of DASC are provided in Figure 5-4. First, the feature extractor part is made up of two groups of convolutional and pooling layers. For every convolutional layer, 1D kernels with zero-padding and ReLU are adopted; for every pooling layer, max-pooling is used. Next, for the classifier part, a fully connected network, comprised of two layers with

dimensions of 256 and C, is used, where C describes the number of health conditions of the rotating machinery. The final objective function of DASC is expressed as:

$$L = L_C + \lambda_D L_D + \lambda_{SC} L_{SC} \tag{5.9}$$

where L_C, L_D, and L_{SC} denote the classification loss, domain-related loss, and semantic clustering loss, respectively; λ_D and λ_{SC} denote the balancing parameters for L_D and L_{SC}. Based on the labeled source data, L_C is calculated only at the output layer of the classifier, according to Equation (5.4); also, L_{SC} is calculated at every pooling layer within the feature extractor, through the use of Equation (5.8). Based on Equation (5.7), L_D, which should be minimized to acquire features that are domain-invariant, is calculated as the MMD between the features from D_S and D_T, which are found as the outcomes of the last convolutional and pooling layers, as can be seen in Figure 5-4. Through minimizing this final objective function, fault diagnosis models learned using label information of D_S can be employed in D_T, and more discriminative features can be learned by making the features semantically well-clustered according to their classes. In addition, one thing to keep in mind is that although L_C and L_{SC} both use the label information from source domain data, their purposes and benefits are different. L_C only considers the location of the labeled data relative to the hyperplane of the classifier. That is, it aims to make the model correctly classify the source domain labeled data. However, L_{SC} considers the labeled data's distribution, the relative distances between samples according to their classes, which results in semantically better clustered features. As a result, based on the proposed DASC method, which considers not only L_D and L_C but also L_{SC}, we can achieve fault diagnosis models with high target

generalization performances. In this paper, we employ the mini-batch gradient descent scheme with momentum to train the diagnosis model. Here, a learning rate is set as $\frac{0.01}{(1+10\times p)^{0.75}}$, which decreases as training progress p increases from 0 to 1 [75]. As σ^2 values for calculating the multi-kernel MMD, powers of 10 between 1e-6 and 1e6 are used. The balancing parameters λ_k, λ_D, and λ_{SC} in Equations (5.8) and (5.9) were chosen within the range of 1e-5 to 1e-1. In addition, more detailed information on the parameters used in the proposed method is presented in Table 5-1.

Table 5-1 Parameter information of the proposed method.

Parameter	Value	Parameter	Value
Epoch	150	Kernel number	20
Batch size	64	Kernel size	3
momentum	0.9	Stride	1
n_f	2	Pooling window size	2
d	100	Pooling stride	2

A flowchart of our DASC-based method of fault diagnosis is shown in Figure 5-5. As explained above, labeled source data is employed to obtain fault diagnosis models applicable in the different domain which is unlabeled. By minimizing the final objective function, discriminative domain-invariant representations can be secured. Thus, target domain fault diagnosis performance is improved.

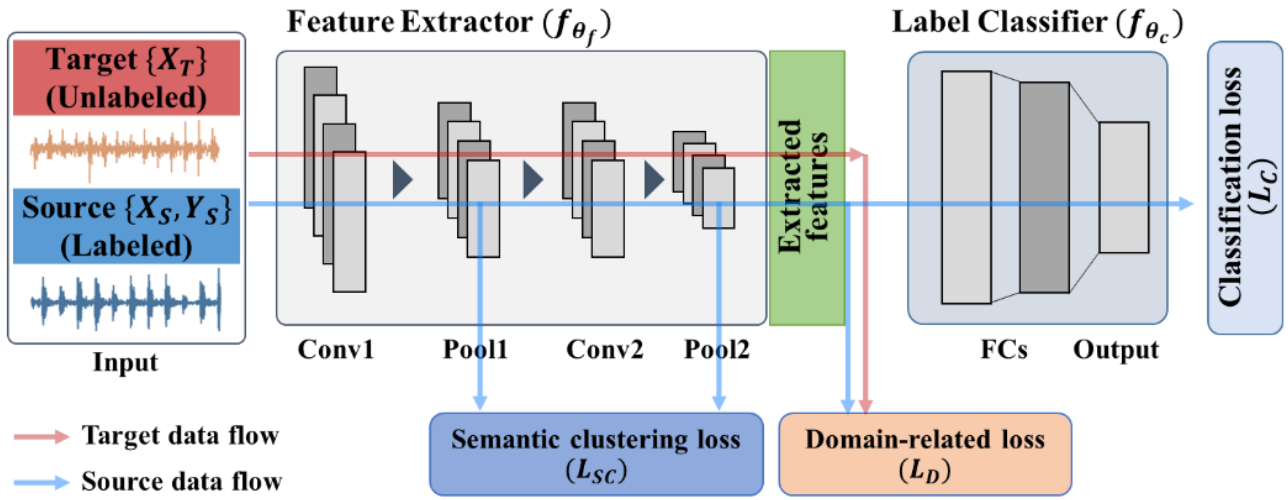


Figure 5-4 The architecture of the DASC-based diagnosis approach.

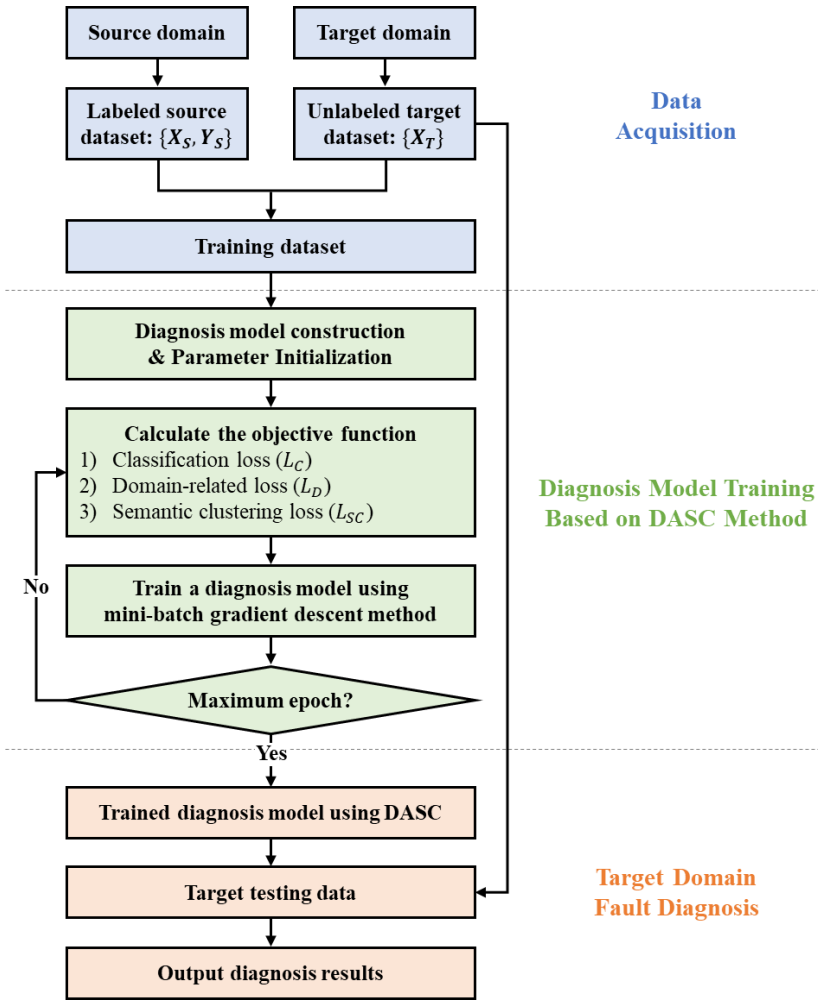


Figure 5-5 Flowchart of the proposed DASC-based fault diagnosis method.

5.6 Experimental Studies and Results

5.6.1 Experiment and Data Description

In this research, bearing data acquired from the systems displayed in Figure 5-6 was used to validate the DASC method. First, we examine a ball-bearing data offered by the Bearing Data Center in Case Western Reserve University (CWRU) [80]. Second, a spherical roller-bearing dataset furnished by the Center for Intelligent Maintenance Systems (IMS) [81] was studied. Finally, a ball-bearing dataset offered by the Xi'an Jiaotong University, Institute of Design Science and Basic Component, and Changxing Sumyoung Technology Co., Ltd. (XJTU-SY) [82] was used. The acceleration signals were obtained from four types of health states: normal (N), ball fault (B), inner raceway fault (IR), and outer raceway fault (OR). For CWRU, faults with diameters 0.007, 0.014, and 0.021 inches were induced to the bearing through electric-discharge machining. For IMS and XJTU-SY, fault data was obtained by run-to-failure experiments. In addition, CWRU and XJTU-SY data were obtained under variable operating conditions. More detailed specifications of the datasets for each type of bearing studied are introduced in Table 5-2.

