Artificial Intelligence Supported EV Electric Powertrain for Safety Improvement

Xiaotian Zhang

PhD

University of York

Physics, Engineering and Technology

September 2023

Abstract

As an environmentally friendly transport option, electric vehicles (EVs) are endowed with the characteristics of low fossil energy consumption and low pollutant emissions. In today's growing market share of EVs, the safety and reliability of the powertrain system will be directly related to the safety of human life. Reliability problems of EV powertrains may occur in any power electronic (PE) component and mechanical part, both sudden and cumulative. These faults in different locations and degrees will continuously threaten the life of drivers and pedestrians, bringing irreparable consequences. Therefore, monitoring and predicting the real-time health status of EV powertrain is a high-priority, arduous and challenging task.

The purposes of this study are to develop AI-supported effective safety improvement techniques for EV powertrains. In the first place, a literature review is carried out to illustrate the up-to-date AI applications for solving condition monitoring and fault detection issues of EV powertrains, where recent case studies between conventional methods and AI-based methods in EV applications are compared and analysed. On this ground this study, then, focuses on the theories and techniques concerning this topic so as to tackle different challenges encountered in the actual applications. In detail, first, as for diagnosing the bearing system in the earlier fault period, a novel inferable deep distilled attention network is designed to detect multiple bearing faults. Second, a deep learning and simulation driven approach that combines the domain-adversarial neural network and the lumped-parameter thermal network (LPTN) is proposed for achieve IPMSM permanent magnet temperature estimation work. Finally, to ensure the use safety of the IGBT module, deep learning -based IGBT modules' double pulse test (DPT) efficiency enhancement is proposed and achieved via multimodal fusion networks and graph convolution networks.

Key words: Electric vehicle; neural network; deep learning; powertrain; IGBT; bearing; PMSM; reliability

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my main supervisor, Dr. Yihua Hu, now working at Kings College London, for granting me the precious opportunity to further myself at the highest level. He has encouraged and helped me to progress academically from a fresh graduate to a professional researcher and a mature individual. I have learned so much from him both academically and personally. Thank him for the invaluable comments and advice on my research as well as my life and career. Secondly, I also thank very much my another supervisor Dr. Mohammad Nasr Esfahani, for his selfless helps to my papers and thesis. It is a great honour for me to be their PhD student, and I hope I have fulfilled his high expectations and faith through my achievements. Thirdly, I would also like to thank my second supervisor Dr. Yue Wang, for the generous support and academic guidance he has provided me, and also thank my annual progression member Dr. Adar Pelah.

I would also like to take this opportunity and thank my parents. They have always been greatly supportive of me. Their continuous help and understanding have made my life full of love, and I am grateful for everything they have done. Additionally, I would like to express my appreciation to my senior Dr. Chao Gong. I am very thankful to his help. Because of his help, I am no longer confused about many things after my Ph.D. enrollment. Until now, he still guides me a lot. I am also very grateful to my co-operators Dr. Jingwei Zhang (China University of Mining and Technology), who have given me many useful suggestions for my research and helped me complete my experimental verifications. Special thanks are also paid to my brilliant and lovely colleagues and friends, in particular to Mr. Wnagjie Lang, Mr. Zeliang Zhang, Mr. Hongzuo Liu, Mr. Yu Nie, Mr. Shangming Mei, Mr. Mengyu Cheng, Mr. Wangde Qiu, Mr Yuxuan Du and Dr. Xing Zhao for many fruitful discussions and enjoyable moments.

Last but not least, the support from the School of Physics, Engineering and Technology at the University of York is gratefully acknowledged. Especially, I am grateful to Ms. Akiko and Ms. Camilla Danese who is really elegant, patient and kind.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Main contents in Chapter 2 have been published to *IEEE Sensors Journal*. Title is "Artificial Intelligence Technique based EV Powertrain Condition Monitoring and Fault Diagnosis: A Review".

Main contents in Chapter 3 have been published to *IEEE Transactions on Transportation Electrification*. Title is "Inferable Deep Distilled Attention Network for Diagnosing Multiple Motor Bearing Fault".

Main contents in Chapter 4 have been submitted and major revised to *IEEE Transactions on Industrial Informatics*. Title is "Domain-Adversarial Adaptation Regression Model for IPMSM Permanent Magnet Temperature Estimation".

Main contents in Chapter 5 have been submitted and major revised to *IEEE Transactions on Industrial Electronics*. Title is "IGBT Module DPT Efficiency Enhancement via Multimodal Fusion Networks and Graph Convolution Networks".

Xiaotian Zhang

Sep. 2023

List of Publications

[1] **X. Zhang**, Y. Hu, A. Yin, J. Deng, H. Xu and J. Si, "Inferable Deep Distilled Attention Network for Diagnosing Multiple Motor Bearing Faults," in *IEEE Transactions on Transportation Electrification*, vol. 9, no. 2, pp. 2207-2216, June 2023, doi: 10.1109/TTE.2022.3211203.

[2] **X. Zhang**, Y. Hu, J. Deng, H. Xu and H. Wen, "Feature Engineering and Artificial Intelligence-Supported Approaches Used for Electric Powertrain Fault Diagnosis: A Review," in *IEEE Access*, vol. 10, pp. 29069-29088, 2022, doi: 10.1109/ACCESS.2022.3157820.

[3] **X. Zhang**, Y. Hu, C. Gong, J. Deng and G. Wang, "Artificial Intelligence Technique-Based EV Powertrain Condition Monitoring and Fault Diagnosis: A Review," in *IEEE Sensors Journal*, vol. 23, no. 15, pp. 16481-16500, 1 Aug.1, 2023, doi: 10.1109/JSEN.2023.3285531.

[4] **X. Zhang**, Y. Hu, J. Zhang, M. Nasresfahani, T. Tilford, S. Stoyanov, "IGBT Module DPT Efficiency Enhancement Via Multimodal Fusion Networks and Graph Convolution Networks," *IEEE Transactions on Industrial Electronics.* (Major Revision)

[5] **X. Zhang**, Y. Hu, H. Xu, M. Alkahtani "Domain-Adversarial Adaptation Regression Model for IPMSM Permanent Magnet Temperature Estimation," *IEEE Transactions on Industrial Informatics.* (Major Revision)

[6] **X. Zhang**, Y. Hu, H. Xu, M. Nasresfahani, " Transferable One-Dimensional Contrastive Self-Supervised Framework for Motor Bearing Multi-Fault Diagnosis," *IEEE Transactions on Transportation Electrification*. (Under Review)

[7] **X. Zhang**, C. Gong, Y. Hu, H. Xu and J. Deng, "Generative Adversarial Network Supported Permanet Magnet Temperature Estimation by Using Random Forest," 16th IEEE Intrenational Conference on compatibility, Power Electronics and Power Engineering, Oct. 2021.

[8] **X. Zhang**, Y. Hu, H. Xu and W. Lang, "Transferable Neural Network Method Used for Permanent Magnet Temperature Estimation of EV motors," *16th IEEE Intrenational Conference on compatibility，Power Electronics and Power Engineering*, UK, 2022.

[9] C. Gong, L. Ding, Y. Li, X. Wu and **X. Zhang**, "Flux Observation Transfer-Based Inductance Identification Technique for Precise FCS-MPCC Used in Surface-Mounted PMSMs," in *IEEE Transactions on Power Electronics*, vol. 38, no. 4, pp. 4241-4245, April 2023, doi: 10.1109/TPEL.2023.3235269.

Abbreviations

EV	Electric vehicle
PE	power electronic
PMSM	permanent magnet synchronous motor
RUL	remaining useful life
ML	machine learning
DL	deep learning
ANN	artificial neural network
DT	decision tree
SVM	support vector machine
Adaboost	adaptive boost
RF	random forest
AE	autoencoder
KNN	K-nearest neighbours
MLP	multilayer perceptron
CNN	convolutional neural network
DBN	deep belief network
RNN	recurrent neural network
GRU	gated recurrent unit
IM	induction motor
AI	artificial intelligence
KD	knowledge distillation
LPTN	lumped-parameter thermal network
CP	coarse predicting
IPMSM	interior PMSM
DPT	double pulse test

MAFN	multimodal attention fusion network
IDDAN	inferable deep distilled attention network
DAAR	domain-adversarial adaptation regression
PU	Paderborn University
GCN	graph convolution network
FT	Fourier transform
DFT	discrete Fourier transform
FFT	fast Fourier transform
IGBT	insulated gate bipolar transistor
QFT	quantum Fourier transform
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
WT	wavelet transform
WPT	wavelet packet transform
WVD	Wigner-Ville distribution
EMD	empirical mode decomposition
EWT	empirical wavelet transform
SK	spectral kurtosis
PCA	principal component analysis
SVM	support vector machine
GDA	generalized discriminant analysis
LMD	local mean decomposition
IMF	intrinsic module function
EEMD	ensemble EMD
WSST	wavelet semi-soft threshold
NPC	neutral point clamped three-level inverter
LCN	local connection network
NSAE	normalized sparse autoencoder

SAE	Sparse autoencoder
LCN	local connection network
VMD	variational mode decomposition
DE	diversity entropy
CRA	cumulative relative accuracy
MODBNE	multi-objective deep belief networks ensemble
EKF	extended Kalman filter
PF	particle filter
CWRU	case western reserve university
DAFD	domain adaptation neural-network-based fault diagnosis
GPR	Gaussian process regression
VSI	voltage source inverter
GAP	global average pooling
SEGLD	sparse elitist group lasso denoising
AMF	adaptive morphological filter
IF	information fusion
DCTLN	deep convolutional transfer learning network
VCTL	varying coefficient transfer learning
LN	varying coefficient transfer learning
MSA	multi-head self-attention
SA	self-attention
GELU	gaussian error linear unit
TNN	thermal neural network
DANN	Domain adversarial neural network
GAN	generative adversarial network
GRL	gradient reversal layer
ELU	exponential linear unit
MSE	mean square error

MAE	mean absolute error
FC	Fully connected
CE	cross-entropy
MMD	maximum mean discrepancy
RMSE	maximum mean discrepancy
RBF	radial basis function
SVR	support vector regression
RFR	random forest regression
PDF	probability density function
t-SNE	t-distributed stochastic neighbour embedding
2D	two-dimensional
FEA	finite element analysis
SPMSM	surface-mounted PMSM
SPICE	simulation program with integrated circuit emphasis
ETM	electrical transient model
FWD	freewheeling diode
DUT	device under test
RRC	reverse recovery characteristic
FFN	feed-forward network
SCC	Spearman correlation coefficient
LE	Logcosh error

Nomenclature

C	the number of channels
D	the constant latent vector size through all of transformer layers
D_s	Source domain
D_t	Target domain
$\mathcal{D}_S$	Labeled source domain dataset
$\mathcal{D}_T$	target domain dataset
$\mathbf{E}$	the patch embedding projection
f_s	the sampling rate
f_t	The target task learning function
G_D	domain classifier
G_S	source domain recursive predictor
G_T	target domain recursive predictor
$J(\cdot)$	the training error at any instance
k	the number of segments after each signal cutting
K	Key matrix
L_{hard}	The cross-entropy of the hard prediction value output by the student model and data label
L_s	the length of each vibration signal
L_{soft}	The cross-entropy between it and the soft prediction value output by the student model
m	the number of iterations
O_{ce}, O_{de}	the output of the two classifiers
P	power losses
P_v	Overall power losses
(P, P)	the resolution of each image patch
$P(\cdot)$	The probability distribution

Q	Query matrix
R	Resistance
R_s	the ohmic resistance
V	Value matrix
V_{new}	down sampled vibration signal
V_s	the original collected vibration signal
$\mathbf{w}_{class}$	the class token
$\mathbf{w}_{distill}$	the distillation token
W^O	The parameter matrix of linear transforms
X	Input matrix
Y_s	The category spaces of the learning objectives of source domain
Y_t	The category spaces of the learning objectives of target domain
$\mathbf{z}_L^0$	the Transformer encoder
α	Weight parameter
β	Weight parameter
δ	the impulse function
θ	Temperature
$\mathcal{L}_D$	Loss function of the domain classifier
$\mathcal{L}_S$	Loss function of the source domain
$\mathcal{L}_T$	Loss function of the target domain
$\phi(\cdot)$	The mapping function
$\sigma(\cdot)$	Softmax function
μ	hyperparameter
λ	hyperparameter
γ	hyperparameter

Chapter 1 **Introduction**

1.1 Motivation

Human life and industry are inseparable from the help of electric powertrain systems, which can bring great convenience to human beings. In today's growing market share of electric vehicles (EVs), the safety and reliability of the powertrain system will be directly related to the safety of human life. Reliability problems of EV powertrains may occur in any power electronic (PE) component and mechanical part, both sudden and cumulative. For instance, permanent magnets of the permanent magnet synchronous motor (PMSM) will be demagnetized if it operates in a high temperature environment for a long time; in the DC-bus, the failure of capacitors and switches will be caused by various reasons such as aging and overstress; the wear and damage of motor bearings can also cause serious safety hazards. These faults in different locations and degrees will continuously threaten the life of drivers and pedestrians, bringing irreparable consequences. Therefore, monitoring and predicting the real-time health status of EV powertrain (Fig. 1-1) is a high-priority, arduous and challenging task [1].

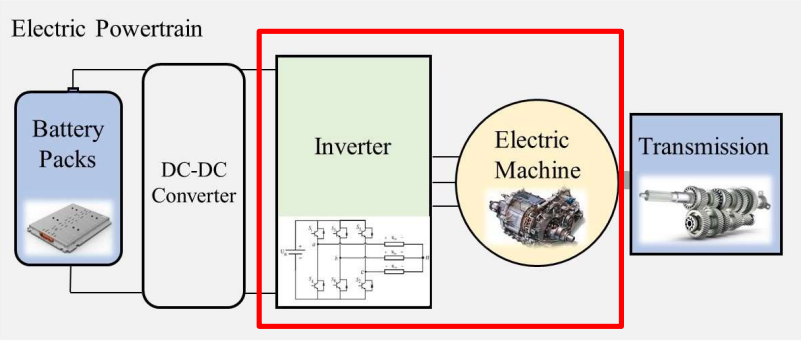


Fig. 1-1. Diagram of regular EV powertrain system (this thesis is main focous on the componments in red square).