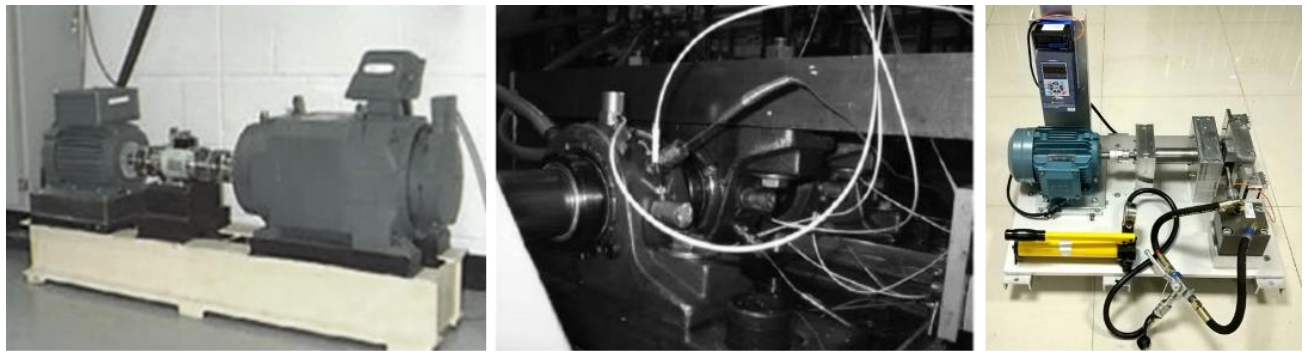


Figure 5-6 Experimental testbed: (a) CWRU testbed; (b) IMS testbed; and (c) XJTU-SY testbed.

Table 5-2 Detailed information about bearing datasets.

Datasets	Sampling rate (kHz)	Rotating speed (RPM)	Load	Health conditions
CWRU1	12	1797	0 (hp)	N / B / IR / OR
CWRU2	12	1772	1 (hp)	N / B / IR / OR
CWRU3	12	1750	2 (hp)	N / B / IR / OR
CWRU4	12	1730	3 (hp)	N / B / IR / OR
IMS	20	2000	26.6 (kN)	N / B / IR / OR
XJTU-SY1	25.6	2100	12 (kN)	N / IR / OR
XJTU-SY2	25.6	2250	11 (kN)	N / IR / OR
XJTU-SY3	25.6	2400	10 (kN)	N / B / IR / OR

Two UDA-based fault diagnosis scenarios were considered to demonstrate DASC's effectiveness. First, the UDA problems between domains with discrepancies caused by different operating conditions are considered in Section 5.6.3. The second scenario describes the development of diagnosis models when source data and target data are collected from distinct mechanical systems; this is explained in Section 5.6.4. For all experiments, the dataset for each health condition consists of 4000 samples of raw vibration signals whose length is 1200 points. Among them, 70% of data samples were dedicated to training; 30% were assigned for use in the test step to estimate the diagnosis performances. Training data from both domains were used to learn the UDA-based fault diagnosis models. Target test samples were used to assess the target domain generalization performance of these UDA-based diagnosis models.

5.6.2 Compared Methods

Target domain diagnosis performances of both DASC and existing methods were examined to validate DASC's effectiveness. First, a CNN trained using source data only was compared for verification of the effectiveness of UDA algorithms. Then, several existing neural-network-based UDA algorithms used for fault diagnosis were compared. Domain-adversarial neural network (DANN) [75] was used as a representative method of an adversarial adaptation approach. For a discrepancy-based method, Deep CORAL [74], which learns the desirable features by reducing the difference between covariances of features in two domains, was compared. Also, the method to minimize the multi-kernel MMD across source features and target features was compared. The method that uses both adversarial

and discrepancy-based strategies simultaneously was also compared; this method is abbreviated as Adv+Disc in the remaining parts. In order to fairly and effectively compare the generalization ability by relying solely on the UDA method used, the same CNN architecture described in Section 5.5 was adopted as the backbone model for all compared methods. Furthermore, the domain classifier for the adversarial adaptation strategy consists of a two-layer, fully connected network with 256 and 2 nodes. The balancing parameter for DANN was chosen from {0,0.5,1.0,1.5,2.0}, and for MMD, the parameter was chosen within the range of 1e-5 to 1e-1. For the method using both adversarial and discrepancy-based strategies simultaneously, the same parameters as used for DANN and MMD were employed, and for Deep CORAL, the parameter was selected within the range of 1e-6 to 1e-2. In addition, for fair comparisons, all compared UDA methods were performed under the same conditions, by using not only the same training and test data but also by using the same implementation details, such as optimizer, learning rate, training epoch, batch size, and momentum. Also, testing of each method was accomplished on the same computer, which was equipped with an NVIDIA GeForce RTX 2080 Ti, Intel i7-8700, and 16 GB of RAM.

5.6.3 Scenario I: Different Operating Conditions

The experimental results of UDA between domains with various operating conditions, (e.g., load and rotational speed) are described in this section. For performance verification of DASC under different operating conditions, the CWRU and XJTU-SY datasets, in which vibration signals were obtained under variable operating conditions as explained in Table 5-2, are used. First, for the CWRU dataset,

we evaluated the proposed and compared methods across all twelve UDA tasks with different source and target domains, which are $1 \rightarrow 2/3/4$, $2 \rightarrow 1/3/4$, $3 \rightarrow 1/2/4$, and $4 \rightarrow 1/2/3$, where each number indicates the dataset with different operating condition presented in Table 5-2. In these experiments, to focus on the discrepancies caused by different operating conditions, the variances due to the fault sizes within each health condition were ignored; this means that samples from the same health condition with different fault sizes were grouped into the same classification category. For these datasets, there are four types of health conditions (N, B, IR, OR); therefore, C is set as 4. The diagnosis accuracies for the target samples under differing conditions of operation, using the dataset provided by CWRU, are presented in Figure 5-7. The values in the figure show the mean accuracy values for three different target tasks with the same source domain; the error bars display the standard error values. Compared with other methods, DASC (our new approach) attains the best diagnosis performance accuracy in all tasks. It is important to note that there are a few cases in which high performance is achieved even if only the source data was used. However, even in these cases, the proposed method always improved the target diagnosis performance consistently, unlike alternative approaches. In general, the DASC approach showed superior performances across all situations. To effectively confirm the generalization capability of DASC, detailed diagnosis performance results for all twelve UDA tasks with different source domains and target domains are presented in Table 5-3.

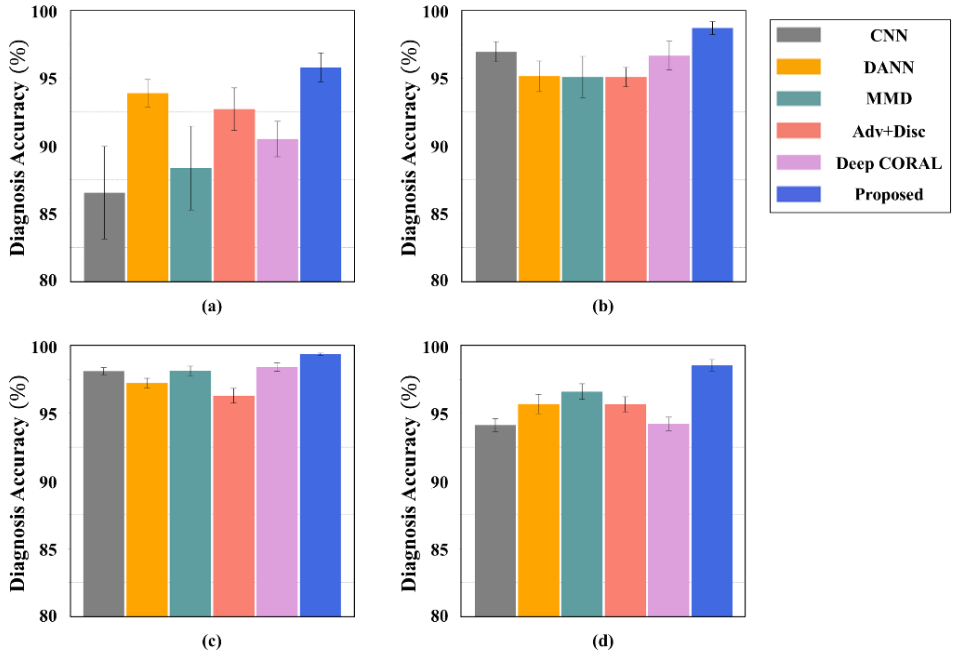


Figure 5-7 Target diagnosis results for scenario I with the CWRU dataset. Mean accuracy values for tasks (a) 1→2/3/4; (b) 2→1/3/4; (c) 3→1/2/4; and (d) 4→1/2/3.

Table 5-3 Average diagnosis accuracy (%) for scenario I with the CWRU datasets.

Source → Target	**CNN**	**DANN**	**MMD**	**Adv+Disc**	**Deep CORAL**	**Proposed**
$1 \rightarrow 2$	97.5 ± 0.15	96.9 ± 0.41	98.5 ± 0.19	95.8 ± 0.52	95.3 ± 0.16	$\mathbf{99.3 \pm 0.06}$
$1 \rightarrow 3$	89.2 ± 0.53	94.8 ± 0.50	89.3 ± 1.10	93.7 ± 0.88	90.3 ± 0.49	$\mathbf{96.2 \pm 0.99}$
$1 \rightarrow 4$	72.9 ± 0.99	89.9 ± 0.46	77.3 ± 3.06	88.7 ± 3.47	85.9 ± 0.48	$\mathbf{91.9 \pm 0.39}$
$2 \rightarrow 1$	98.8 ± 0.11	97.9 ± 0.64	98.5 ± 0.10	96.5 ± 0.60	98.2 ± 0.35	$\mathbf{99.5 \pm 0.19}$
$2 \rightarrow 3$	98.1 ± 0.21	96.7 ± 0.11	97.6 ± 0.16	96.3 ± 0.77	99.4 ± 0.14	$\mathbf{99.8 \pm 0.01}$
$2 \rightarrow 4$	93.8 ± 0.08	90.8 ± 1.21	89.2 ± 1.91	92.5 ± 0.49	92.4 ± 0.66	$\mathbf{96.8 \pm 0.45}$
$3 \rightarrow 1$	97.5 ± 0.31	96.1 ± 0.27	97.6 ± 0.24	94.8 ± 0.44	97.3 ± 0.11	$\mathbf{99.1 \pm 0.13}$
$3 \rightarrow 2$	99.0 ± 0.08	97.5 ± 0.55	99.5 ± 0.04	96.7 ± 0.98	99.6 ± 0.03	$\mathbf{99.7 \pm 0.02}$
$3 \rightarrow 4$	97.8 ± 0.29	98.1 ± 0.34	97.2 ± 0.16	97.4 ± 0.52	98.3 ± 0.11	$\mathbf{99.3 \pm 0.03}$
$4 \rightarrow 1$	93.0 ± 0.26	94.3 ± 1.23	94.4 ± 0.56	95.1 ± 1.07	93.3 ± 0.10	$\mathbf{96.9 \pm 0.25}$
$4 \rightarrow 2$	93.5 ± 0.53	95.1 ± 0.91	97.2 ± 0.08	94.4 ± 0.23	93.1 ± 0.45	$\mathbf{98.9 \pm 0.10}$
$4 \rightarrow 3$	95.8 ± 0.48	97.7 ± 0.33	98.2 ± 0.05	97.4 ± 0.18	96.2 ± 0.29	$\mathbf{99.9 \pm 0.02}$
Average	93.9	95.5	94.6	94.9	95.0	**98.1**

Next, we evaluated diagnosis performance of different UDA methods using the XJTU-SY dataset. For this dataset, there are six available tasks, including $1 \rightarrow 2/3$, $2 \rightarrow 1/3$, and $3 \rightarrow 1/2$. For these tasks, C is set as 3, since there are only three kinds of health conditions (i.e., N, IR, and OR) available in the XJTU-SY1 and 2 datasets, as shown in Table 5-2. The target diagnosis performances under different operating conditions for XJTU-SY are displayed in Figure 5-8. The accuracy values in the figure represent the average values for two different tasks with the same source domain; the error bars display standard error values. In examining the XJTU-SY dataset, as well, DASC – which adopts the semantic clustering loss term – consistently improved the target diagnosis performances. Detailed diagnosis results for all six UDA tasks are shown in Table 5-4. As seen in the results, for most cases, DASC leads to the best performances among the several UDA methods studied.

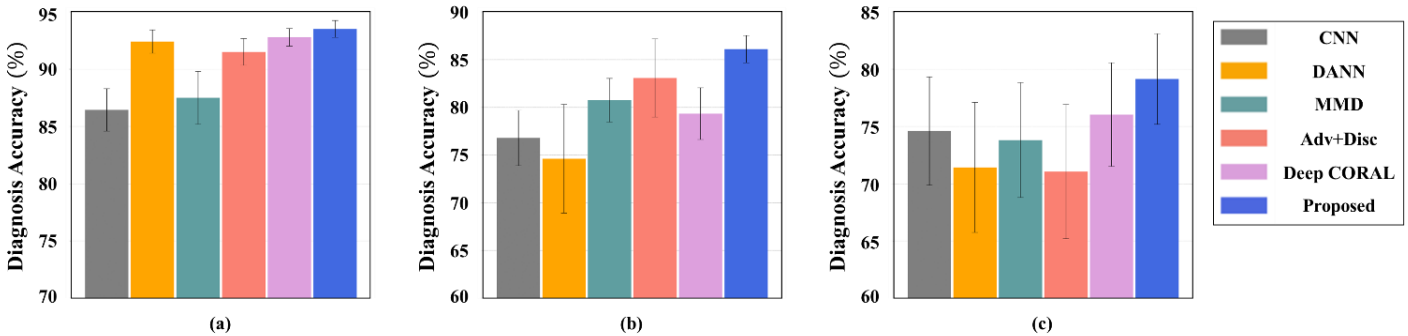


Figure 5-8 Target diagnosis results for scenario I with the XJTU-SY dataset. Mean accuracy values for tasks (a) 1→2/3; (b) 2→1/3; and (c) 3→1/2.

Table 5-4 Average diagnosis accuracy (%) for scenario I with the XJTU-SY datasets.

Source → Target	**CNN**	**DANN**	**MMD**	**Adv+Disc**	**Deep CORAL**	**Proposed**
$1 \rightarrow 2$	94.7 ± 0.02	96.5 ± 0.15	96.3 ± 0.17	95.3 ± 0.80	96.2 ± 0.02	$\mathbf{96.6 \pm 0.07}$
$1 \rightarrow 3$	78.2 ± 0.25	88.3 ± 0.82	78.7 ± 2.36	87.7 ± 1.40	89.3 ± 0.02	$\mathbf{90.6 \pm 0.67}$
$2 \rightarrow 1$	63.8 ± 0.19	56.8 ± 8.11	70.7 ± 0.81	74.5 ± 7.23	67.3 ± 0.39	$\mathbf{81.6 \pm 2.04}$
$2 \rightarrow 3$	89.7 ± 0.12	$\mathbf{91.8 \pm 0.56}$	90.8 ± 0.06	91.6 ± 0.44	91.4 ± 0.03	91.1 ± 0.15
$3 \rightarrow 1$	53.4 ± 0.09	46.2 ± 1.16	51.5 ± 0.50	45.5 ± 1.32	55.9 ± 0.25	$\mathbf{61.6 \pm 0.76}$
$3 \rightarrow 2$	95.7 ± 0.09	96.6 ± 0.47	96.2 ± 0.20	96.7 ± 0.33	96.2 ± 0.25	$\mathbf{96.7 \pm 0.24}$
Average	79.2	79.5	80.7	81.9	82.7	**86.3**

5.6.4 Scenario II: Different Rotating Machinery

This section describes the results of UDA tasks between domains obtained from different rotating machinery. First, the UDA tasks between the CWRU and IMS datasets were considered. Each dataset has four health conditions; therefore, C needs to be 4. To confirm the generalization performance and robustness of DASC thoroughly, validations were conducted on all UDA tasks between CWRU data obtained from four operational states (see Table 5-2) and IMS data. Therefore, we evaluated all methods across 8 UDA tasks, as can be seen in Table 5-5. Since the IMS fault dataset was obtained under run-to-failure experiments, CWRU data with the smallest fault size was considered for those tasks. Thorough results are provided in Table 5-5; average diagnosis performances for $\mathrm{CWRU} \rightarrow \mathrm{IMS}$ tasks and $\mathrm{IMS} \rightarrow \mathrm{CWRU}$ tasks can be seen in Figure 5-9. The values in the table represent the mean accuracy, as well as the standard error values of ten trials for each task; Figure 5-9 also shows mean and standard error values of diagnosis performances for each $\mathrm{CWRU} \rightarrow \mathrm{IMS}$ and $\mathrm{IMS} \rightarrow \mathrm{CWRU}$ task. In most cases, it was confirmed that – by utilizing UDA methods – the target diagnosis performances are enhanced over those achieved by a CNN model trained only with source data. In addition, DASC provides better performance than existing methods, for all cases.

Table 5-5 Average diagnosis accuracy (%) for scenario II between the CWRU and IMS datasets.