Condition monitoring and fault diagnosis are important topics in the field of health monitoring. For the sake of high reliability, EV's powertrain needs two classes of online condition monitoring in real time: the first-class is the prediction that does not directly cause failures, such as battery remaining useful life (RUL) prediction; the second-class is the online real-time

monitoring also before faults occur, such as motor bearing damage fault monitoring. The fault diagnosis can achieve the effect of the second-class condition monitoring, that is, real-time identification of fault occurrence, type and location. There are many techniques that are sufficient to support powertrain condition prediction and fault diagnosis. It is mainly divided into conventional methods based on mathematical models and data-based method. However, conventional methods usually lack not sensitivity to fault time and are hard to make a prognosis for fault time. Therefore, academia and industry gradually began to use AI-based technology to overcome and solve these problems. Almost all AI-based methods require the support of signal data. For example, the methods in [2], [3] collect and utilize vibration signals; papers [4], [5] rely on motor current signals to identify health status. These two types of signals are also the most common sources of data. Besides, researchers have also tried using thermal images [6], [7], [8], infrared information [9], [10] and acoustics signals [11], [12], etc. Particularly, paper [13] suggests a combinational-logic-based method to confirm the possibility of EV powertrain fault appearance. They are widely used to diagnose many types of faults in motors, inverter, batteries and even gearboxes [4], [5].

In EVs, powertrain components require the intervention of health monitoring methods to reduce the possibility of catastrophic issues and losses, as well as provide warnings for predictive maintenance. In other words, some AI techniques could make the electric powertrain more intuitive and transparent, allowing people to early deal with possible failures. Therefore, people need the help of AI technology to solve many problems in the powertrain system. This is also inseparable from the advantages of machine learning (ML) and deep learning (DL) algorithms.

In past years, expert systems [14], fuzzy logic systems [15] and simple fully-connected artificial neural networks (ANN) were early attempts at powertrain troubleshooting. Recently, modified and improved supervised learning algorithms based on decision tree (DT), support vector machine (SVM), some ensemble algorithms (such as adaptive boost (Adaboost), random forest (RF), etc.) and some deep learning methods are often employed. Unsupervised learning algorithms such as autoencoder (AE), K-nearest neighbours (K-NN) and some clustering algorithms are able to skip the step of manual data labeling during training. Usually, neural network-based methods such as multilayer perceptron (MLP), convolutional neural network (CNN), deep belief network (DBN), recurrent neural network (RNN) (including long and short-term memory network (LSTM) and gated recurrent unit (GRU)) are becoming more and more

popular among researchers because these can avoid the feature extraction stage (typical application is in Fig. 1-2). Many scholars have made partial reviews on the application of AI technology in the monitoring of machine health conditions. For instance, [16] summarizes the induction motor (IM) as the target, not for EV operation, nor for the entire electric powertrain, and the review has outdated. In [17], the field described is relatively broad, and no detailed discussion is given for the application in specific EVs. [18] reviews emphatically from the perspective of rotating machinery, and the diagnosis methods for PE components are not included. Overall, there are insufficient review works on AI technique-based condition monitoring and fault diagnosis for the whole EV powertrains. In addition, the intrinsic properties (e.g., advantages and challenges, etc.) of the AI-supported methods were not clearly discussed in comparison with the traditional ones previously. Hence, it is significant and urgent to carry out a relatively comprehensive review to provide guidelines for future study concerning AI-supported health improvement of EV powertrains.

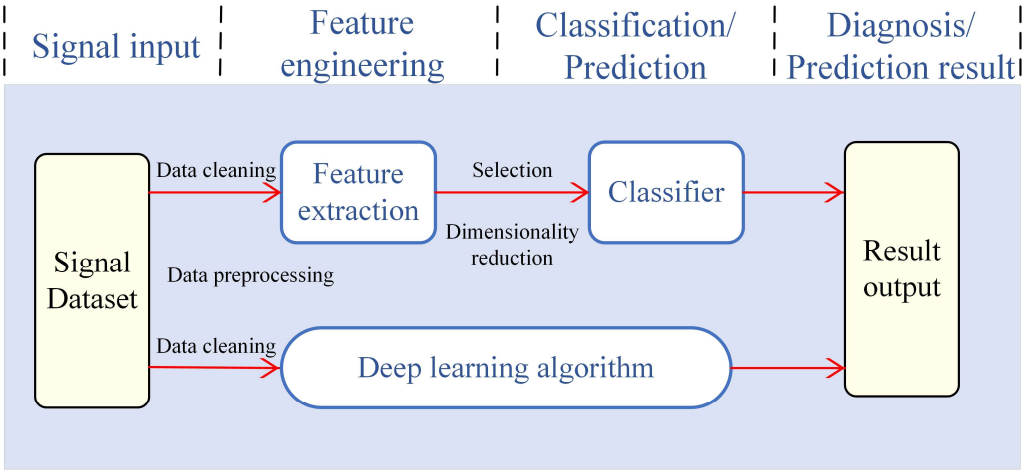


Fig. 1-2. Typical applications of AI approaches for condition monitoring and fault detection of electric powertrains.

This research focuses on AI-supported approaches (especially health condition monitoring and fault diagnosis techniques) for EV powertrain components safety improvement applications. The motor is the most important part of an EV powertrain. Therefore, the objective of this thesis research is on the application of motor-related components, such as motor bearings, rotor permanent magnets in the motor and IGBTs in the inverter used for motor control. The progressive objectives of this study can be summarized as follows:

i. To **review AI-based fault detection and condition monitoring approaches for EV powertrains to reduce personal and maintenance cost and ensure human's life**. Electric powertrain used in electric vehicles (EVs), which is constituted by motor, transmission unit, inverter and battery packs, etc., is a highly-integrated system. Its reliability and safety are not only related to industrial costs, but more importantly to the safety of human life. This review contributes to comprehensively summarizing artificial intelligence (AI)-based/AI-supported approaches in EV powertrain condition monitoring and fault diagnosis that can be used in EV applications. The application of AI on PE in EV is a new attempt, which can solve many issues with better performance than traditional methods, and even achieve functions that the conventional methods cannot achieve.

ii. To **propose an AI-based method to solve the health monitoring problem of mechanical bearings in electric powertrain.** Bearing, as a vital component in electric powertrains, is increasingly used globally such as in electric vehicle (EV). Their damages and faults may bring huge cost loss to owners and even threaten personal safety. The AI-supported inferable deep distilled attention network (IDDAN) method is proposed to solve this problem. Firstly, it attempts to apply the self-attention-based network to build an advanced fault diagnosis model. Secondly, this paper optimizes the structure of networks through knowledge distillation (KD) technique to require a lighter and fast model. Thirdly, this paper proposes a new data augmentation strategy for 1-D vibration signals to provide large-scale pre-training samples for IDDAN.

iii. To **develop a deep learning based permanent magnet temperature (PMT) estimator without target machine labelled data.** PMSMs are vulnerable to damage from high temperatures, which can cause the demagnetization of the permanent magnets (PMs). To prevent demagnetization, accurate real-time estimation of the PM temperature (PMT) is crucial. The PMT measurement in practical applications is still challenging because the PM is in the rotor and the rotor is rotating. Firstly, due to the difficulty in acquiring labeled PM data, lumped-parameter thermal network (LPTN)-based simulation provides a convenient way to obtain a large amount of simulation data. Secondly, this simulation dataset serves as the source domain data to train the proposed modified Transformer-based long sequence time-series regression (MT-LSTR) model, which generates coarse predicting labels (CP-labels). The CP-labels can be used to label the target domain dataset collected from a real interior permanent-magnet synchronous

motor (IPMSM). Finally, the proposed domain-adversarial adaptation regression (DAAR) model is trained with both feature-level and domain-level adaptations to obtain a well-trained online PMT estimator.

iv. To **develop a novel double pulse test (DPT) efficiency enhancement method, which can estimate and supplement switching transient waveforms of all working conditions.** Thereby, dynamic electrical characteristics of the IGBT are obtained by estimated waveforms of DPT. Double pulse is a basic experimental method for analyzing the dynamic characteristics of power switching devices, which runs through the research and development of devices, applications and the design of drive protection circuits. IGBT dynamic electrical characteristics are of great significance in practical application and are usually obtained through DPT. However, DPTs of IGBTs under various working conditions is time-consuming and laborious. Traditional estimation methods are based on detailed physical parameters and complex formula calculations, making deployment difficult. We propose a multimodal attention fusion network (MAFN) to capture and fuse the features of switching transient waveforms between different positions thereby improving the expressive power and performance of the model. On the other hand, this method first proposes to utilize the graph convolutional network (GCN) to embed DPT data under multiple working conditions into graph structure, which can use the graph structure information to fuse the features of spatially correlated working conditions data to obtain reliable estimation result.

1.2 Dissertation Outline

The dissertation consists of seven chapters. These chapters are all focus on the objects in EV motor: bearing, rotor PM and inverter. An outline of the structure is organized as follows:

Chapter 1 introduces the motivation of this study. Besides, the main objectives and a brief introduction of the research contents are illustrated.

Chapter 2 reviews comprehensively AI-based fault detection and condition monitoring approaches for EV powertrains. This review is the first contribution to comprehensively summarize both the feature engineering methods and artificial intelligence (AI) algorithms (including machine learning, neural networks and deep learning) in electric powertrain condition monitoring and fault diagnosis approaches and also explains the motivation,

advantages, shortcomings and challenges of AI-supported methods through summary, classification, comparison and quantitative analysis of cases between conventional methods and AI-based methods.

Chapter 3 designs an inferable deep distilled attention network (IDDAN) method which is a self-attention mechanism and transfer learning-based method to diagnose and classify multiple bearing faults in various motor drive systems efficiently and accurately. Compared with convolutional networks, the self-attention-based network can better extract the global feature information and easier to benefit from large amounts of pre-training data. Its significance is to accurately classify various faults of the target machine when the labeled data of the target machine is not enough to directly train the diagnosis model. Experiments show that the self-attention mechanism-based model is more likely to benefit from large-scale data. After testing, compared with many methods and other exist similar methods, the proposed method achieves higher classification accuracy and better performance.

Chapter 4 designs a simulation-driven unsupervised transfer learning approach that leverages the domain-adversarial adaptation regression (DAAR) model to enable online and accurate estimation of PMT. This PMT estimator can transferred knowledge from the model that trained by simulated data. In experiments, simulation data generated from MATLAB/Simulink is used for offline training, while the effectiveness of the proposed method is validated using the IPMSM temperature dataset provided by Paderborn University (PU).

Chapter 5 designs a novel DPT efficiency enhancement method based on graph convolution network (GCN) and feature fusion technology, which can estimate and supplement switching transient waveforms of all working conditions. Thereby, dynamic electrical characteristics of the IGBT are obtained by estimated waveforms of DPT. The method is verified to be effective and accurate on real dataset collected on two batches of IGBTs.

Chapter 6 concludes all chapters of this dissertation and discusses the future trends for this research field. The future works includes that the use of advanced AI algorithms. For solve more types of engineering issues, it is necessary to apply more novel algorithms to solve practical industrial problems.

Chapter 2 Review of AI-based Applications for EV Powertrain Safety Improvement

Electric powertrain used in electric vehicles (EVs), which is constituted by motor, transmission unit, inverter and battery packs, etc., is a highly-integrated system. Its reliability and safety are not only related to industrial costs, but more importantly to the safety of human life. This review contributes to comprehensively summarizing artificial intelligence (AI)-based/AI-supported approaches in EV powertrain condition monitoring and fault diagnosis that can be used in EV applications. The application of AI on PE in EV is a new attempt, which can solve many issues with better performance than traditional methods, and even achieve functions that the conventional methods cannot achieve. This review is the first contribution to comprehensively summarize both the feature engineering methods and artificial intelligence (AI) algorithms (including machine learning, neural networks and deep learning) in electric powertrain condition monitoring and fault diagnosis approaches and also explains the motivation, advantages, shortcomings and challenges of AI-supported methods through summary, classification, comparison and quantitative analysis of cases between conventional methods and AI-based methods.

AI-based methods usual contains two parts: data preprocessing (feature engineering) and AI algorithm. The following sections include feature engineering at section 2.1 and AI methods at section 2.2-2.4.

2.1 Feature Engineering

Feature extraction is usually used to analyse the waveform signal such as current signal and vibration signal [19], [20], but if the extracted feature information is not completely contributed to the fault classification, it is necessary to use feature selection technology to find representative features to achieve the purpose of reducing the feature size and enhancing fault diagnosing efficiency [19].

2.1.1 Feature Extraction

The step after data preprocessing is often to extract features of the vibration signal, current signal, etc., which extract the fault information contained in the signal. The most famous and important method in signal processing is the Fourier transform (FT), which is widely used signal processing because it has the function of transforming the signal back and forth from the time domain to the frequency domain and helping analyze the components of the signal. For the signal that is discretized into discrete sampling points, the FT becomes the discrete Fourier transform (DFT). However, due to a large amount of DFT calculation and the high computational complexity, the earliest version of fast Fourier transform (FFT) was proposed by J. Cooley and T. Tukey in 1965 [21], which reduced the amount of calculation by several orders of magnitude. This allows computers to process signals more quickly, thereby promoting the rapid development of communications and signal processing. FFT has made great progress in fault diagnosis methods based on Fourier analysis [22]. Yang *et al.* [23] applies FFT in MCSA to analyze the stator current data of IMs and calls the obtained feature the FFT-ICA feature of stator current. Romero-Troncoso *et al.* [24] uses the improved FFT after fractional resampling for IM periodic monitoring tasks. In the paper [25], FFT banded RMS value input CNN for gearbox fault identification. FFT is also used to extract the fault spectrum characteristics of the three-phase current when diagnosing the multiple insulated gate bipolar transistors (IGBTs) open-circuit faults and current sensor faults in the three-phase pulse width modulation inverter [26]. Under necessary circumstances, quantum Fourier transform (QFT) can accelerate the FFT exponentially [27]. Short-time Fourier transform (STFT) contribute to deal with the problem of FT about losing all time-domain information. E. H. E. Bouchikhi [28] applies STFT to analyse IM stator current to diagnose bearing faults.

Continuous wavelet transform (CWT) introduces the wavelet mother function to achieve dynamic resolution in the time domain and frequency domain. Using CWT can get the complete time-frequency domain information of signals, avoiding the loss of information of the original signal to the utmost extent [29]. Discrete wavelet transform (DWT) is a new type of spectrum analysis tool that discretizes the scale and translation of basic wavelets. It can not only observe the frequency domain characteristics of local time-domain processes but also observe the time domain characteristics of local frequency domain processes, so even those non-stationary processes can be transformed and processed well. The type and number of

features used in fault diagnosis methods could be extracted by DWT, but when extracting fault characteristics of motor bearings, key information may be lost near the fault characteristic frequency. Some research on IM proves that DWT can make fault identification more accurate than traditional FFT [30]. Therefore, it is also one of the time-frequency domain techniques preferred by researchers. In [31], the author uses DWT to diagnose faults in motor transient operation. Frequency domain methods often fail to detect faults from nonstationary signals, such as bearing vibration signals [32].