Source → Target	**CNN**	**DANN**	**MMD**	**Adv+Disc**	**Deep CORAL**	**Proposed**
CWRU1 → IMS	72.9 ± 1.75	81.7 ± 0.85	86.5 ± 0.02	82.6 ± 1.27	84.3 ± 0.01	$\mathbf{90.1 \pm 0.03}$
CWRU2 → IMS	76.6 ± 0.00	81.5 ± 1.41	86.0 ± 0.03	84.9 ± 1.16	85.3 ± 0.26	$\mathbf{89.6 \pm 0.00}$
CWRU3 → IMS	78.0 ± 0.41	79.6 ± 2.03	87.2 ± 0.44	82.2 ± 1.78	86.4 ± 0.30	$\mathbf{89.1 \pm 0.05}$
CWRU4 → IMS	73.4 ± 0.00	74.3 ± 2.77	86.6 ± 0.00	72.3 ± 3.43	83.7 ± 0.08	$\mathbf{90.1 \pm 0.09}$
Average	75.2	79.3	86.5	80.5	84.9	**89.7**
IMS → CWRU1	47.2 ± 0.92	56.4 ± 3.06	66.4 ± 2.77	59.0 ± 2.42	60.4 ± 3.13	$\mathbf{73.5 \pm 0.18}$
IMS → CWRU2	49.2 ± 0.08	56.8 ± 3.20	68.1 ± 2.41	55.9 ± 1.97	55.8 ± 2.75	$\mathbf{72.5 \pm 0.90}$
IMS → CWRU3	50.1 ± 0.14	54.3 ± 2.04	64.5 ± 2.60	58.4 ± 2.57	49.2 ± 1.47	$\mathbf{70.3 \pm 1.01}$
IMS → CWRU4	50.1 ± 0.00	57.1 ± 2.75	72.0 ± 0.98	57.9 ± 2.28	50.5 ± 0.12	$\mathbf{74.7 \pm 0.65}$
Average	49.1	56.2	67.8	57.8	54.0	**72.8**

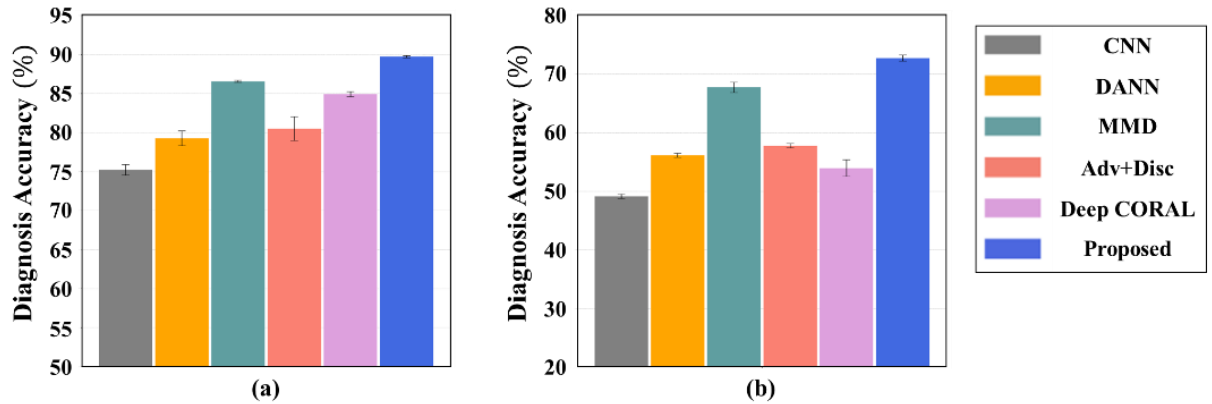


Figure 5-9 Target diagnosis results for scenario II between the CWRU and IMS datasets. Mean accuracy values for tasks (a) CWRU→IMS; (b) IMS→CWRU.

Next, the UDA tasks between two different datasets obtained through the run-to-failure experiments, which are IMS and XJTU-SY, were considered. In this case, to match the number of health conditions C between the two datasets, XJTU-SY3, which has four fault types (the same as the IMS data) was used. The target diagnosis results between the IMS and XJTU-SY datasets are presented in Figure 5-10. The average and standard error values for ten trials are shown in this figure. As seen in the prior cases, these outcomes confirm that DASC, proposed herein, shows the best performance among the many UDA methods. Detailed diagnosis results can be found in Table 5-6.

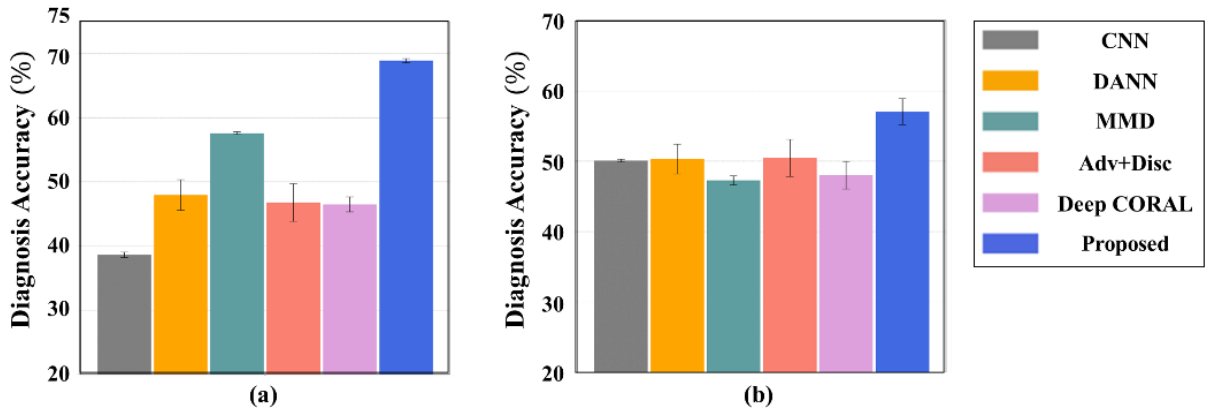


Figure 5-10 Target diagnosis results for scenario II between the IMS and XJTU-SY datasets. Mean accuracy values for tasks (a) XJTU-SY→IMS; (b) IMS→XJTU-SY.

Table 5-6 Average diagnosis accuracy (%) for scenario II between the XJTU-SY and IMS datasets.

Source → Target	**CNN**	**DANN**	**MMD**	**Adv+Disc**	**Deep CORAL**	**Proposed**
XJTU-SY → IMS	38.6 ± 0.43	48.0 ± 2.38	57.6 ± 0.21	46.8 ± 2.95	46.5 ± 1.16	$\mathbf{68.9 \pm 0.33}$
IMS → XJTU-SY	50.0 ± 0.17	50.3 ± 2.12	47.3 ± 0.64	50.5 ± 2.63	48.0 ± 1.95	$\mathbf{57.0 \pm 1.85}$
Average	44.3	49.2	52.5	48.7	47.3	**60.6**

5.6.5 Analysis and Discussion

Through many of the UDA tasks described in the two prior sections, it was confirmed that – when domains are different – diagnosis performance for mechanical systems can be improved by using UDA methods. Also, it was shown that, in all cases, the proposed DASC method outperforms other UDA methods thanks to adopting the additional loss term called semantic clustering loss and reducing it at multiple feature levels. This is the result of semantically well-clustered, discriminative features whose inter-class distance is large enough, and simultaneously, whose intra-class distance is minor, learned by DASC. To improve understanding and demonstrate the efficacy of DASC, additional analyses are explored and described below.

First, the efficacy of DASC can be confirmed through the visualization of the learned feature distributions. In this paper, learned feature distributions are visualized using the t-distributed Stochastic Neighboring Embedding (t-SNE) method [63]. Figure 5-11 shows the visualization results for task $\mathrm{CWRU1} \rightarrow \mathrm{CWRU3}$ of scenario I. This figure shows that the feature distribution obtained by DASC best separates data by class. This is because the DASC method learned the most discriminative features by adopting semantic clustering loss. As a result, it can be seen that the overlap between different classes is the least and that the same classes are well-clustered.

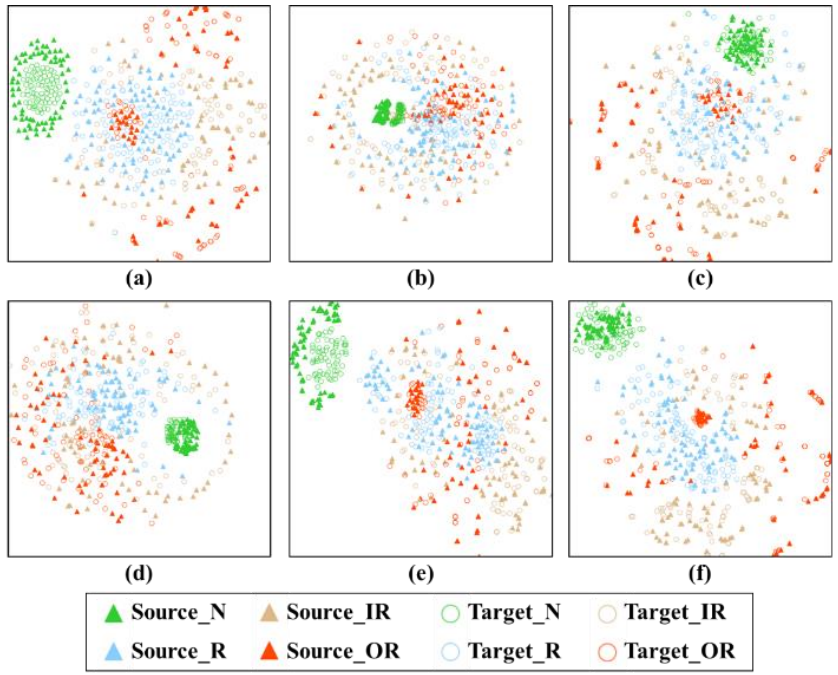


Figure 5-11 Feature visualization results using t-SNE for scenario I with the CWRU dataset: (a) CNN; (b) DANN; (c) MMD; (d) Adv+Disc; (e) Deep CORAL; and (f) DASC.