In addition to STFT and wavelet transform (WT), other signal time-frequency domain analysis techniques, such as wavelet packet transform (WPT), Wigner-Ville distribution (WVD) [33], empirical mode decomposition (EMD) [34], empirical wavelet transform (EWT) and spectral kurtosis (SK) analysis, etc. are also widely employed.

WT is a very successful method in time-frequency domain EV powertrain health monitoring, which can also reduce noise in the noisy working environment. However, the high-frequency band is not split where the modulation information of machine fault exists. WPT can extract key components through a band-pass filter to decompose the signal into different levels, which is helpful for the correct extraction of fault feature information. In paper [35], the CAA algorithm was used to improve the feature correlation according to the weight of the feature after the author extracted the fault features used by this method from the time domain, statistical feature, and frequency domain through wavelet packet transform (WPT). Then it reduces the dimensionality of fusion features through principal component analysis (PCA) and the support vector machine (SVM) is a diagnostic classifier. This is the basic process of a fault diagnosis method based on fault characteristics. WPT obtains a variety of the wavelet packet feature quantities through its multi-scale time-frequency analysis: energy, fluctuation coefficient, skewness, and margin factor. To pursue higher diagnostic accuracy, T. Gao [36] uses generalized discriminant analysis (GDA) to fuse wavelet packet features to eliminate redundant information. However, the feature extraction of WPT will not be able to extract effective fault information due to lack of adaptability [37], [38]**.** The WFM method combines WPT with manifold learning, which can suppress background noise in the time-frequency domain and enhance the signal [39]. The small-amplitude transient pulse at the beginning of the bearing failure can be extracted by the WFM method [40].

Methods such as FFT, STFT and WVD are more suitable for processing linear signals with high stability. WT has great advantages in analyzing non-stationary and non-linear signals, but it also has its shortcomings such as the certain bands of the defect information and the selection of the base function [41]. In addition, there are EMD, local mean decomposition (LMD) [42] and other adaptive methods are applied to the time-frequency domain analysis of powertrain fault signals. EMD is a signal decomposition method that can adaptively decompose any signal into a set of intrinsic module functions (IMF) with different frequency characteristics. This is a major advancement in analysing non-stationary signals. Ge *et al.* proposed a rolling bearing fault diagnosis method based on ensemble EMD (EEMD) [43], wavelet semi-soft threshold (WSST) signal reconstruction and multi-scale entropy (MSE). First, use EEMD to decompose the bearing vibration signal into IMF, and then use the Pearson correlation coefficient to filter the high-frequency IMFs that contain more noise information. The WSST method denoises the high-frequency IMFs for signal reconstruction. The feature vector is constructed by using the MSE method to calculate the MSE value of the reconstructed signal. In paper [44], the author uses the EMD-SVM method to diagnose the neutral point clamped three-level inverters (NPC). The EMD-based feature extraction method is widely used in the intelligent diagnosis of bearings and rotating machinery [45], [46] in EV, and the stator as well [47]. However, EMD methods decompose the signal according to its time-scale characteristics. If the signals collected from different sensors are processed under the condition of without a pre-set basis function, the position of the fault feature in the EMD will be uncertain. Therefore, the best choice is to use WT related methods with a fixed wavelet basis function to decompose all signals when using deep learning diagnosis methods.

Empirical wavelet transform (EWT) combines the adaptability of EMD decomposition and the advantages of WT, which is a useful adaptive tool for vibration signal processing and can also decompose the original signal into different modes [48]. In the paper [49], there is a hybrid automatic bearing fault detection method that combines EWT and fuzzy logic system (FLS) to locate the early degradation of the bearing state under different working conditions. Paper [50] proved that EWT is more effective than EMD in the diagnosis of rolling bearings. However, the EWT method still has some problems. On the one hand, it is difficult to determine adaptive and robust boundaries of the EWT segments; on the other hand, the filtered signal still has noise and redundant vibrations, which will mask weaker fault features, which is not conducive

to detecting early-stage faults [51]. There is also spectral kurtosis (SK) analysis in time-frequency domain technology that can handle both stationary and nonstationary signals [52].

To take advantage of the powerful ability of convolutional neural network (CNN) (Fig. 2-1) to automatically extract features, time-domain vibration signals which include adequate once or more types of faults information are transformed to two-dimensional (2D) grey-scale images through continuous wavelet transform (CWT) [53]. In addition, CNN is good at processing continuous wavelet transform scalogram (CWTS) generated by CWT. S. Guo [54] adds the Pythagorean spatial pyramid pooling (PSPP) layer to the top layer of CNN, so that the fault features obtained by the PSPP layer from CWTS (Fig. 2-2) can be passed to the convolutional layer below for secondary extraction. Feature extraction methods related to neural networks include AE, but traditional AEs cannot stably obtain various meaningful signals from vibrations. Sparse autoencoder (SAE) is another neural network-based method for failure features extraction. The author of the paper [55] proposed to use normalized sparse autoencoder (NSAE) constructs local connection network (LCN), namely NSAE-LCN. NSAE-LCN overcomes two shortcomings of traditional autoencoders: machine feature extraction may learn similar features and shift variant properties may lead to machine health misclassification.

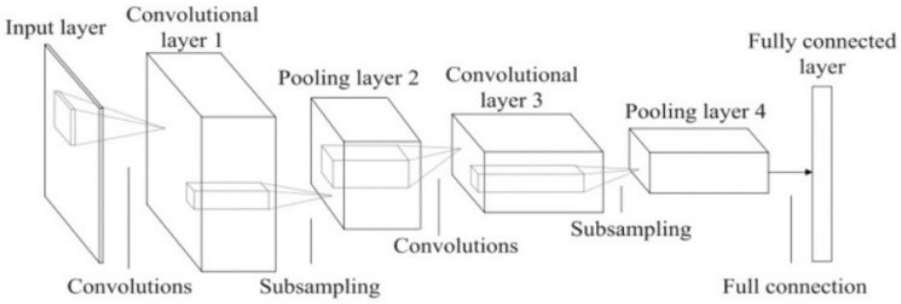


Fig. 2-1. Schematic diagram of the classical CNN.

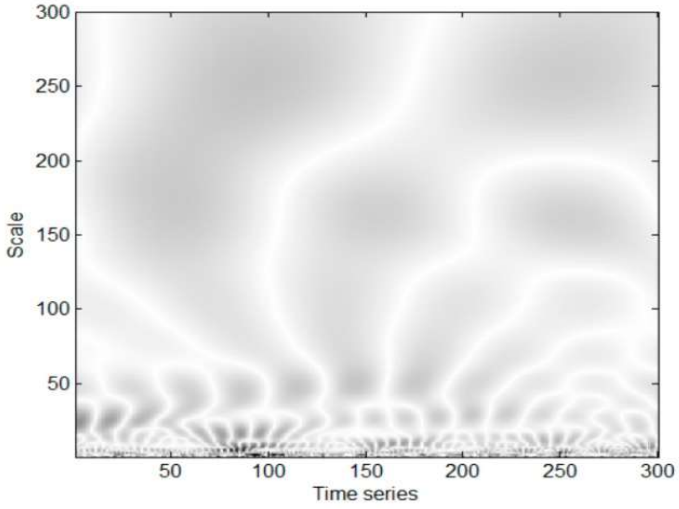


Fig. 2-2. Continuous Wavelet Transform Scalogram (CWTS) (The darker pixels correspond to larger wavelet coefficients) of a ball fault bearing signal [54]**.**

Entropy-related methods are also an important tool for feature extraction in fault diagnosis, and there are now many derivative methods. A diversity entropy (DE) based novel method called multiscale diversity entropy (MDE) is extended to deal with multiscale analysis for a comprehensive feature description by combining with the coarse graining process [56]. This entropy method is designed for fault diagnosis of rotating machinery such as rotor and gearbox [57]. In the paper [58], the frequency band entropy (FBE) based on information entropy (IE) and STFT are introduced to extract the fault feature frequency of rolling element bearings. This method in the paper [59] is improved by WPT based on the Daubechies wavelet. Some entropy methods can play an auxiliary role in the construction of feature vectors. Permutation entropy (IPE), multi-scale entropy (MSE), multi-scale permutation entropy (MPE), weighted permutation entropy (WPE), and fine-sorted dispersion entropy (FSDE) calculate the entropy values of signal reconstructed by EMD and variational mode decomposition (VMD) [60], [61] to construct the feature vectors which will be used as the input data and classification basis for subsequent AI fault classification methods such as types of support vector machine (SVM). Summary of pros and cons from all literatures mentioned before are shown in TABLE 2-1.

TABLE 2-1 Summary of Pros and Cons of Feature Extraction Methods

Method	Related Papers	Advantages	Shortcomings

FT/DFT/FFT	[22], [23] [24], [26] [27]	• To be suitable for signal with characteristics of discreteness and finiteness • Computer calculation convenient • Good frequency domain resolution	• Time domain information completely lost • The transient signal of the fault characteristic cannot be displayed • Low noise resistance • Sensitive to low signal-to-noise ratio • Overlapping of closely-located spectral components, nonstationary signals and spectral leakage
STFT	[28]	• Time domain information is retained after transform	• Fixed time and frequency resolution • Choice difficultly of window width
CWT	[29]	• Dynamic time and frequency resolution	• Computer cannot perform CWT on the continuous signal
DWT	[30], [31], [32]	• Computer processing acceptable • Controllable wavelet decomposition scale	• Less retained information than CWT • The loss of the invariant translation property
WPT	[35], [36], [37], [38]	• Decomposition of high frequency components more detailed than CWT	• Usually more computational time than CWT
WVD	[33]	• High resolution in time and frequency domain • To combine with EMD and EEMD to analyze signal IMFs	• Cross-terms generating frequently
EMD	[34], [43], [44], [45], [46], [47]	• Signal decomposition based on its time scale characteristics	• Uncertain location of the fault features • To consume a lot of computing resources to decompose the calculation steps • To be suffered from the mode mixing, distorted components and the end effects
VMD	[62], [63]	• To avoid effectively the mode mixing problem of EMD and LMD • Higher computational efficiency and good noise robustness	• To setting the number of decomposition modes and the penalty factor difficultly • Loss of information or excessive decomposition probably occur
LMD	[42], [64]	• To get the calculation of instantaneous frequency (IF) and instantaneous amplitude without Hilbert transform	• Difficult data processing from multiple channels in health monitoring

Hilbert–Huang transform (HHT) [65]	[66], [67]	• To calculate the instantaneous amplitude and instantaneous frequency of intrinsic mode functions (IMFs) • To highlight the IMF time-frequency representation and visualize the modulation of each component	• To be suffered from end effects and mode mixing
Singular value decomposition (SVD) [68]	[69], [70]	• Non-linear filtering to effectively eliminating random noise components	• The quality of the noise reduction signal affected by different reconstruction orders • Information loss when weak faults
EWT	[48], [49], [50], [51]	• High efficiency, accuracy and reliability. • To reduce signal noise	• To determine difficultly adaptive and robust boundaries of the EWT segments • Noises and redundant vibrations exist after filtering
CNN	[53], [54], [71]	• To extract features automatically • Adjustable extraction level	• To find the most suitable network structure difficultly • To be more suitable for image signals

In summary, different feature methods has different advantages and disadvantages. Therefore, in practical applications, the appropriate method needs to be selected based on the actual situation.

2.1.2 Dimensionality Reduction and Feature Selection

Principal Component Analysis (PCA) prepares for fault diagnosis before entering the dataset and convert existing features to low-dimensionality, reduce feature space, and avoid high-dimensional data redundancy [72], [73]. PCA use a group of fundamental function to reasonably optimize the minimum error of data model. The author introduces the algorithm of combining independent component analysis (ICA) in the kernel technique to improve the feature extraction of condition monitoring and fault diagnosis in IMs. ICA is formulated in the kernel-inducing feature space and developed through two-phase kernel ICA algorithm: whitened using kernel principal component analysis (kernel PCA) plus ICA. Kernel PCA spheres data and makes the data structure become as linearly separable as possible by virtue of an implicit nonlinear mapping determined by kernel. ICA seeks the projection direction in the kernel PCA whitened space, making the distribution of the projected data as non-Gaussian as possible. In [73], the performance of the classification process due to the choice of kernel

function is presented to show the excellent characteristic of the kernel function. And this research demonstrates the clustering feature using ICA is better than PCA does. In the article [74], a machine learning-based fault diagnosis method for single VFD-fed IMs have been developed. Gharavian *et al.* [75] compared the FDA-based and PCA-based feature selection results in the fault diagnosis of EV gearboxes. PCA may not be the optimal feature reduction method in a specific problem.

PCA and ICA are the most popular unsupervised learning feature dimensionality reduction techniques. For feature selection through supervised learning methods, Linear Discriminant Analysis (LDA) and Random forest Decision tree (RFDT) can be used [76]. What the difference is PCA follows the direction of maximum variance for optimal reconstruction, whereas LDA is to obtain the optimal low-dimensional representation result of the original data set by maximizing the between-class scatter matrix meanwhile minimizing the within-class scatter matrix. Since LDA operates based on labelled information, it can obtain better results than PCA when there are sufficient labelled samples [77]. LDA has been used in lots of fault detection works. Many works of literature explore the application of the combination of LDA and MCSA methods to various parts, online monitoring of operating conditions through voltage and current analysis, and estimation of the severity of the fault through the magnitude of the sub-harmonic amplitude. Paper [78] solved the problem of difficulty in detecting faults with sub-harmonics in the MCSA-LDA method, using harmonic amplitude instead of sub-harmonics as fault detection and classification function. Jin *et al.* [77] proposed a combination of trace ratio linear discriminant analysis (TR-LDA) and VA, which was applied to the bearing fault analysis of IM and brushless DC motors. TR-LDA is a variant of LDA based on the TR criterion. It is an excellent solution to the TR problem and can be extended to deal with non-Gaussian data sets encountered in many real-world fault diagnosis problems. LDA can also be used in the global spectrum analysis of vibration signals to improve the effect of fault diagnosis of ball bearings [79].

In addition, AI-based optimization algorithms have outstanding performance in feature selection, such as particle swarm optimization (PSO) and differential evolution (WBDE) [80]. Lee *et al.* [81] pointed out through experiments that PSO is the key technology to find the optimal weight of damage-sensitive feature vectors in the bearing system. In [82], the GA is utilized to reduce the number of features and select the most important ones from the feature database to reduce power consumption and reduce model computational complexity. In the

paper [83], GA combined with random forest (RF) algorithm to produce a new method RFOGA for machine fault detection, which is more accurate than only RF.