The visualization results for the $\mathrm{CWRU1} \rightarrow \mathrm{IMS}$ task of scenario II are shown in Figure 5-12. As shown in Figure 5-12 (a), for a CNN that only used the source data, since it does not consider the relationship with the target data, the domain discrepancy is large. On the other hand, by using UDA methods, the discrepancy between D_S and D_T is reduced to learn domain-invariant features that can be shared in both domains, as displayed in Figure 5-12 (b)-(f). For DASC, both discriminative and domain-invariant representations can be obtained through learning semantically well-clustered features according to their health conditions. Thus, the learned features have the characteristics of significant inter-class variations and minor intra-class variations. Consequently, based on those well-learned robust features, the best generalization performances in the target domain can be obtained.

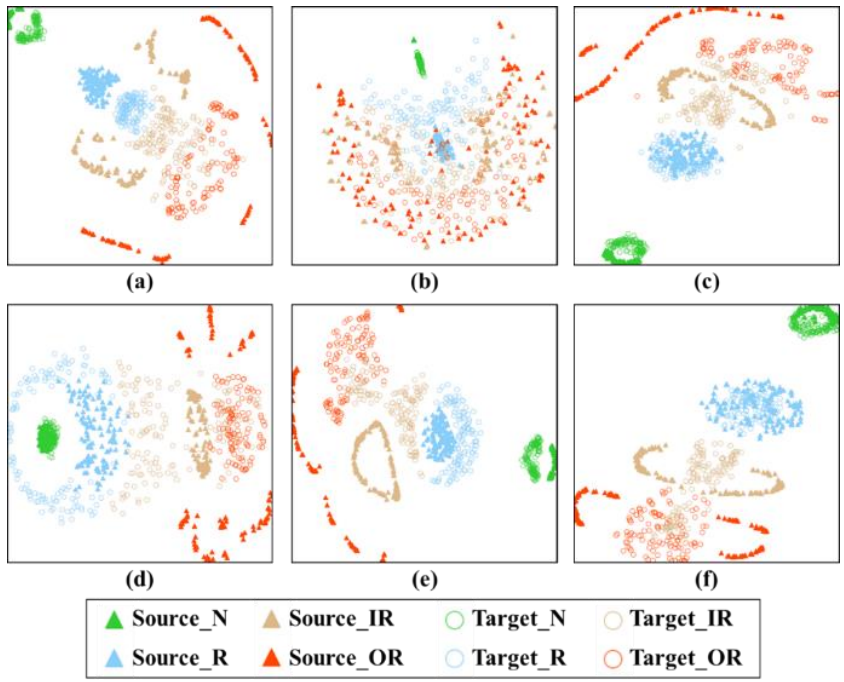


Figure 5-12 Feature visualization results using t-SNE for scenario II, CWRU→IMS task: (a) CNN; (b) DANN; (c) MMD; (d) Adv+Disc; (e) Deep CORAL; and (f) DASC.

Next, the effectiveness of our method was verified through an ablation study. In this section, for the ablation study, four different methods were compared: (1) CNN using classification loss only; (2) UDA using domain-related loss; (3) Last_SC using semantic clustering loss only at the last layer of the feature extractor; and (4) our proposed DASC method. The diagnosis accuracy results for all tasks of scenario I and scenario II are shown in Figure 5-13. As shown in this figure, by employing domain-related loss, the diagnosis performance in the target domain can be improved, as compared to the CNN models that are trained with only source domain data. In addition, by considering the semantic clustering loss term, which makes the samples from each class semantically well-clustered, the diagnosis performance can be further improved. Lastly, the highest performances were consistently obtained by applying the semantic clustering loss at multiple feature levels.

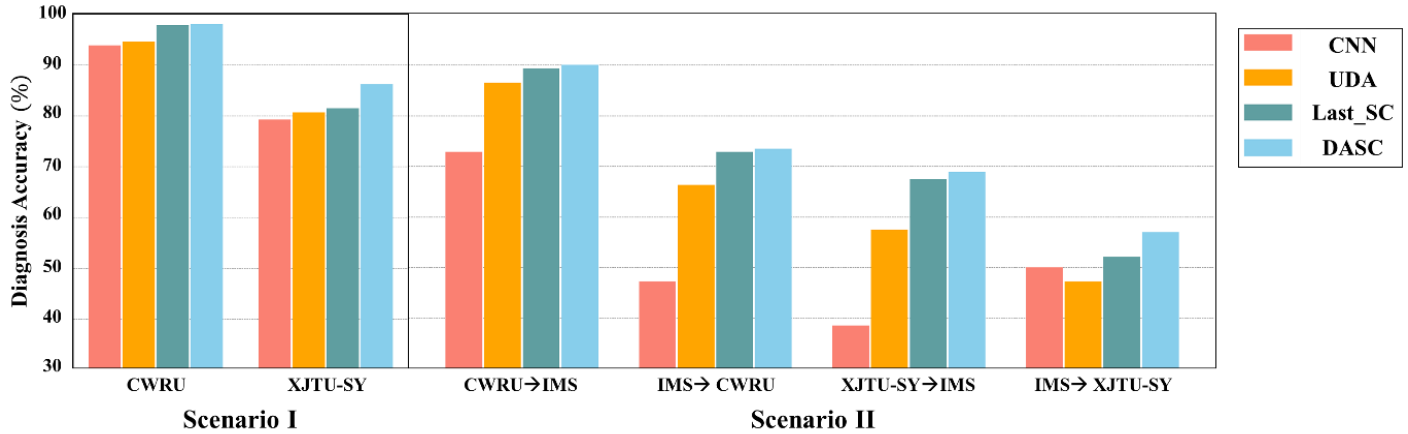


Figure 5-13 Average diagnosis accuracy (%) of each method, as found by the ablation study.

Furthermore, to evaluate how well the target features are clustered semantically, which is the property that we purpose to achieve by using the proposed method, an index called the semantic clustering index (SCI) is defined. The SCI is defined as the ratio between averaged inter-class distances and averaged intra-class distances for target data, which is expressed as:

$$SCI = \frac{2}{(C-1)} \cdot \frac{\sum_{l_1,l_2=1}^{C} \left\| \mathrm{cent}_{l_1} - \mathrm{cent}_{l_2} \right\|_2}{\sum_{l=1}^{C} \frac{1}{n_l} \sum_{i=1}^{n_l} \left\| f_{\theta_f}(x_i) - \mathrm{cent}_l \right\|_2} \tag{5.10}$$

where C represents the number of health conditions studied for the target rotating machinery; n_l represents the number of samples with health condition l; $f_{\theta_f}(x_i)$ denotes the feature vector of sample x_i extracted by a feature extractor; and cent_l signifies the center of samples with label l in the extracted feature space. A larger SCI value means that the target samples are semantically better clustered according to their classes. In other words, it means that the target sample features have smaller within-class variations and bigger between-class distances. Therefore, we expect that a UDA method that has a large SCI value will have high target generalization performances, especially for data located near the boundary of each class.

The SCI values of all compared UDA methods for UDA tasks of the scenario I and II are presented in Table 5-7, and the averaged SCI values are displayed in Figure 5-14. The figure shows that DASC consistently has the largest SCI values among the compared UDA methods, for every task in both scenarios. This means that the features learned by this method have the most discriminative distributions. This is because the DASC method learns semantically well-clustered features by applying an additional semantic clustering loss term at multiple feature levels; these results

explain well why our method shows the highest generalization performances in the target domain. From these results, which analyze SCI values, we can confirm that semantic clustering loss, based on source domain labeled data only, can enhance the class-wise clustering performance of different, but related, target domain data. Thus, the proposed approach leads to improved target domain generalization performance.

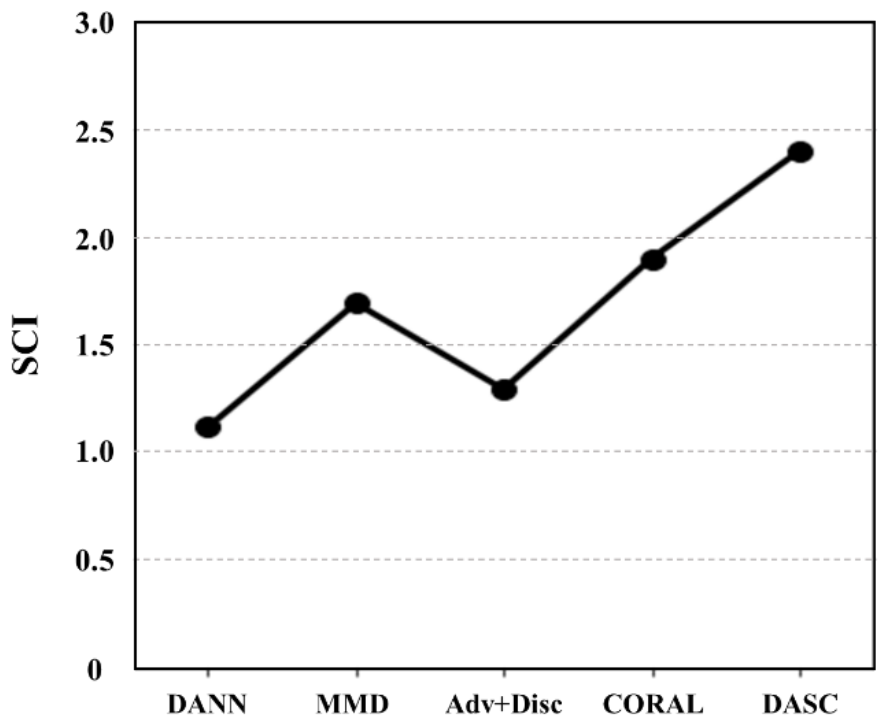


Figure 5-14 Averaged semantic clustering index (SCI) values.

Table 5-7 Semantic clustering index (SCI) for all UDA tasks of scenario I and II.

SCI	DANN	MMD	Adv+Disc	Deep CORAL	Proposed
CWRU	1.069	1.455	0.923	1.683	**1.702**
XJTU-SY	0.944	1.379	1.137	1.620	**1.706**
CWRU → IMS	1.173	2.160	2.219	2.758	**3.277**
IMS → CWRU	0.519	1.899	1.734	1.794	**1.970**
XJTU-SY → IMS	1.281	1.897	0.992	1.788	**2.233**
IMS → XJTU-SY	1.721	1.373	0.760	1.766	**3.520**
Average	1.118	1.694	1.294	1.902	**2.401**

Finally, as shown in Table 5-3 to Table 5-6, DASC shows the smallest performance variances in most tasks, which means that it can provide the most stable and robust diagnosis performances for target domain data. Through the results of these various analyses provided in this section, we can confirm that DASC, proposed herein, offers the most accurate and robust UDA-based diagnosis models for mechanical systems, in settings where it is problematic to acquire label information.

5.7 Summary and Discussion

In the research outlined in this chapter, an advanced UDA approach, domain adaptation with semantic clustering (DASC), for diagnosis of mechanical systems is offered. Our approach pursues to learn discriminative and domain-invariant representations to improve generalization performance in the D_T using information from D_S. The effectiveness of DASC is verified based on two UDA-based fault diagnosis scenarios using experimental bearing data. Results show that, in most tasks, DASC offers better target diagnosis accuracy performances than any other approaches, by applying semantic clustering loss at multiple feature levels. In addition, through visualization of learned features and an ablation study, this research demonstrates the superiorities of the proposed DASC method. Furthermore, by defining a new index called SCI, it is possible to confirm the superior characteristics of the learned features by the DASC approach.

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

1) **Myungyon Kim**, Jin Uk Ko, Jinwook Lee, Byeng D. Youn, Joon Ha Jung, and Kyung Ho Sun, "A Domain Adaptation with Semantic Clustering (DASC) Method for Fault Diagnosis of Rotating Machinery," *ISA Transactions*, 2021.

Chapter 6

Conclusion

6.1 Contributions and Significance

The proposed research in this doctoral dissertation aims at developing the methods to maximize the use of information for deep learning based fault diagnosis techniques. This doctoral dissertation is composed of three research: (1) a direct connection based convolutional neural network which enhances the gradient information flow; (2) a robust and discriminative feature learning method for fault diagnosis by transferring the pre-trained model and making the features better separated by their classes; and (3) a domain adaptation method based on the semantic clustering loss for learning more discriminative domain-invariant features. It is expected that the proposed research offers the following potential contributions and broader impacts in the fields related to fault diagnosis techniques.

Contribution 1: Suggestion of an advanced CNN-based architecture to improve the gradient information flow within the deep learning model

This doctoral dissertation proposes a direct connection based CNN (DC-CNN) architecture to greatly increase the training efficiency and diagnosis performance. Based on the directly connected convolutional module, the gradient information flow can be maximized. As a result, the diagnosis models based on the deeper network architecture can be trained well. In addition, the training efficiency can be also enhanced by adopting the dimension reduction module. Consequently, this DC-CNN architecture enables learning of enriched and superior features based on the enhanced information flow, which results in higher and more stable diagnosis performances.

Contribution 2: Suggestion of a method to learn well-generalized diagnosis models even under insufficient and noisy data conditions

This doctoral dissertation proposes a robust and discriminative feature learning method to improve the diagnosis performances even under insufficient and noisy data conditions. Effective and robust feature learning is possible through the pre-trained model learned from the source domain with abundant data. In addition, by additionally utilizing the triplet loss during the model training process, a semantically well-separable and discriminative feature can be learned. Through the use of the pre-trained model, information obtained from abundant data in the source domain is transferred and maximally utilized. Also, the distribution of samples by their classes was considered. As a result, by taking advantage of transfer learning and metric learning concepts, an effective diagnosis model can be learned.

Contribution 3: Suggestion of a diagnosis method which provides high generalized performance for unlabeled target domain data

This doctoral dissertation proposes a novel domain adaptation based fault diagnosis technique to improve the diagnosis performance for the unlabeled target domain by maximally using the label information obtain from the source domain. By minimizing the domain discrepancy metric, domain-invariant features are obtained. In addition, by applying newly devised semantic clustering loss at multiple feature levels, semantically well-clustered features can be secured. As a result, more discriminative domain-invariant features can be learned based on the proposed method. Furthermore, the new metric for evaluating how well the features clustered semantically was suggested. In conclusion, based on the proposed DASC method, we can enhance the fault diagnosis performance for the target domain by transferring the label information obtained from the source domain.

6.2 Suggestions for Future Research

Although the technical advances proposed in this doctoral dissertation successfully address some issues in the field of deep learning based fault diagnosis techniques, there are still several research topics that further investigations and developments are required to enhance the fault diagnosis performance more. Specific suggestions for future research are listed as follows.

Suggestion 1: Physics-informed artificial intelligence (Physics-informed AI) for fault diagnosis techniques

As available data increases, research on data-driven fault diagnosis methods is increasing. However, if we make good use of physical and domain knowledge, we can obtain information that may be difficult to obtain by data-driven methods or support a data-based method. Therefore, research to develop physics-informed AI-based fault diagnosis techniques that combine usable physical knowledge and data-driven methods should be conducted.

Suggestion 2: Fault diagnosis techniques considering class imbalance problems

In real industrial fields, in many cases, the amount of data that can be obtained is different for each health condition, and this situation is called the class imbalance problem. In this case, it may be difficult to learn a high-performance diagnosis model compared to having sufficient data for all states. Therefore, research on developing the methods which can effectively learn the high-performance diagnosis models even under these class imbalance circumstances should be conducted.

Suggestion 3: Domain adaptation method that maximally uses the information from the target domain

In the UDA problem, since target data is unlabeled, the information that can be

obtained from the target domain is limited. However, the ultimate goal of UDA is to learn a diagnosis model with high generalization performances in the target domain. Therefore, it is necessary to conduct research to develop a method that effectively and maximally uses the information extracted from the unlabeled target domain for improving the target diagnosis performance.

Suggestion 4: Research on an integrated methodology that properly combines several techniques

In this doctoral dissertation, various methods were presented to maximize the use of information for deep learning based fault diagnosis techniques. Since the problem situations in which each method is used are different, there may be a limit to using these methods simultaneously. However, there may be some possibilities to obtain higher diagnosis performances by integrating the component techniques used in each method. Therefore, in order to maximize the diagnosis performance of the target system, it is necessary to conduct research on an integrated methodology that properly combines and uses several techniques according to the problem settings.

References

1. Z. Gao, C. Cecati, and S. X. Ding, “A survey of fault diagnosis and fault-tolerant techniques-part I: Fault diagnosis with model-based and signal-based approaches,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

2. Z. Gao, C. Cecati, and S. X. Ding, “A survey of fault diagnosis and fault-tolerant techniques-part II: Fault diagnosis with knowledge-based and hybrid/active approaches,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.

3. A. K. Jalan and A. R. Mohanty, “Model based fault diagnosis of a rotor–bearing system for misalignment and unbalance under steady-state condition,” *Journal of Sound and Vibration*, vol. 327, no. 3–5, pp. 604–622, Nov. 2009.