2.2 Condition Monitoring and Fault Diagnosis for Motors

The motor is the vital part of the EV powertrain system. With respect to the demand for motor performance in different environments, the motors should be endowed with fast speed, high output power and high efficiency. Now, the mainstream motors used in EVs include PMSMs and IMs [84], which are mainly discussed in this section. It is important to ensure their correct operation.

As the working environments and working conditions of the EV motors are ever-changing, the motor parameters change constantly. For this situation, AI-based methods are well-suited to accurately diagnose faults of different parts (stator winding, rotor PM, bearing) of EV motors and can show marked performance [85]. The stator, rotor and bearings will be described separately for the motor. Table 2-2 shows the main contributions and applied AI algorithm of each literature with used signals and feature methods.

TABLE 2-2 Summary of Contribution and AI Algorithms of Each Paper with Signals and Feature Methods

Parts	Paper	Signals	Feature Method	AI Algorithm	Contributions
Bearing	[53]	Vibration	CWT-CNN	CNN-RF	• To converted into 2D gray-scale images by CWT • the ensemble of multiple RF classifiers used
	[54]	Vibration	CWTS-CNN	PSPP-CNN	• To suggest adding PSPP layer on the top of CNN
	[86]	Vibration	SDAE	SDAE-LSTM	• To discover the initial anomaly in advance • Sliding window algorithm used in SDAE • To use LSTM for predict the behavior of data in advance
	[87]	Vibration	DWAE	DWAE-ELM	• To enhance unsupervised learning ability by constructing DWAE • To diagnose multiple types and levels of rolling bearing faults
	[71]	Vibration	CNN	CNNEPDCNN	• To propose new bearing fault detection model CNNEPDNN • Features with different discriminative abilities fused by CNNEPDNN

[4]	Stator current	WPT	SVM	• To propose a hybrid method based on optimized stationary WPT • SVM combining with artificial immune system
[88]	AE	DWPT	MCSVM	• To take the relative wavelet packet energy (RWPE) and wavelet packet node kurtosis (WPNK) as fault characteristics • To adopt OAA MCSVM classifier
[89]	Vibration	-	RF	• To solve the problem of slow diagnosis and repeated voting in traditional RF • RF algorithm processed on SPARK platform
[2]	Vibration	WPT	RF	• To propose a novel hybrid RF approach • To enhance the anti-noise ability of the RF algorithm
[90]	Vibration	1D-LBP	GRA	• To suggest a new method of bearing fault diagnosis combining 1D-LBP and GRA
[91]	AE	TVMREA	GA-DFA	• To use GA to select the most discriminative features • GADFA improves the classification accuracy rate by 7.3%-46.6%
[92]	Rotor speed	-	AVPCA	• To propose a concise diagnosis method based only on rotor speed • To suggest a modified AVPCA algorithm
[93]	Vibration	Deep residual network	DNN	• Simultaneous fault location detection and severity identification • To visualize DNN
[94]	Vibration	ResNet	DeepResNet	• To propose IFMs transfer method • To use ResNet to solve the problem of CNN gradient disappearance
[95]	Vibration	Short-time periodogram	k-NN	• To detect faults based on GMS • Diagnosis is based on the same distribution of training data and test data
[96]	Vibration	1-D CNN	DCTLN	• Transfer learning based • To learn and recognize features automatically by 1-D CNN
[97]	Vibration	SK based, MFCCs	CNN	• To use MFCCs and delta cepstrum to feature extraction • Filter method based on SK to suppress noise
[98]	Vibration	VCN	VCN	• To design 1-D vision ConvNet (VCN) • Different working conditions adaptable

	[99]	Vibration	CNN	CNN-SVM	• To construct CNN-SVM diagnostic model • To apply residual learning to prevent CNN gradients from disappearing • To use transfer learning to solve the problem of insufficient data
Rotor	[100]	Motor current	SWPT	SVDD-SVM	• To embed SVDD in MSVM algorithm • The training decision boundary based on hyper-sphere
	[101]	Stator current	RUWPT	DAGSVM	• To combine RUWPT and DAGSVM • To extract one new parameter by RUWPT
	[102]	Stator current	Hilbert transform (HT)	MLP	• To replace DFT by HT • the amplitude and frequency of the harmonic from stator current envelope (SCE) spectrum as NN inputs
	[103]	Vibration	DBN	DBN	• Vibration imaging as feature source • DBN applied to feature extraction and classification
	[104]	Vibration	DC-CNN	DC-CNN	• Vibration imaging as feature source • DC-CNN applied to feature extraction and classification
	[8]	Thermal images	CNN	CNN	• Thermal Image as input of CNN • To apply parameter transfer for limited training data
	[105]	Vibration	SDP-CNN	SDP-CNN	• Vibration signal transfer to SDP image
	[106]	Displacement	SVI-MCNN	SVI-MCNN	• To construct shaft vibration image (SVI) through displacement signal
Stator	[107]	Vibration	DTCWT	PNN	• Orthogonal Least Squares Regression (OLSR) applied to optimize the size of the PNN • DTCWT applied
	[108]	Stator current	-	ANN	• To binary code input signal • To design new process addition and mean of the set of same rank
	[109]	Stator current	-	PSO-ANN	• PSO optimized ANN • Simulation in SIMULINK environment
	[110]	Stator current	CNN	CNN	• The original three-phase current is input • Steady-state current fault detection • No load affect
	[111]	Stator current	ST	SVM	• Two SVM model for two faults • ST applied

[112]	Stator current	SWT, DWT	ANN	• Several ANN build a modular neural network (MNN) • Feature extraction by combining SWT and DWT
[85]	Stator current	DWT	ENN	• To detect and locate faults early • For ITSC fault • Discrete Wavelet Energy Ratio (DWER) as feature

In summary, although the same task is solved, the reference signals that can be relied on may be different. The specific reference signal to be used needs to be determined based on the actual situation and practical application.

2.2.1 Stator Inner-turn Short-circuit Fault Diagnosis

Stator faults mainly include inner-turn short-circuit (ITSC), phase-to-phase short-circuit, and ground short-circuit. According to surveys [113], the ITSC accounts for about 33 percent of all IM failures.

EV wheel traction motors have relatively high requirements for power/torque density, and their working conditions are complex and poor, which are more likely to cause serious ITSC fault. When ITSC is serious, it will evolve into phase-to-phase short-circuit fault and ground short-circuit fault, which will lead to high-temperature operation of the motor and irreversible demagnetization of the magnets, and catastrophic accidents will occur. Fig. 2-3 illustrates the principle of ITSC fault, which assumes that one branch of phase C has ITSC faults and the fault winding in series.

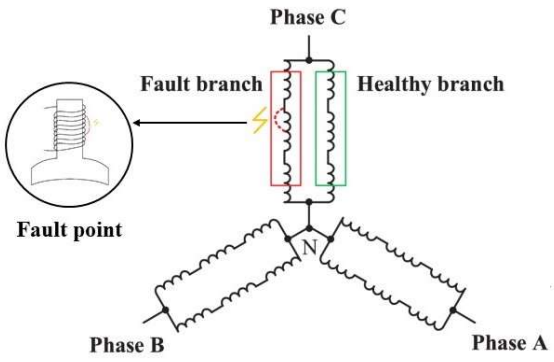


Fig. 2-3. Schematic diagram of ITSC fault of PMSM stator.

For EV applications, the conventional method to detect ITSC faults is based on electromagnetic characteristics and thermal characteristics. Paper [114] introduces a Volterra kernel identification-based on-line fault diagnosis method, which refers to the variation of the stator branch voltage and current in the event of ITSC. However, the disadvantage of this type of method is that it is difficult to detect when ITSC failures are early or weak.

Paper [115] is a method based on the finite element model and thermal network coupling model, which uses an open-loop physics-based back electromotive force (EMF) estimator for EV PMSM ITSC fault detection. Similarly, in paper [116], the ITSC fault state of the PM alternator (PMA) was analysed by proposed the electromagnetic-thermal coupled model-based method. It can realize that the shortcoming includes requiring extremely strong prior knowledge and model knowledge, weak fault identification is difficult, and limited fault degree identification.

Due to no prior knowledge of the process or its model parameters is required, these AI-based data-driven methods have recently been applied to resolve ITSC failures. It is analysed here through a case study containing multiple representative methods. Paper [85] classifies ITSC faults based on the evidence of discrete wavelet energy ratio (DWER) by using Elman neural network (ENN). Paper [107], this researcher employs the dual-tree complex wavelet transform (DTCWT) to obtain fault features from vibration signals and import features into the probabilistic neural network (PNN) whose size is optimized by the orthogonal least squares regression (OLSR) algorithm. Paper [111] uses the Stockwell transform (ST) to extract different fault features in different frequency bands and proposes a parallel and different SVM model to diagnose two ITSC faults simultaneously. The severity of ITSC can also be quantified. Paper [110] quantifies the level of ITSC failure for the line start PMSM (LSPMSM). It can be concluded that, people are accustomed to supervised training on data through deep learning models to avoid model knowledge.

TABLE 2-3 Performance of AI-Based Methods and Suitable Working Conditions

Methods	Application targets	Working conditions	The best accuracy
DTCWT-PNN [107]	IM	Early faults	95.25%
DWER-ENN [85]	IM	Early faults and locating	≈1
SVM-based [111]	IM	2 types of ITSC	96%

CNN-based [110]	LSPMSM	4 types of ITSC	97.75%

As obtained from Table 2-3, mentioned AI-based methods have a good ability to achieve high-precision early diagnosis of ITSC faults and multi-level fault diagnosis. Therefore, the motivation of employing AI-based methods is to make up for the shortcomings of traditional methods.

As for the disadvantage mentioned in [115], it is a good choice to perform calculations and build a new EMF estimator for each machine before using the AI based method in the actual EV motor. Through transfer learning and meta-learning technologies, the problem that the same estimator cannot take care of the diagnosis work of a large number of machines can be solved. This is a possible future trend for solving such faults.

In summary, for the ITSC fault, the training data of most AI-based methods are stator currents, which advantage is that it is very easy to measure and collect when EV driving.

2.2.2 Rotor PM Temperature Monitoring

PM temperature is essential to ensure high performance and reliable control of PMSM for EV applications. Conventional PM temperature estimation methods can be divided into three major types: thermal model-based methods [117], back electromotive force (BEMF)-based methods [118], and high-frequency voltage injection-based methods [119].

Due to the rotation of the rotor, it is hard to measure the PM temperature, especially in actual EVs. In a laboratory environment, PM temperature can be measured by battery-powered equipment, infrared sensors, slip rings, etc. These measurement methods cannot be performed in actual EV driving.

The data-driven PM temperature estimation method needs a huge amount of training data collected from the practical EV powertrains for the AI model [120]. Therefore, the application of AI in the PM temperature prediction of EV rotors still faces huge challenges. In this article, the proposed deep learning-based will solve this problem to a certain extent.

2.3 Health monitoring in Mechanical Bearing

2.3.1 Bearing RUL prediction

RUL predictions of rolling bearings play a vital role in reducing unplanned maintenance fees and improving the reliability, availability, and safety of EVs.

Bearing failure prediction methods mainly include conventional model-based methods, AI-based data-driven methods and hybrid prognostics methods. Among them, the model-based approach is to establish mathematical and physical models to describe the degradation process of bearings. Recently, paper [121] extracts features from the time-domain and time-frequency domain of bearing vibration signals and generates an extended Kalman filter (EKF) estimator by learning and determining parameters. Qian *et al.* [122] proposes an enhanced particle filter (PF) and designs the RUL online prediction framework.

However, the shortcomings and limitations of the model-based conventional method are that it requires a priori knowledge of the estimated parameters, and assumptions and simplifications will lead to uncertainty in the estimation. AI-based data-driven methods can solve this problem.

In paper [123], a multi-objective deep belief networks ensemble (MODBNE) method is proposed, which uses a multi-objective evolutionary algorithm combined with traditional DBN training technology to simultaneously evolve multiple DBNs and combine the final evolved DBNs to build an integrated model for RUL estimation. In addition, RVM and CNN-based models are also used in some papers [124], [125], [126] which can guarantee superior performance in the test.

Table II compares the performance of five state-of-the-art methods using the same vibration data from accelerated degradation tests of rolling element bearings. In the experiment, cumulative relative accuracy (CRA) is used to express the accuracy of the method. The closer the CRA value is to 1, the more accurate the RUL estimation result of the prediction method is. Analyzing the data in the case, the hybrid method seems to provide faster and more accurate RUL predictions.

TABLE 2-4 Quantitative Comparison of Five Different Methods

Categories	Methods	Test target	The best convergence speed rank	The best accuracy (CRA)
	EKF-based [121]		2	0.88

Conventional model-based method	PF-based [122]		3	0.783
AI-based data-driven method	MODBNE [123]	Real experimental data of LDK UER_{204} rolling element bearings	5	0.901
	RVM-based		4	0.657
Hybrid prognostics method	[127]		1	0.961

It should be noticed that the use of AI-based data-driven methods for bearing RUL prediction is still a challenge because it is difficult to obtain marked life test data under real EV operating conditions. Almost all current AI-based research can only be verified in the laboratory.

2.3.2 Multiple fault diagnosis

Motor bearing damages are usually divided into six main types: fatigue, wear, corrosion, electrical erosion, plastic deformation, and fracture and cracking [128]. Each type of damages can have different development states and lead to sudden failure [129]. Rolling element bearing is one of the most used bearings in EV motors. Fig. 2-4 provides the example of bearing failures and faults positions on bearing. As shown in Fig. 2-4, bearing damages can occur at the inner ring, rolling elements, cage, and outer ring.

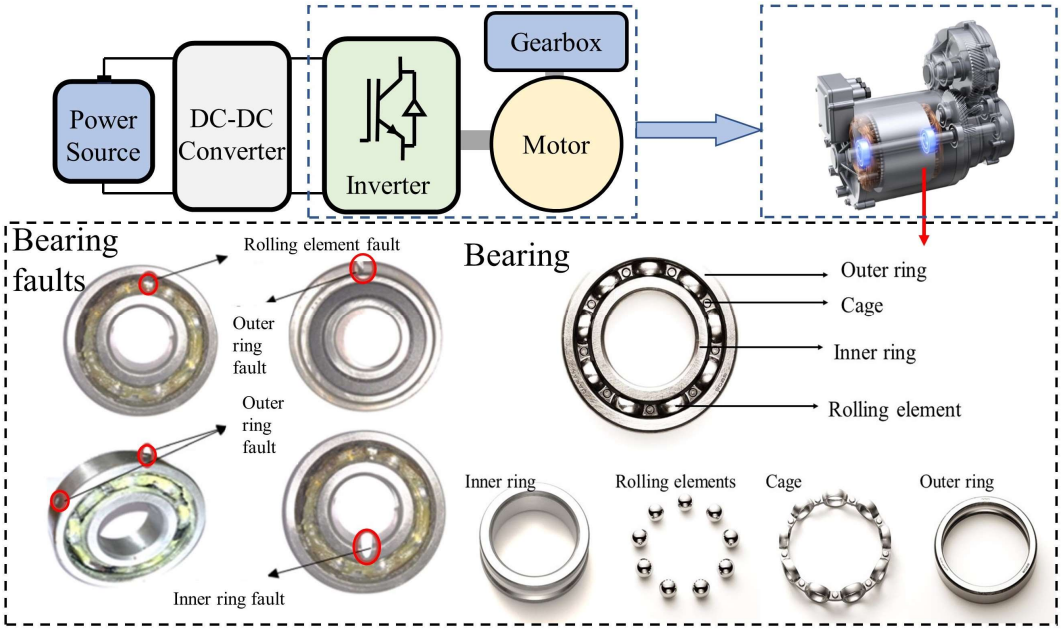


Fig. 2-4. Rolling bearing faults in the structure of EV powertrains.

Recent conventional methods for bearing faults diagnosis mainly include dictionary construction-based methods [130], [131] and the sparse denoising methods in time-domain and time-frequency domain [132]. These methods have unavoidable disadvantages in EV applications:

1) The calculation process of the dictionary construction-based method is complicated and time-consuming, and it is difficult to diagnose online real-time faults in EVs.

2) Building a sparse denoising model for bearing fault diagnosis requires prior knowledge of bearing parameters, which may hinder its application in the EV industry.

3) These conventional methods have many hyperparameters, which makes it difficult to maintain the models.

Therefore, relying solely on traditional PE methods, it is difficult or impossible to clearly detect the specific fault location and fault type. However, during EV driving, it is necessary to know the health status of the bearing in real time to avoid driving accidents in time.

The commonly used machine learning classifiers suited for EV bearing issues are SVM and RF. They are usually only used to classify faults based on existing features but cannot directly extract deep fault features from the signal. This existing feature needs to be manually extracted from the initial signal using some signal processing methods, and then input into the classifier for classification work [4] [88] [89] [2]. However, AI-based approaches which combine feature extraction with AI classification algorithms not only consume a lot of time and manpower for manual data labelling before model training but also lack flexibility for different machines.

CNN is a neural network that has attracted much attention due to its powerful functions. Convolutional layers and pooling layers in CNN have powerful automatic feature extraction capabilities, while fully-connected layers are generally applied for fault identification. The number of layers depends on the complexity of the application task. Here are two typical cases.

Paper [53] uses a combined method of CNNs and machine learning classifiers. On the one hand, LeNet-5 based convolutional layers extracts the shallow to deep fault features of time-domain vibration signals; on the other hand, multiple independent RF classifiers are separated by multiple local and global feature information obtained in different convolutional layers. Fig. 5

show the flowchart of the proposed signal-to-image method in [53]. In paper [54], a special variant of CNN, Pythagorean spatial pyramid pooling CNN (PSPP-CNN), can also deeply extract and analyse the deep fault information from continuous wavelet transform scalogram (CWTS), so as to achieve high-precision fault diagnosis under the condition of variable speed and full speed.

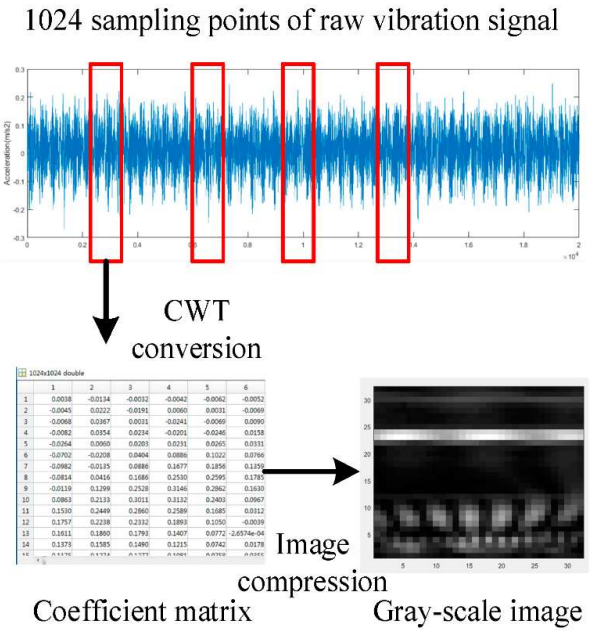


Fig. 2-5. One example of method for signal-to-image conversion [53]**.**

CNN model based could classify fault types directly [54] [97].Meanwhile, CNN could also be a part of diagnosis model as well. In paper [99] and [71], CNN is just as a powerful feature extractor. Therefore, SVM and NNs can also be followed by the convolutional layers for classification of fault types [99] [53]. Fig. 2-6 illustrates fault diagnosis process of convolutional layer with ML classifier. The ML classifier is used to accurately classify the deep fault feature information extracted and output from the feature image by the 2-D CNN model. In the paper [71], the author combines deep neural network (DNN) and convolutional layers to get a fault diagnosis architecture CNNEPDNN for bearings, which recognize faults according to features with different discriminative powers. To address the gradient decay problem of deep CNN models, papers [94] and [93] combine residual structures with convolutional layers to avoid gradient disappearance during CNN training.

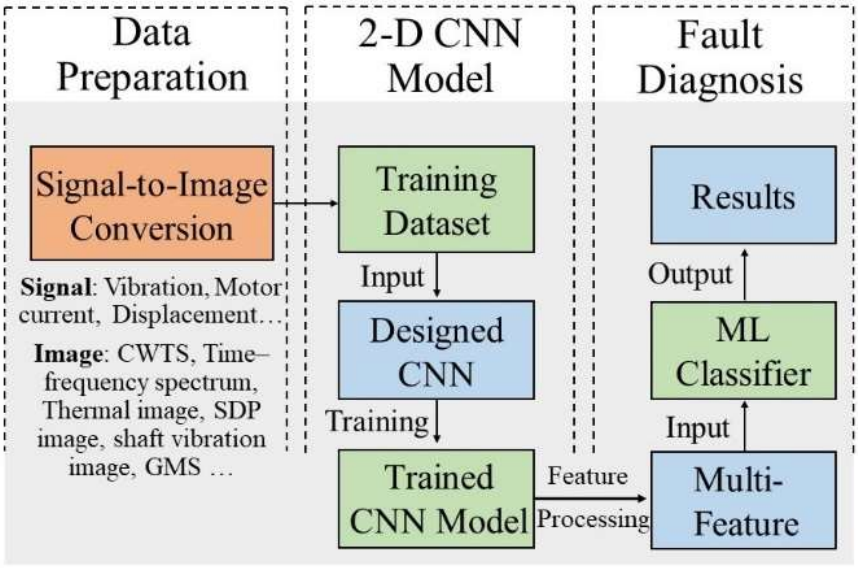


Fig. 2-6. Flowchart of 2-D CNN and ML classifie based fault diagnosis structure.

In order to show the difference between various signal-to-image conversion methods, this paper uses the bearing vibration signal in the CWRU dataset to demonstrate the difference between methods (coutinuous 1-D wavelet transform, short-time Fourior transform, Hilbert-Huang transform, spectrogram, specgram and constant-Q Gabor transform) and is summarized in Fig. 2-7. These methods also work in other collected signals.

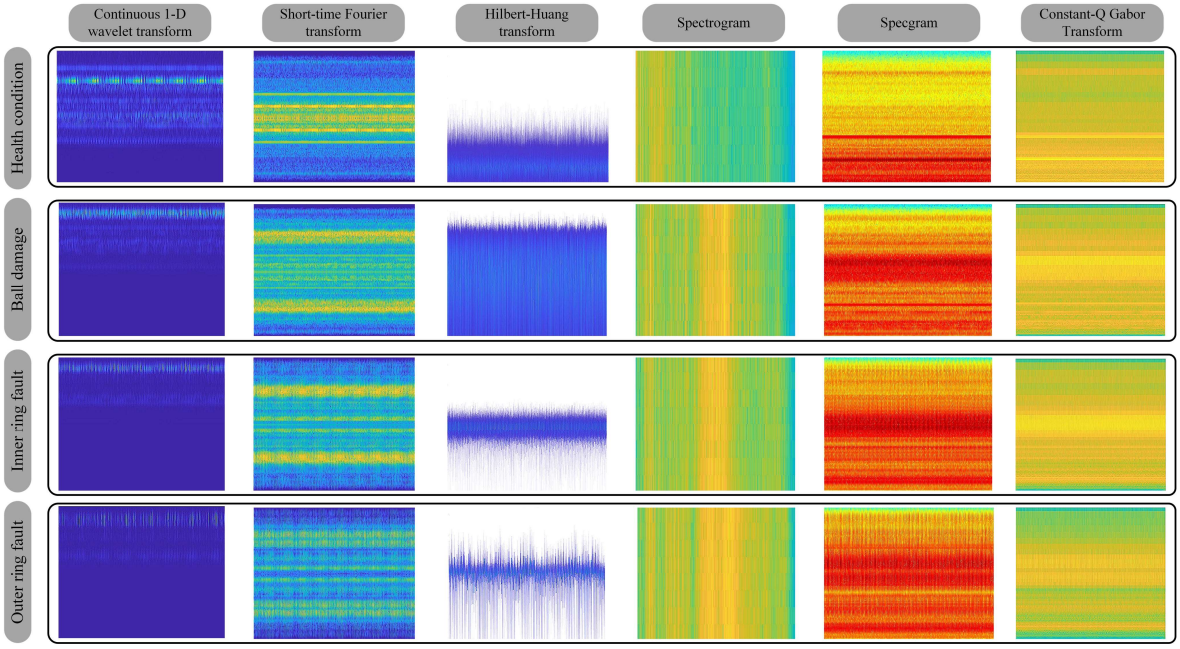


Fig. 2-7. Bearing vibration signal conversion using different signal-to-image methods (i.e. Coutinuous 1-D wavelet transform, short-time Fourior transform, Hilbert-Huang transform, Spectrogram, Specgram and Constant-Q Gabor transform) on CWRU rolling element bearing vibration signal dataset.

The 1-D CNN model was used to extract the deep features and final classification of the one-dimensional signals in the paper [96] and [98]. 1-D CNN can directly process 1-D bearing vibration signals, eliminating the need for a complex signal imaging step. Thence, compared with 2-D CNN, 1-D CNN based model is more suitable for EV real-time applications. The 1-D convolution is to take a dot product between a kernel $\mathrm{w}_c \in R^m$ and the jth segmented signal (jth piece of sub-signal from the whole signal) $s^i_{j-m+1:j} \in R^m$ to get convolution features:

$$c_j = \mathrm{Relu}\left(\sum_{i=1}^{n} \mathrm{w}_c * s^i_{j-m+1:j} + b_c\right) \tag{2-1}$$

where $*$ is a 1-D convolution operator, w_c is referred to as the convolution kernel, b_c is the corresponding bias, n is the number of kernels, and c_j is the j th output point of the convolutional layer. $\mathrm{Relu}(\cdot)$ is an activation function. Fig. 2-8 explains why 1-D CNN is more advantageous in actual EV applications. Table 2-5 aims to summarize and compare the performance of the CNN-based method of the above method with the same dataset which is a labeled bearing dataset from case western reserve university (CWRU). It can be seen from Table 2-5 that 1-D CNN, the accuracy of 1-D CNN with transfer learning technique is slightly lower than other CNN-based methods. This is because other methods use the CWRU dataset as both training data and verification data, but the diagnosis model in paper [96] is a general model which transfers fault-related knowledge from other bearing systems and tests directly on the CWRU dataset. Therefore, it does not mean that the diagnostic effect of the 1-D CNN model is not as good as other CNN models. This is because the performance of the model on different tasks depends on the number of parameters of the model, hyperparameter settings and data quality, etc.

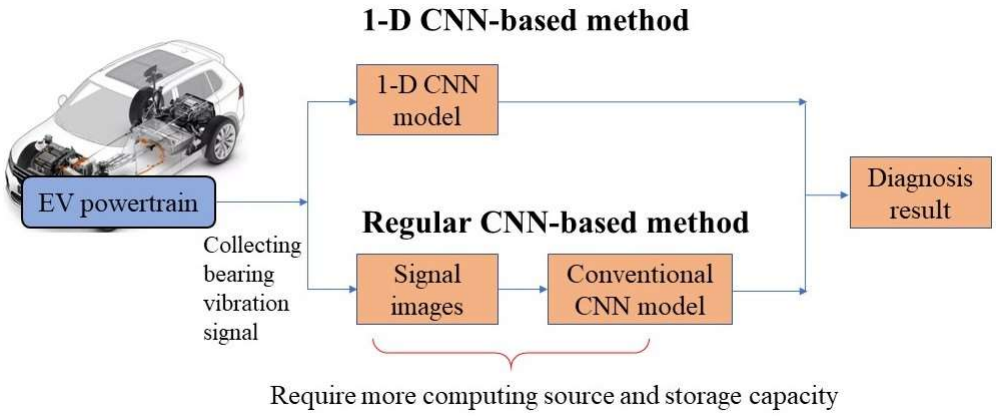


Fig. 2-8. Architecture comparison between regular CNN based method and 1-D CNN based method (1-D CNN can omit the signal-to-image step than regular CNN).

TABLE 2-5 Performance of CNN-based Methods under CWRU Bearing Data

Papers	Methods	Test target	Top-one accuracy
[53]	CNN+RF		99.08%
[54]	PSPP-CNN		97.79%
[71]	CNNEPDNN	CWRU dataset	98.1%
[99]	CNN+SVM		98.75
[96]	1-D CNN+Transfer learning		90.8%

Data are a vital part of AI-based diagnosis methods. Due to the difficulty of collecting the bearing fault data for a specific type of EV, the most suitable AI technology would be transfer learning. Transfer learning makes the model trained in the source domain still applicable in the target domain. And greatly saves the time to construct a bearing health recognition model. Fig. 2-9 demonstrates the principle of transfer learning.