4. J. Poon, P. Jain, I. C. Konstantakopoulos, C. Spanos, S. K. Panda, and S. R. Sanders, “Model-based fault detection and identification for switching power converters,” *IEEE Transactions on Power Electronics*, vol. 32, no. 2, pp. 1419–1430, Feb. 2017.

5. J. Luo, M. Namburu, K. R. Pattipati, L. Qiao, and S. Chigusa, “Integrated model-based and data-driven diagnosis of automotive antilock braking systems,” *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 40, no. 2, pp. 321–336, Mar. 2010.

6. C. Hu, B. D. Youn, and P. Wang, “Engineering Design under Uncertainty and Health Prognostics,” 2019.

7. A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mechanical Systems and Signal Processing*, vol. 20, no. 7. pp. 1483–1510, Oct-2006.

8. J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, “Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications,” *Mechanical Systems and Signal Processing*, vol. 42, no. 1–2, pp. 314–334, Jan. 2014.

9. D. E. Bently, C. T. Hatch, and B. Grissom, *Fundamentals of Rotating Machinery Diagnostics*. ASME Press, 2002.

10. J. M. Ha, B. D. Youn, H. Oh, B. Han, Y. Jung, and J. Park, "Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 161–175, Mar. 2016.

11. J. Park, B. Jeon, J. Park, J. Cui, M. Kim, and B. D. Youn, "Failure prediction of a motor-driven gearbox in a pulverizer under external noise and disturbance," *Smart Structures and Systems*, vol. 22, no. 2, pp. 185–192, 2018.

12. Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, May 2016.

13. S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowledge-Based Systems*, vol. 140, pp. 1–14, Jan. 2018.

14. J. U. Ko, J. H. Jung, M. Kim, H. B. Kong, J. Lee, and B. D. Youn, "Multi-task learning of classification and denoising (MLCD) for noise-robust rotor system diagnosis," *Computers in Industry*, vol. 125, p. 103385, Feb. 2021.

15. Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, 27-May-2015.

16. J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, May 2018.

17. Z. Li, M. Dong, S. Wen, X. Hu, P. Zhou, and Z. Zeng, "CLU-CNNs: Object detection for medical images," *Neurocomputing*, vol. 350, pp. 53–59, Jul. 2019.

18. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

19. R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training Very Deep Networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 2377–2385.

20. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings - 30th IEEE Conference*

on Computer Vision and Pattern Recognition, CVPR 2017, 2017, vol. 2017-Janua, pp. 2261–2269.

21. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 770–778.

22. S. Soetens, A. Sarris, and K. Vansteenhuyse, "A Discriminative Feature Learning Approach for Deep Face Recognition," in *European conference on computer vision*, 2016, pp. 499–515.

23. N. Jin, J. Wu, X. Ma, K. Yan, and Y. Mo, "Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification," *IEEE Access*, vol. 8, pp. 77060–77072, 2020.

24. Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin, "Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features," *IEEE Access*, vol. 7, pp. 105146–105158, Jan. 2019.

25. L. Wen, X. Li, L. Gao, and Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

26. Q. Liu and C. Huang, "A Fault Diagnosis Method Based on Transfer Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 171423–171430, 2019.

27. L. S. Maraaba, A. S. Milhem, I. A. Nemer, H. Al-Duwaish, and M. A. Abido, "Convolutional Neural Network-Based Inter-Turn Fault Diagnosis in LSPMSMs," *IEEE Access*, vol. 8, pp. 81960–81970, 2020.

28. Y. Kim, T. Kim, B. D. Youn, and S.-H. Ahn, "Machining quality monitoring (MQM) in laser-assisted micro-milling of glass using cutting force signals: an image-based deep transfer learning," *Journal of Intelligent Manufacturing 2021*, pp. 1–16, Apr. 2021.

29. M. Xia, T. Li, L. Xu, L. Liu, and C. W. De Silva, "Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.

30. M. Kim, J. H. Jung, J. U. Ko, H. B. Kong, J. Lee, and B. D. Youn, "Direct Connection-Based Convolutional Neural Network (DC-CNN) for Fault Diagnosis of Rotor Systems," *IEEE Access*, 2020.

31. C. Cheng, B. Zhou, G. Ma, D. Wu, and Y. Yuan, “Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data,” *Neurocomputing*, vol. 409, pp. 35–45, Oct. 2020.

32. S. Haidong, D. Ziyang, C. Junsheng, and J. Hongkai, “Intelligent fault diagnosis among different rotating machines using novel stacked transfer auto-encoder optimized by PSO,” *ISA Transactions*, vol. 105, pp. 308–319, Oct. 2020.

33. S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, “Analysis of Representations for Domain Adaptation,” in *NIPS’06: Proceedings of the 19th International Conference on Neural Information Processing Systems*, 2006, pp. 137–144.

34. S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J. Wortman Vaughan, S. R. Ben-David David Cheriton, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. Vaughan, “A theory of learning from different domains,” *Mach Learn*, vol. 79, pp. 151–175, 2010.

35. S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10. pp. 1345–1359, 2010.

36. M. Kaya and H. Ş. Bilge, “Deep metric learning: A survey,” *Symmetry*, vol. 11, no. 9, p. 1066, Aug. 2019.

37. J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu, “Learning fine-grained image similarity with deep ranking,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1386–1393.

38. S. Chopra, R. Hadsell, and Y. LeCun, “Learning a similarity metric discriminatively, with application to face verification,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. I, pp. 539–546.

39. Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” in *Advances in Neural Information Processing Systems*, 2014, vol. 3, no. January, pp. 1988–1996.

40. E. Hoffer and N. Ailon, “Deep metric learning using triplet network,” in *International Workshop on Similarity-Based Pattern Recognition*, 2015, vol. 9370, pp. 84–92.

41. J. Ni, J. Liu, C. Zhang, D. Ye, and Z. Ma, “Fine-grained patient similarity measuring using deep metric learning,” in *International Conference on*

Information and Knowledge Management, Proceedings, 2017, pp. 1189–1198.

42. C. Chen, Z. Chen, B. Jiang, and X. Jin, "Joint Domain Alignment and Discriminative Feature Learning for Unsupervised Deep Domain Adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, no. 01, pp. 3296–3303.

43. X. Li, L. Yu, C.-W. Fu, M. Fang, and P.-A. Heng, "Revisiting Metric Learning for Few-Shot Image Classification," in *CoRR abs/1907.03123*, 2019.

44. J. Wang, K. C. Wang, M. T. Law, F. Rudzicz, and M. Brudno, "Centroid-based Deep Metric Learning for Speaker Recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, vol. 2019-May, pp. 3652–3656.

45. X. Wang and F. Liu, "Triplet Loss Guided Adversarial Domain Adaptation for Bearing Fault Diagnosis," *Sensors*, vol. 20, no. 1, p. 320, Jan. 2020.

46. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*, 2014, no. PART 1, pp. 818–833.

47. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

48. T. Tan, Y. Qian, H. Hu, Y. Zhou, W. Ding, and K. Yu, "Adaptive Very Deep Convolutional Residual Network for Noise Robust Speech Recognition," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 26, no. 8, pp. 1393–1405, Aug. 2018.

49. Y. Bengio, P. Simard, and P. Frasconi, "Learning Long-Term Dependencies with Gradient Descent is Difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

50. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of Machine Learning Research*, vol. 9, pp. 249–256, 2010.

51. Q. Xu, M. Zhang, Z. Gu, and G. Pan, "Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs," *Neurocomputing*, vol. 328, pp. 69–74, Feb. 2019.

52. S. Gupta, R. Gupta, M. Ojha, and K. P. Singh, “A comparative analysis of various regularization techniques to solve overfitting problem in artificial neural network,” in *Communications in Computer and Information Science*, 2018, vol. 799, pp. 363–371.

53. J. H. Jung, B. C. Jeon, B. D. Youn, M. Kim, D. Kim, and Y. Kim, “Omnidirectional regeneration (ODR) of proximity sensor signals for robust diagnosis of journal bearing systems,” *Mechanical Systems and Signal Processing*, vol. 90, pp. 189–207, Jun. 2017.

54. H. Oh, J. H. Jung, B. C. Jeon, and B. D. Youn, “Scalable and Unsupervised Feature Engineering Using Vibration-Imaging and Deep Learning for Rotor System Diagnosis,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3539–3549, Apr. 2018.

55. S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, 2015, pp. 448–456.

56. Z. kai Feng, W. jing Niu, Z. yang Tang, Z. qiang Jiang, Y. Xu, Y. Liu, and H. rong Zhang, “Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization,” *Journal of Hydrology*, vol. 583, p. 124627, Apr. 2020.

57. Z. Chai and C. Zhao, “Enhanced random forest with concurrent analysis of static and dynamic nodes for industrial fault classification,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 54–66, Jan. 2020.

58. S. Yan and X. Yan, “Using Labeled Autoencoder to Supervise Neural Network Combined with k-Nearest Neighbor for Visual Industrial Process Monitoring,” *Industrial and Engineering Chemistry Research*, vol. 58, no. 23, pp. 9952–9958, Jun. 2019.

59. W. He, Y. He, B. Li, and C. Zhang, “A Naive-Bayes-Based Fault Diagnosis Approach for Analog Circuit by Using Image-Oriented Feature Extraction and Selection Technique,” *IEEE Access*, vol. 8, pp. 5065–5079, 2020.

60. Y. Li, Y. Wei, K. Feng, X. Wang, and Z. Liu, “Fault diagnosis of rolling bearing under speed fluctuation condition based on vold-kalman filter and RCMFE,” *IEEE Access*, vol. 6, pp. 37349–37360, Jul. 2018.

61. B. C. Jeon, J. H. Jung, B. D. Youn, Y. W. Kim, and Y. C. Bae, “Datum unit optimization for robustness of a journal bearing diagnosis system,” *International Journal of Precision Engineering and Manufacturing*, vol. 16,

no. 11, pp. 2411–2425, Oct. 2015.

62. Z. kai Feng, W. jing Niu, R. Zhang, S. Wang, and C. tian Cheng, "Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization," *Journal of Hydrology*, vol. 576, pp. 229–238, Sep. 2019.

63. L. Van Der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

64. Q. Zhang, Y. N. Wu, and S. C. Zhu, "Interpretable Convolutional Neural Networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8827–8836.

65. B. C. Jeon, J. H. Jung, M. Kim, K. H. Sun, and B. D. Youn, "Optimal vibration image size determination for convolutional neural network based fluid-film rotor-bearing system diagnosis," *Journal of Mechanical Science and Technology*, vol. 34, no. 4, pp. 1467–1474, 2020.

66. W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA Transactions*, vol. 95, pp. 295–305, Dec. 2018.

67. M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, and B. Liang, "A Deep Transfer Model with Wasserstein Distance Guided Multi-Adversarial Networks for Bearing Fault Diagnosis under Different Working Conditions," *IEEE Access*, vol. 7, pp. 65303–65318, 2019.

68. B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 692–706, May 2019.

69. L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.

70. T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application," *ISA Transactions*, vol. 97, pp. 269–281, Feb. 2020.

71. P. Ma, H. Zhang, W. Fan, and C. Wang, "A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains," *ISA Transactions*, 2019.

72. T. Han, C. Liu, W. Yang, and D. Jiang, “A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults,” *Knowledge-Based Systems*, vol. 165, pp. 474–487, Feb. 2019.

73. E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep Domain Confusion: Maximizing for Domain Invariance,” in *CoRR abs/1412.3474*, 2014.

74. B. Sun and K. Saenko, “Deep CORAL: Correlation Alignment for Deep Domain Adaptation,” in *In ICCV workshop on Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV)*, 2016.

75. Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, U. Dogan, M. Kloft, F. Orabona, T. Tommasi, and A. Ganin, “Domain-Adversarial Training of Neural Networks,” *Journal of Machine Learning Research*, vol. 17, pp. 1–35, 2016.

76. A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A Kernel Two-Sample Test,” *Journal of Machine Learning Research*, vol. 13, pp. 723–773, 2012.

77. I. Steinwart, “On the influence of the kernel on the consistency of support vector machines,” *Journal of machine learning research*, vol. 2, no. 1, pp. 67–93, 2001.

78. M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” in *32nd International Conference on Machine Learning, ICML 2015*, 2015, vol. 1, pp. 97–105.

79. M. Kim, J. U. Ko, J. Lee, B. D. Youn, J. H. Jung, and K. H. Sun, “A Domain Adaptation with Semantic Clustering (DASC) method for fault diagnosis of rotating machinery,” *ISA Transactions*, Mar. 2021.

80. K. A. Loparo, “Case Western Reserve University Bearing Data Center,” 2003. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/home.

81. H. Qiu, J. Lee, J. Lin, and G. Yu, “Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics,” *Journal of Sound and Vibration*, vol. 289, no. 4–5, pp. 1066–1090, Feb. 2006.

82. B. Wang, Y. Lei, N. Li, and N. Li, “A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings,” *IEEE Transactions on Reliability*, 2020.

국문 초록

딥러닝 기반 고장 진단을 위한 정보 활용 극대화 기법 개발

서울대학교 대학원
기계항공공학부
김 명 연

기계 시스템의 예기치 않은 고장은 많은 산업 분야에서 막대한 사회적, 경제적 손실을 야기할 수 있다. 갑작스런 고장을 감지하고 예방하여 기계 시스템의 신뢰성을 높이기 위해 데이터 기반 고장 진단 기술을 개발하기 위한 연구가 활발하게 이루어지고 있다. 고장 진단 기술의 목표는 대상 기계 시스템의 고장 발생을 가능한 빨리 감지하고 진단하는 것이다. 최근 합성곱 신경망 기법을 포함한 딥러닝 기반 고장 진단 기술은 자율적인 특성인자(feature) 학습이 가능하고 높은 진단 성능을 얻을 수 있다는 장점이 있어 활발히 연구되고 있다.

그러나 딥러닝 기반의 고장 진단 기술을 개발함에 있어 해결해야 할 몇 가지 문제점들이 존재한다. 먼저, 신경망 구조를 깊게 쌓음으로써 풍부한 계층적 특성인자들을 배울 수 있고, 이를 통해 향상된 성능을 얻을 수 있다. 그러나 기울기(gradient) 정보 흐름의 비효율성과 과적합

문제로 인해 모델이 깊어질수록 학습이 어렵게 된다는 문제가 있다. 다음으로, 높은 성능의 고장 진단 모델을 학습하기 위해서는 충분한 양의 레이블 데이터(labeled data)가 확보돼야 한다. 그러나 실제 현장에서 운용되고 있는 기계 시스템의 경우, 충분한 양의 데이터와 레이블 정보를 얻는 것이 어려운 경우가 많다. 따라서 이러한 문제들을 해결하고 진단 성능을 향상시키기 위한 새로운 딥러닝 기반 고장 진단 기술의 개발이 필요하다.

본 박사학위논문에서는 딥러닝 기반 고장 진단 기술의 성능을 향상시키기 위한 세가지 정보 활용 극대화 기법에 대한 연구로 1) 딥러닝 아키텍처 내 기울기 정보 흐름을 향상시키기 위한 새로운 딥러닝 구조 연구, 2) 파라미터 전이 및 삼중항 손실을 기반으로 불충분한 데이터 및 노이즈 조건 하 강건하고 차별적인 특성인자 학습에 대한 연구, 3) 다른 도메인으로부터 레이블 정보를 전이시켜 사용하는 도메인 적응 기반 고장 진단 기법 연구를 제안한다.

첫 번째 연구에서는 딥러닝 모델 내 기울기 정보 흐름을 개선하기 위한 향상된 합성곱 신경망 기반 구조를 제안한다. 본 연구에서는 다양한 계층의 아웃풋(feature map)을 직접 연결함으로써 향상된 정보 흐름을 얻을 수 있으며, 그 결과 진단 모델을 효율적으로 학습하는 것이 가능하다. 또한 차원 축소 모듈을 통해 학습 파라미터 수를 크게 줄임으로써 학습 효율성을 높일 수 있다.

두 번째 연구에서는 파라미터 전이 및 메트릭 학습 기반 고장 진단 기법을 제안한다. 본 연구는 데이터가 불충분하고 노이즈가 많은 조건 하에서도 높은 고장 진단 성능을 얻기 위해 강건하고 차별적인 특성인자 학습을 가능하게 한다. 먼저, 풍부한 소스 도메인 데이터를 사용해 훈련된 사전학습모델을 타겟 도메인으로 전이해 사용함으로써 강건한

진단 방법을 개발할 수 있다. 또한, semi-hard 삼중항 손실 함수를 사용함으로써 각 상태 레이블에 따라 데이터가 더 잘 분리되도록 해주는 특성인자를 학습할 수 있다.

세 번째 연구에서는 레이블이 지정되지 않은(unlabeled) 대상 도메인에서의 고장 진단 성능을 높이기 위한 레이블 정보 전이 전략을 제안한다. 우리가 목표로 하는 대상 도메인에서의 고장 진단 방법을 개발하기 위해 다른 소스 도메인에서 얻은 레이블 정보가 전이되어 활용된다. 동시에 새롭게 고안한 의미론적 클러스터링 손실(semantic clustering loss)을 여러 특성인자 수준에 적용함으로써 차별적인 도메인 불변 기능을 학습한다. 결과적으로 도메인 불변 특성을 가지며 의미론적으로 잘 분류되는 특성인자를 효과적으로 학습할 수 있음을 증명하였다.

주요어: 고장 진단
딥러닝
합성곱 신경망
전이학습
비지도 도메인 적응
정보 활용 극대화

학 번: 2015-22708