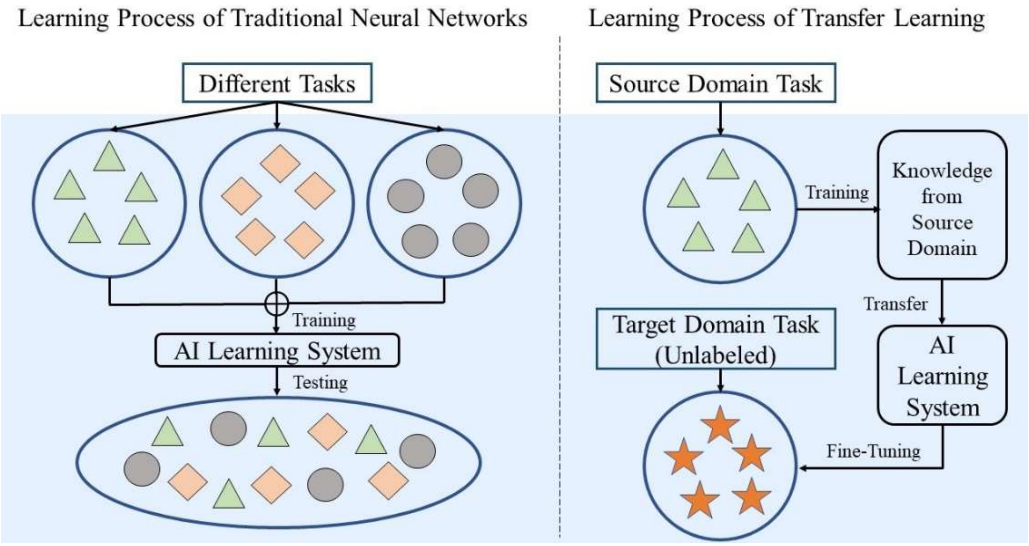


Fig. 2-9. The principle of transfer learning and the difference with traditional NNs.

For transferring fault diagnosis tasks where there is no available labeled data in the target domain, the method based on transfer learning is better than the classic method without transfer learning. Paper [96] proposed a deep convolutional transfer learning network (DCTLN) based on transfer learning technology, which can learn features directly from vibration signals.

AI-based methods should preferably use the same dataset for comparative analysis. Therefore, Table 2-6 analyses and compares five transfer learning methods suitable for EV bearing fault diagnosis on the CWRU bearing dataset, which includes deep domain confusion (DDC) [133], domain adversarial training of NNs (DANN) [134], transfer component analysis (TCA) [135] and deep domain adaptation neural-network-based fault diagnosis (DAFD) [136]. It can be concluded that transfer learning is a promising tool in the field of EV bearing diagnosis. In addition, the DCTCL method has relatively high accuracy and is suitable for intelligent fault diagnosis of unmarked data of EV bearings.

TABLE 2-6 Performance of Transfer Learning-Based Method Based on CWRU Dataset

Papers	Methods name	Feature learning	Top-one accuracy
[96]	DCTCL	Directly from vibration signal	90.8%
[133]	DDC	Directly from vibration signal	88.5%

[134]	DANN	Directly from vibration signal	88.5%
[135]	TCA	Manually extract features	42%
[136]	DAFD	Vibration signal spectrum	54.9%

Thermal images-based AI diagnosis is an advanced and novel attempt for motor bearing failure. For example, in paper [8], thermal images of bearing failure can be analysed directly by the proposed modified CNN model without any signal pre-processing. However, due to the particularity of the working environment of the EV motor and the high integration of EV components, this method is hardly implemented on the EV currently.

2.4 Condition Monitoring and Fault Diagnosis in Inverter

The rapid growth of the EV industry makes people pay more attention to the safety of low-voltage and high-power inverters. In fact, the health of parts such as the power transistors in the inverter is closely related to the reliability of the entire device. The commonly used inverter fault diagnosis techniques mainly include three types: model-based methods, expert systems and AI-based methods [137]. Therefore, the amount of literature on AI-based EV inverter methods is relatively small compared to other methods.

Currently, most of the power devices in EV inverters are insulated gate bipolar transistors (IGBTs). This is because the IGBT has the advantages of low state voltage drop and high input impedance. IGBT is taken as the main research object in this section. Fig. 2-10 shows three periods of IGBT failure: early failure period, random failure period and wear failure period with different failure rates.

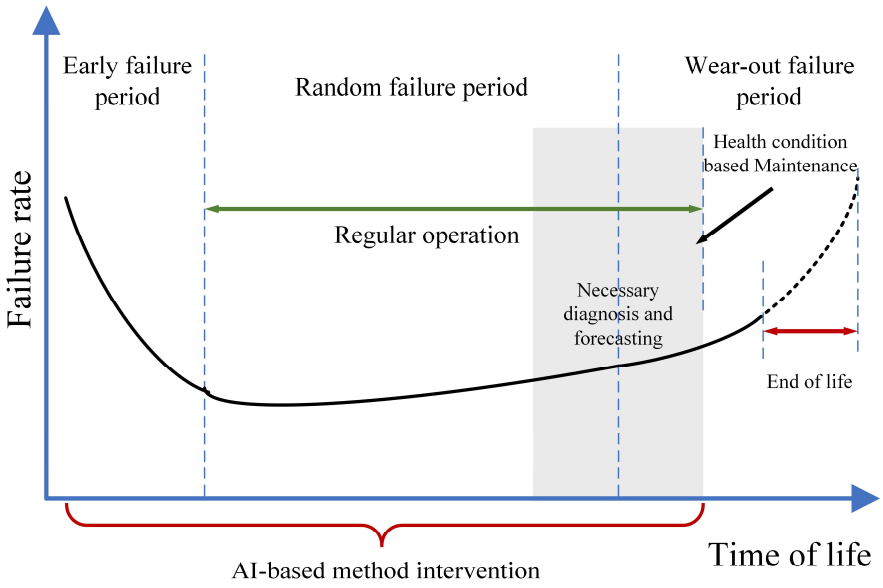


Fig. 2-10. Power devices/IGBTs failure rate in lifetime and AI intervention timing.

2.4.1 RUL prediction for IGBT

EVs use IGBTs for efficient power conversion. IGBTs are exposed to electrical, thermal, and mechanical stresses during operation, and these stresses can cause performance degradation, which in turn can lead to wire-bond lift-off and solder fatigue.

In EV applications, they may experience increased thermal cycling (temperature rise and fall), resulting in thermomechanical stress in the weak part of the IGBT package interconnection. Unexpected failures of IGBT will endanger the reliability of EV. By monitoring the trend of the fault precursor trajectory, this degradation can be identified at an early stage.

IGBT RUL estimation methods are roughly divided into physics-based analysis methods and AI data driving methods. Among them, the physics-based method can provide a universal fault propagation method for the IGBT under test. However, these general physical models cannot represent the degradation of all IGBTs. Furthermore, due to the cumulative random behavior (CRB) of thermal stress and mechanical stress, these methods have poor adaptability to dynamic environments.

The data-driven methods predicted by IGBT RUL are divided into conventional data-driven methods and AI-based data driving methods. Both two types of data-driven methods can avoid complex physics and random uncertainty. Paper [138] proposes an RUL estimation approach

based on the auxiliary particle filter (APF) to reduce the estimated variance by increasing the dimensionality of the sample and maintaining the diversity of the sample. Paper [139] also proposed a PF-based method for building a fault precursor by fusing the junction temperature (T_j) and collector-emitter on voltage ($V_{CE(on)}$). Paper [140] uses Gaussian process regression (GPR) for applying a Bayesian inference (BI) on RUL estimation of the IGBT device.

Bayesian-based estimation methods can accurately track linear trends. PF can track non-linear trends, but when the trajectory has both linear and non-linear trends, their performance will decrease, especially when subjected to harsh measurement noise. In addition, these algorithms are computationally expensive and require a large number of samples to perform complex prior probability estimation. Paper [141] develops an algorithm that can be embedded and supports real-time decision-making on board. The probability of the state of health in the degradation process is obtained by performing a random process on the estimation result. Table 2-7 compare and analyze the performance of the conventional data-driven methods and the AI-based method.

Fig. 2-11 presents the comparison of six feature values of the five measurable signals in EV inverter. In general, $V_{CE(on)}$ has a great comprehensive performance.

TABLE 2-7 Quantitative Analysis of Data-Driven Cases

Types	Method	Experiment target	Reference parameter	Accuracy (RMSE)
Conventional data-driven method	PF-based [139]	Simulation data from the IGBT module based on Hefner physical model from Saber software	Fusing $V_{CE(on)}$ and T_j	≤0.357
	APF-based [138]	Infineon Technologies $FS_{30}R_06W_1E_3$ IGBT data	$V_{CE(on)}$	0.178
AI-based data-driven method	TDNN-based [141]	Power cycling tests data from Nottingham university	$V_{CE(on)}$	≤1.15

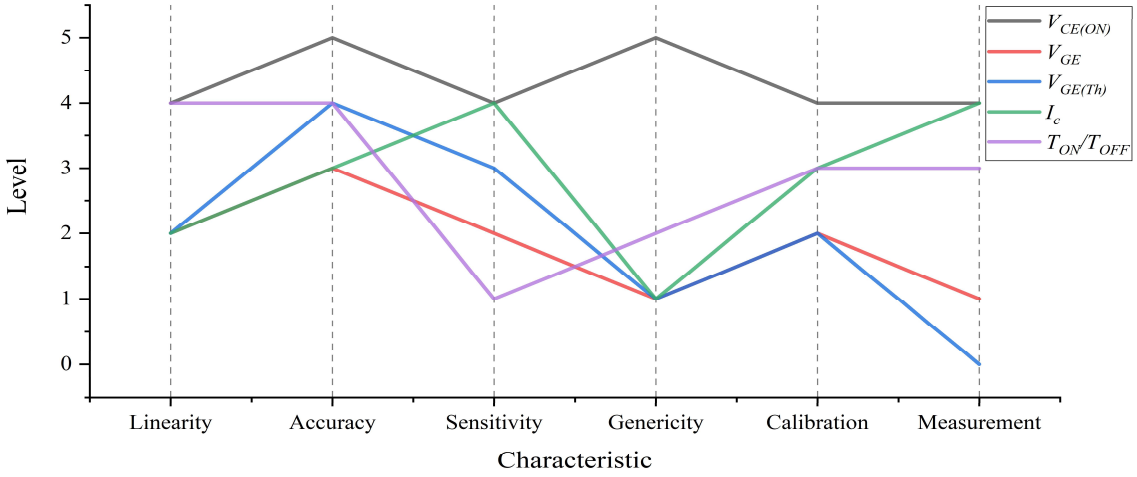


Fig. 2-11. Performance comparisons of the fault evidence signals of IGBT ($V_{CE(on)}$ is the collector-emitter on voltage, V_{GE} is the gate-emitter voltage, $V_{GE(Th)}$ is the gate-emitter threshold voltage, I_c is the collector current, T_{ON}/T_{OFF} donates the ratio of switch turn on time and turn off time).

2.4.2 IGBT open-circuit fault monitoring and detection

The open-circuit fault of IGBT is a relatively long timing problem, which will cause the whole system to gradually overheat and eventually collapse. This is also the reason why IGBT open-circuit fault is easier to detect and diagnose than short-circuit fault. When an IGBT is short-circuited, an abnormal short-circuit current usually occurs for only a few milliseconds, so it is difficult for the AI algorithm to detect a short-circuit fault. Even if the occurrence of a short-circuit fault is detected, it is difficult to take action to remedy before the system is damaged.

Fig. 2-12 illustrates the internal circuit principle of three cases of IGBT open circuit fault, which presents that the current flows wrongly when open-circuit of T_1 happens in three cases.

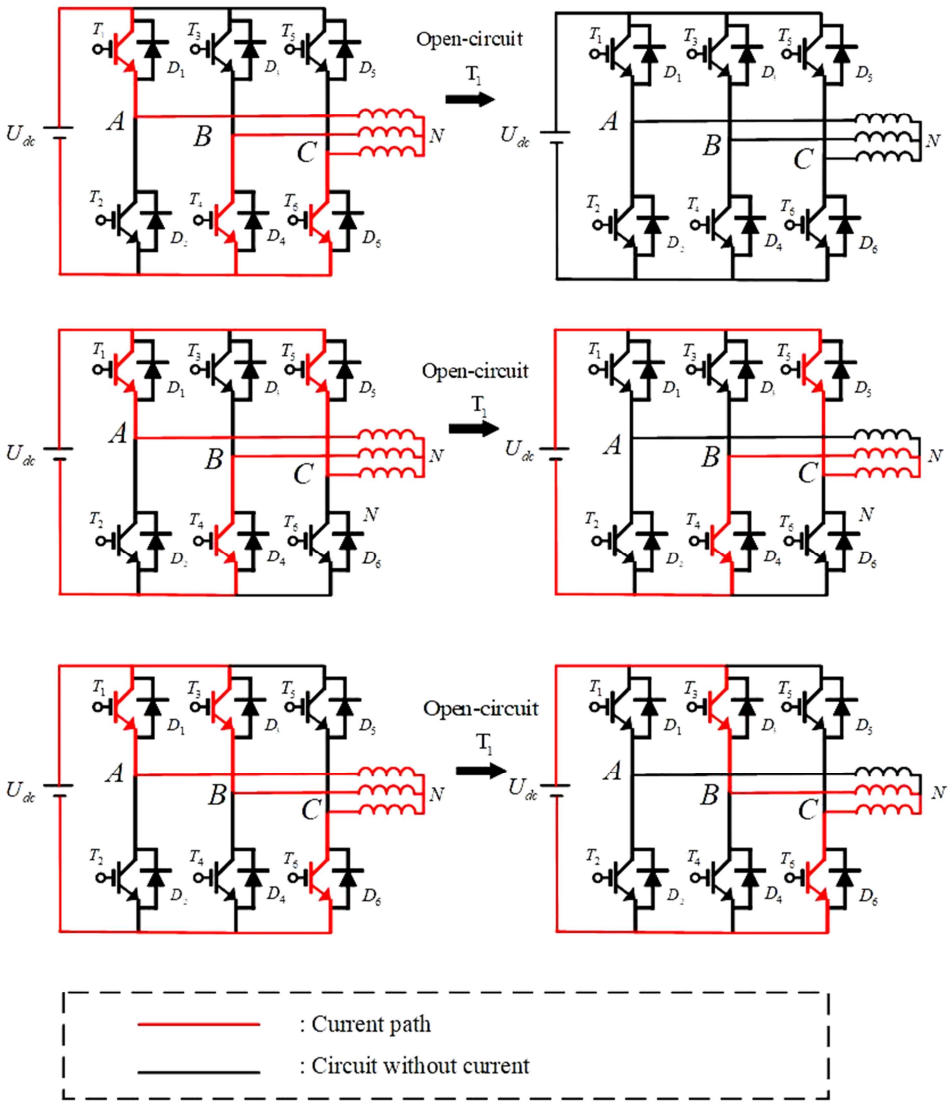


Fig. 2-12. Schematic diagram of IGBT open-circuit fault principle (assuming T_1 open-circuit).

At present, in the field of electrical inverter fault diagnosis, the mainstream diagnosis methods include methods based on mathematical models of fault mechanisms, methods based on the expert system, and data-driven AI-based methods. The traditional methods have been applied and have achieved considerable results in diagnosing open-circuit faults. However, some shortcomings of these methods still exist. Model-based diagnosis methods need to establish a high-precision mathematical model describing the fault evolution mechanism. This method can achieve better results in actual applications. However, it is very difficult to establish a high-precision system fault model with complex multi-variable. In addition, the established fault

mathematical models are usually designed to solve specific device and are difficult to transplant to solve similar or other devices. Furthermore, with the increasing complexity of electromechanical devices, fault diagnosis methods based on mathematical models are restricted in practical application. Therefore, with the rise of AI, people began to try intelligent diagnosis methods based on ML and NNs.

In the application of EV powertrain, SVM is a commonly used ML algorithm when diagnosing IGBT open circuit faults. For example, paper [142] uses SVM as a classifier to obtain fault types. Similarly, paper [143] compares algorithms such as MLP, SVM, and K-means and believes that SVM can obtain better classification results. The difference is that [142] uses fast Fourier transform (FFT) to extract fault features from inverter output voltage, while [143] uses wavelet transform (WT) to extract fault information from three-phase currents of DC/AC inverter. Table 2-8 is a good case to compare the performance of AI-based methods (K-means, self-organizing map (SOM), MLP, SVM) for EV powertrain voltage source inverter (VSI). The advantages and disadvantages of these ML technologies for EV inverter open-circuit fault detection could be summarized in Table 2-9.

TABLE 2-8 Quantitative Analysis of Four ML Technologies Used in The Case

Methods	Test target	Accuracy	Training time	Complexity
MLP [143]	EV powertrain VSI inverter	High	5s	High
SVM [143]		High	1s	Low
SOM [143]		-	5s	Medium
K-means [143]		-	2s	Medium

TABLE 2-9 The Advantages And Disadvantages of Four AI Technologies Used for IGBT Open-Circuit Fault Detection in EV Powertrains

Technologies	Advantages	Disadvantages
MLP [143]	Can solve all the problems with its complex and nonlinear relationships. No limit on the number of outputs. The high accuracy performance.	Needs a very heavy and complex training. Due to split the data with a line, it has lowest margin of safety and highest risk.

SVM [143]	Low training process. Low programming complexity. Low Implementation and testing time. Due to the data separation zone, has the most safety marginalized and least risky.	To solve very complex problems. Doesn't have a particular way to set the variables.
SOM [143]	The training algorithm is simple and easy and does not need to determine the complex output.	Doesn't have accuracy in complex and non-linear data.
K-means [143]	The training algorithm is simple and easy and does not need to determine the complex output.	Doesn't have accuracy in complex and non-linear data.

Furthermore, for detecting open-circuit faults of neutral point clamped (NPC) inverters, paper [144] uses modern compressed sensing theory to process the voltage signal into the proposed least-squares SVM with gradient information (G-LS-SVM) for fault classification. The author of paper [145] innovatively uses the upper, middle and down bridge voltages signals as the classification evidence for multi-layer SVM. Differently, [146] apply NN model is used as the classifier.

The frequency of use of NN based deep learning models in related research is gradually increasing [147] [148] [149] [150]. Its most representative model is the ANN used in paper [151]. Other variants of NN have also received much attention. For example, the feed-forward back propagation neural network (FFBPNN) in paper [152] is used to diagnose IGBT open-circuit faults under variable load conditions at different frequencies and has good performance.

Diagnosis methods based on SVM, MLP, ANN and other algorithms still require complicated and time-consuming manual extraction and marking of fault features, which could be avoided by deep learning algorithms with the ability to automatically extract features such as CNN. In paper [153], the author proposes to apply global average pooling (GAP) for modifying CNN model to diagnose the open-circuit fault of the DC-DC inverter, which extracts 12 types of fault features from signal transformed images. We use the 12 types of IGBT open-circuit faults dataset to test the performance of various ML and DL algorithms, including: SVM, K-NN, BPNN, DNN, CNN, CNN-GAP (Table 2-10). More related papers can refer Table 2-11.

TABLE 2-10 Quantitative Analysis of Six AI Technologies Used in Cases

Methods	Test target	Average accuracy	Training time	Decision time
SVM [153]	12 types of IGBT open-circuit faults data from the DC-DC inverter simulation model	96.13%	22.79s	4.14s
K-NN [153]		95.15%	0.04s	0.07s
BPNN [153]		80.32%	78.19s	0.05s
DNN [153]		97.01%	396.5s	0.27s
CNN [153]		99.27%	901.26s	2.95s
CNN-GAP [153]		99.95%	824.89s	2.49s

TABLE 2-11 Summary of Contribution and AI Algorithms of Each Paper with Signals and Feature Methods

Paper	Signal Input	Feature Method	AI Algorithm	Contributions
[144]	Voltage		G-LS-SVM	• To use the compressed sensing theory to sparsely express the voltage fault signal
[152]	Current	DWT	FFBPNN	• To be a new method of three phase VSI fault diagnosis • Threshold needless under variable load conditions
[146]	Voltage	JADE–ICA	NN	• To use JADE-ICA algorithm to overcome the effects of nonlinearity and time difference
[143]	Current	WT	MLP, SVM, SOM, K-means	• To analyze 5 dynamic modes and 5 static modes
[142]	Voltage	FFT, RPCA	SVM	• To select the output voltages of the inverter as the fault characteristic signal
[153]	1-D time-series	CNN	CNN-GAP model	• To learn features directly from the original one-dimensional time-series data • To be fast and more accurate

	voltage and current			
[147]	Three-phase current	ICA	NN	• To use ICA for three-phase current signal processing
[145]	Upper, middle and down bridge voltages	DFT	multi-layer SVM	• To Use a novel multi-layer SVM to detect the open circuit fault of the inverter
[148]	Upper, middle and down bridge voltages	PCA	multi-layer ANN	• To propose a novel multilayer neural network to diagnose all possible open-circuit faults
[149]	Current	CNN	CNN	• To learn fault features independently through a convolutional network
[150]	Three-phase currents	Park's transform	BPNN	• To apply Park's transform to obtain the three-phase current base wave amplitude as the feature variable for fault detection
[151]	Three-stator currents	CEEMD	HHT-ANN	• To use HHT to detect the harmonic characterizing the fault based on CEEMD of the three-stator currents

2.5 Summary

This chapter reviews four aspects about AI-supported EV powertrain safety improvement topics. With the rapid development of AI, support methods that are developed based on this technology can further enhance the electric powertrain condition monitoring and fault diagnosis. This study presents a comprehensive review of this, and it can be learned from this paper that AI-based data-driven method can avoid the need for accurate physical models of the system, and AI can also help build accurate system models. A large number of experiments and case studies have proved that both the AI data-driven method and the AI-supported model-based method have high accuracy and performance, and some advanced AI technologies can achieve functions and effects that cannot be achieved by conventional methods. Therefore, the application of AI in PE has important practical significance. The main contributions include:

1) This review focuses on feature engineering and three main components of EV powertrain and summarizes the recent research on EV. The motivation, advantages and challenges of using AI methods in PE problems are explained.

2) This review carries out the quantitative comparative analysis on condition monitoring and fault diagnosis methods in specific fields. In each application field, specific practical case studies on EV and quantitative comparative analysis with traditional methods are provided. Based on the characteristics of EV, this review analyses the feasibility, advantages and disadvantages of each method in the practical application of EV.

3) Usually, for bearing fault diagnosis tasks, the vibration signal is a commonly used signal because the vibration sensors are widely installed in powertrains. Similarly in gearbox, the vibration signal is a best choice for fault diagnosis tasks. For health monitoring tasks in motor and inverter, stator current signal is a better choice because this signal is easiest to collected and obtained.

Chapter 3 Inferable Deep Distilled Attention Network for Diagnosing Multiple Motor Bearing Faults

This chapter aims to propose an AI-based method to solve the health monitoring problem of mechanical bearings in electric powertrain. Bearing, as a vital component in electric powertrains, is increasingly used globally such as in electric vehicle (EV). Their damages and faults may bring huge cost loss to the industry and even threaten personal safety. This chapter proposes an inferable deep distilled attention network (IDDAN) method which is a self-attention mechanism and transfer learning-based method to diagnose and classify multiple bearing faults in various motor drive systems efficiently and accurately. Compared with convolutional networks, the self-attention-based network can better extract the global feature information and easier to benefit from large amounts of pre-training data. Its significance is to accurately classify various faults of the target machine when the labeled data of the target machine is not enough to directly train the diagnosis model. Firstly, this chapter attempt to apply the self-attention-based network to build an advanced fault diagnosis model. Secondly, this paper optimizes the structure of networks through knowledge distillation (KD) technique to require a lighter and fast model. Thirdly, this paper proposes a new data augmentation strategy for 1-D vibration signals to provide large-scale pre-training samples for IDDAN. Experiments show that the self-attention mechanism-based model is more likely to benefit from large-scale data. After testing, compared with many methods and other exist similar methods, the proposed method achieves higher classification accuracy and better performance.

3.1 Introduction

The electric motor drive system has been widely used in industry and human life. The reliability and safety of its components bear the responsibility of human life and industrial cost. The reliability issues of the electric powertrain may appear on any components. The bearing plays a critical and necessary role in motor drive system. According to incomplete statistics, 40-70% motor and electric powertrain faults are caused by various degrees of rolling bearing damage [154]. Such faults are leading to the higher costs in industrial applications and its maintenance.

Therefore, real-time condition monitoring and bearing fault diagnosis (BFD) in all motor drive system is gradually becoming more important and higher-priority work.

The safety and stability of bearings have attracted increasing attention in both academic and engineering. Scholars and engineers have employed many traditional methods to detect which type of bearing fault in some motor drive systems. The conventional method of bearing fault diagnosis mainly relies on advanced signal processing technology to extract effective features for analysis. Paper [155] proposes an adaptive morphological filter (AMF) to analyze the vibration and acoustical signals of the bearing to determine the fault type. Paper [156] proposes the sparse elitist group lasso denoising (SEGLD) algorithm to online diagnose bearing faults in industry. This is because part of the information contained in the motor stator current signal or the bearing vibration signal collected by the measurement is not related to the bearing fault, such as the supply fundamental and its harmonics, noise, etc. The core contributions of the paper [157] and [158] are both to solve this problem. On the other hand, judging from the current trend of big data, artificial intelligence (AI) can bring about better convenience and more advantageous new ideas for bearing fault diagnosis.

The AI-based BFD methods have characteristics of model independence, does not require professional mechanical knowledge, and has excellent performance. Paper [159] uses graph-mapped spectrum (GMS) to represent fault information in bearing vibration signals and applies K-nearest neighbour (K-NN) classifier to identify fault types. Paper [160] combines information fusion (IF) technology with convolutional neural network (CNN) to diagnose bearing faults. Paper [161] is also based on supervised learning of AI. The method proposed in this paper is based on CNN to identify damage to rotor bearings from infrared images. Supervised learning is a classic method in AI-based methods. However, for some machines, it is difficult to obtain sufficient labeled data for supervised training in real situations. People need new ways to solve this problem.

To solve the above problem of lack of labeled data, transfer learning technology is used in bearing multiple fault diagnosis. It obtains a pre-trained model with rich domain knowledge from a machine, and then adapts it to the target machine with a small amount of data. Due to the difficulty in obtaining bearing fault data, there are many similar methods based on transfer learning recently. Paper [162] proposes an intelligent bearing fault diagnosis system combining AlexNet and transfer learning technology. The deep convolutional transfer learning network (DCTLN) proposed in paper [163] can make the model still effective in the target domain.

These methods are almost all proposed based on CNN [164], [165]. Paper [166] uses Bayesian network as a fault diagnosis model for bearing fault detection and proposes varying coefficient transfer learning (VCTL) to obtain knowledge and correlation from the resource domain. Paper [167] uses the bearing vibration data obtained by computer simulation to pre-train the model, and then achieve the general effect of the model through transfer learning. However, traditional network models such as SVM, and CNN have the ability of automatic feature learning, but still face many challenges. For example, CNN cannot focus on learning to important discriminate features of faults and ignore useless features. Furthermore, global information cannot be extracted due to the limitation of convolution kernel size. The new type of neural network based on self-attention mechanism will improve the current situation.

Based on the transfer learning and a new self-attention mechanism, an inferable deep distilled attention network (IDDAN) is proposed to diagnose bearing multiple faults. The network can accumulate the advantage of optimized feature mapping across the network through the intervention of the self-attention mechanism, which uses global information to adaptively enhance more discriminative features and suppress irrelevant features. With the proposed data augmentation technique, it is possible to explore how much each method benefits from large amount of data samples. In addition, the knowledge distillation (KD) technique [168] makes the trained network lighter and increases the inference speed. The contributions of this chapter are summarized as follows:

1) It originally proposed a new data augmentation method of 1-D vibration signals and applied it to the bearing fault diagnosis framework. Models can benefit from data augmentation to become more generalizable. Experiments show that this strategy cooperates well with the self-attention module to obtain accurate diagnosis results.

2) Self-attention mechanism based neural network is introduced in transfer learning-based intelligence diagnosis method. The types of bearing faults are complex and diverse, which poses a higher challenge to the feature recognition ability of the diagnosis model. Self-attention mechanism based can enhance more discriminative features and suppress irrelevant features during training.

3) An advanced bearing multiple fault diagnosis method, IDDAN, is proposed based on data augmentation technique, self-attention networks, transfer learning and KD technique and works under various working conditions. The model size of the method is lighter than commonly used models so as to increase computing efficiency.

In this chapter, Section 3.2 introduces related theories for further understanding the architecture introduced in Section 3.3. Section 3.3 describes the proposed method in detail. Section 3.4 presents the case study and experiment result. Finally, Section 3.5 provides a conclusion of this chapter.

3.2 Self-attention Mechanism and Related Theories

3.2.1 Self-Attention Mechanism

Self-attention is a special form of attention mechanism. The output of attention mechanism (see Fig. 3-1) could be presented as (3-1) [169].

$$Attention_{Output} = Attention(Q, K, V) \tag{3-1}$$

where Q, K and V respectively stand for the *Query* matrix, *Key* matrix, and *Value* matrix. Let input vector X be the input and get $Q = W_Q X$, $K = W_K X$, $V = W_V X$ (W_Q, W_K and W_V are parameter matrixes). The dimension of X, Q, K and V can be set for each task.

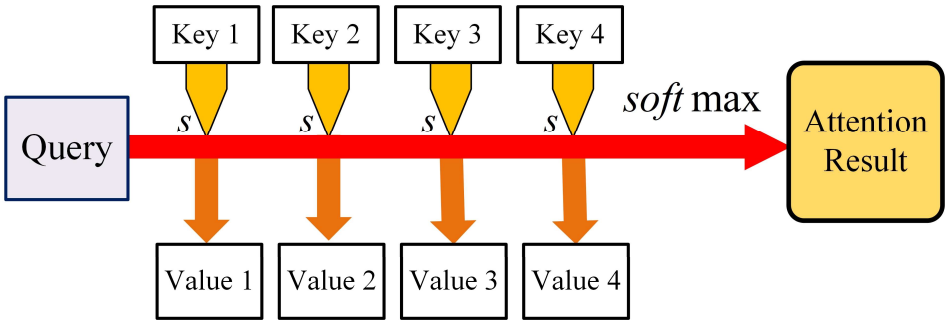


Fig. 3-1. Diagram of computing process of attention mechanism.

The scaled dot-product is used in the calculation process as shown in (3-2):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3-2}$$

Among (3-2), d_k is the dimension of K, the *softmax* function is a function that turns a vector of K real values into a vector of K real values that sum to 1, which formula could be presented as (3-3). The input values can be positive, negative, zero, or greater than one, but the *softmax* transforms them into values between 0 and 1, so that they can be interpreted as probabilities. It

is usual to append a *softmax* function as the final layer of the neural network to convert the scores to a normalized probability distribution.

$$\sigma(\vec{Z})_i = \frac{e^{Z_i}}{\sum_{j=1}^{K} e^{Z_j}} \tag{3-3}$$

3.2.2 Multi-Head Attention Mechanism

Multi-head attention is to project the h group Q, K, V through different linear transforms, and connect the final result. In self-attention mechanism, each group of Q, K, V is the same.

$$Multihead(Q,K,V) = Concat(head_1, head_2, \dots, head_h)W^O \tag{3-4}$$

$$Head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{3-5}$$

In both (3-4) and (3-5), W^O represents for the parameter matrix of linear transforms. In detail, $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$ and $W_i^O \in \mathbb{R}^{hd_v \times d_{model}}$ ($d_k = d_v = d_{model}/h$).

3.2.3 Classic KD Theory

Neural network models can solve a variety of complex problems, but these models are usually huge and have a large number of parameters, making it difficult or impossible to deploy to edge devices. KD is a new method of compressing neural models. The obtained new smaller network trained through KD technology can achieve the same or similar effect as the original network [170].

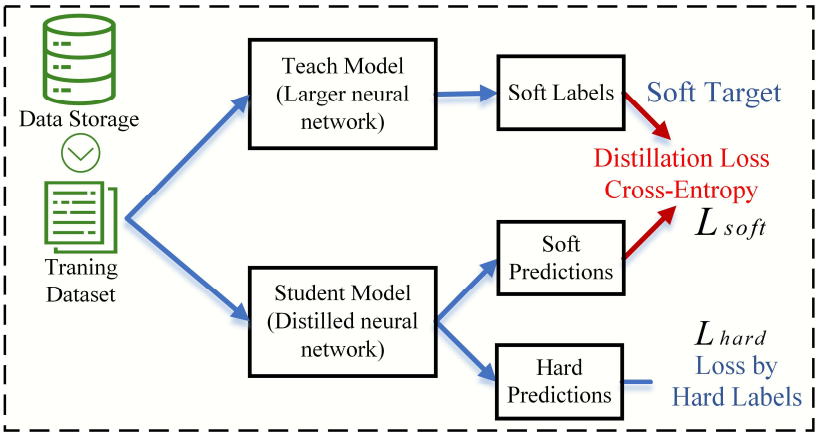


Fig. 3-2. Schematic diagram of complete KD process.

Fig. 3-2 shows that the essence of KD is the process of pre-trained larger teacher network teaching smaller student network. The training dataset is applied to both the teacher network and the student network at the same time. The soft label is the output of the teacher network in each layer. The cross-entropy between it and the soft prediction value output by the student model is L_{soft}. L_{har} is defined as the cross-entropy of the hard prediction value output by the student model and data label. Therefore, the objective function of the KD process is composed of weighted L_{soft} and weighted L_{hard}:

$$L = \alpha L_{soft} + \beta L_{hard} \tag{3-6}$$

where α and β are weights of L_{soft} and L_{hard}, respectively. The propose of the KD strategy is to minimize L.

3.2.4 Transfer Learning Problem

The essence of the model pre-training is transfer learning [171]. To clearly explain the proposed architecture in Section 3.3, here it is necessary to introduce two types of domains in transfer learning: the source domain D_s and the target domain D_t. Both domains are composed of the feature space X and the probability distribution $P(X)$ of the data. The category spaces of the learning objectives of transfer learning in D_s and D_t are represented by Y_s and Y_t, respectively. When $D_s \neq D_t$, the data distribution before and after transfer is also different. Transfer learning can improve the performance of the target task learning function f_t when $D_s \neq D_t$. The following Fig. 3-3 shows the principle of transfer learning to process unlabelled data task and what is different process between transfer learning and traditional neural networks.

This paper aims to looking for a multi-fault diagnosis architecture that can be quickly deployed on any machine and monitor its health conditions. From the perspective of transfer learning, the data used for pre-training $X = \{x_1, x_2 \cdots x_n\}$ and the corresponding label space $Y_s = \{y_{x_1}, y_{x_2} \cdots y_{x_n}\}$ compose source domain $D_s = \{X, Y_s\}$. The unlabelled data of the target machine is $X_t = \{x_{t1}, x_{t2} \cdots x_{tn}\}$, where x_{tn} is the data samples of the target task. The more adequate samples in D_s, the larger the category space, the stronger generalization ability, and the better performance of the model after pre-training transfer.

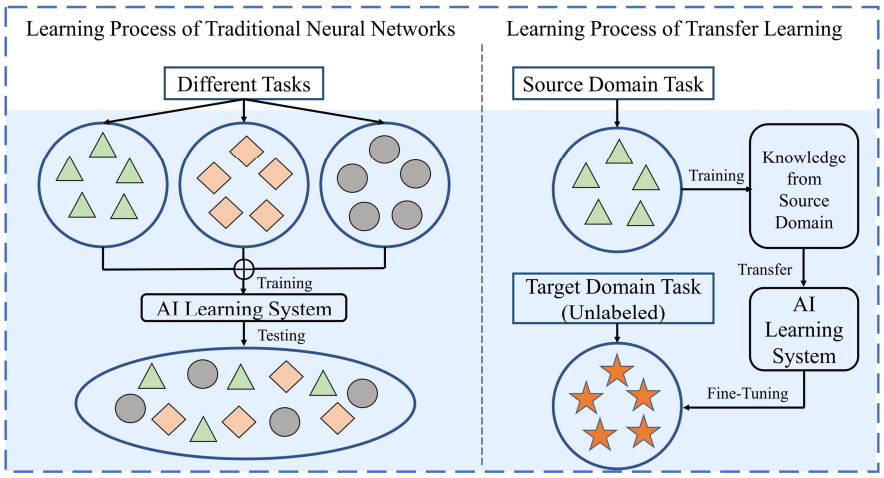


Fig. 3-3. The comparison and difference of the learning process between traditional neural networks and transfer learning.

3.3 Intelligent Diagnosis Framework Based on Inferable Deep Distilled Attention Network (IDDAN)

The proposed diagnosis method consists of four main modules (as shown in Fig. 3-4): data augmentation, backbone network, distillation strategy and transfer inference. Transfer inference helps IDDAN become more adaptable to the target domain through fine-tuning by the very small amount of data after the pre-training stage. Particularly, the first and second parts of Fig. 3-4 show how the bearing vibration signal is collected and pre-processed in the context of the current application.

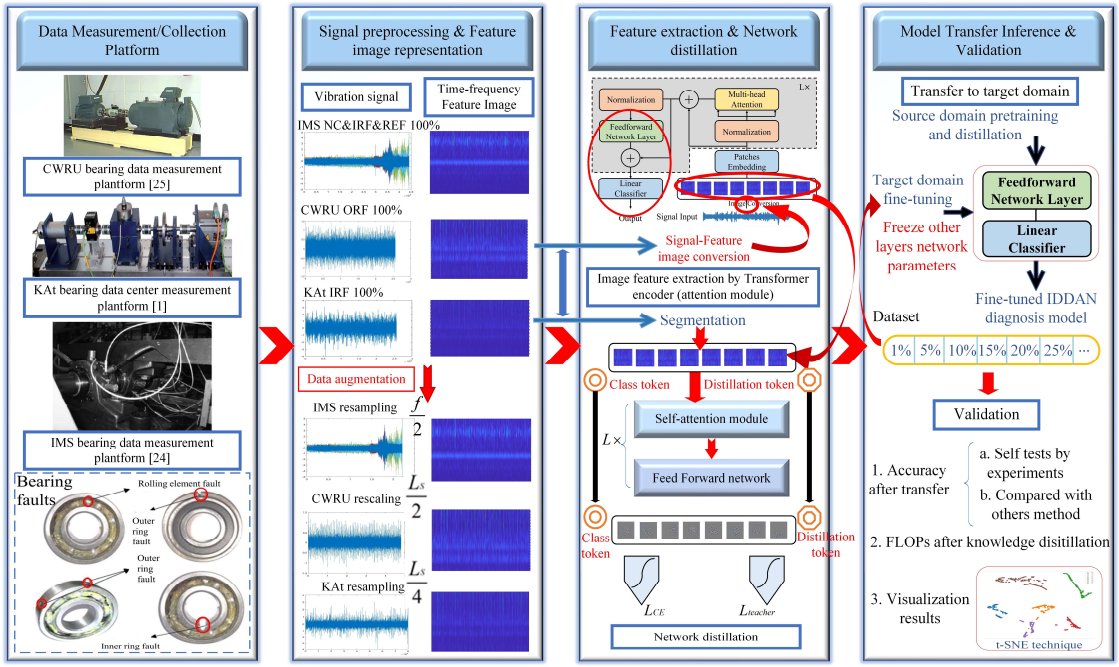


Fig. 3-4. The architecture demonstration of the proposed method.

3.3.1 Data Augmentation Strategy

The data augmentation method of IDDAN's pre-training data in the source domain adopts multi-scale and multi-timescale signal conversion which is a special data augmentation method proposed for the collected 1-D bearing vibration signals.

Multiscale signal conversion is similar to the idea of cropping and zoom in computer vision, which can improve the generalization ability of pre-trained IDDAN. In other words, the domain adaptation ability of the IDDAN can be improved when transferring knowledge from the source domain to the target domain. Defining the length of each vibration signal as L_s and the sampling rate of as f_s, the principle of data enhancement can be expressed as Fig. 3-5.

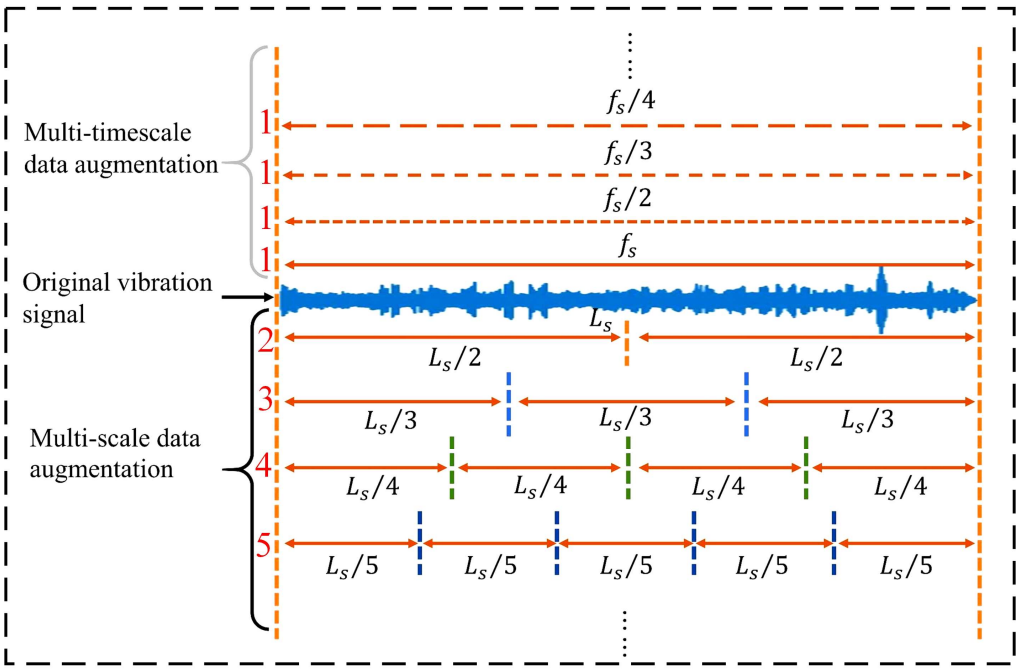


Fig. 3-5. The data augmentation strategy diagram of the data used for IDDAN pre-training.

It can be obtained from Fig. 3-5 that the strategy of data augmentation could be presented by equation (3-7):

$$\begin{cases} V_{segmented} = \frac{L_s}{k} V_{origianl}, \ k \in \mathbb{N}^* \\ V_{new}(t) = \sum_{n=-\infty}^{\infty} V_s(nT)\delta(t-nT) \end{cases} \tag{3-7}$$

where k is the number of segments after each signal cutting, $V_{origianl}$ is the original vibration signal, $V_{segmented}$ is the segmented vibration signal, $V_{new}(t)$ is down sampled vibration signal, $V_s(t)$ is the original collected vibration signal, δ represents the impulse function, $T = 1/f_s$. In this data augmentation step, collected 1-D vibration signals are mainly processed in two ways:

1) Each vibration signal used for IDDAN pre-training step is segmented multiple parts, and the size of each segment is $L_s/2$, $L_s/3$, $L_s/4$, $L_s/5$, etc.

2) Each vibration signal used for IDDAN pre-training step is resampled by sampling rate $f_s/2$, $f_s/3$, $f_s/4$, etc.

Through this data augmentation method, the amount of data can be increased efficiently, providing sufficient data for IDDAN pre-training to improve the effect of knowledge transfer.

3.3.2 Backbone Network

The bearing health status recognition is realized through self-attention based network module as the main body. The backbone of IDDAN is a modified transformer network, which includes one transformer encoder and a linear classifier. The transformer encoder is constructed by one multi-head attention module, one feed-forward module and two normalization layers, which can automatically learn global features with the help of the self-attention mechanism [172]. A linear classifier is used to identify and distinguish bearing health conditions. As mentioned in Section 3.1, the self-attention mechanism can be processed parallelly with global capabilities, long-distance information will not be weakened [173]. Self-attention could be considered as a CNN with a learnable receptive field. In other words, self-attention can learn receptive field automatically, but the receptive field of CNN needs manual adjustment and optimization of parameters.

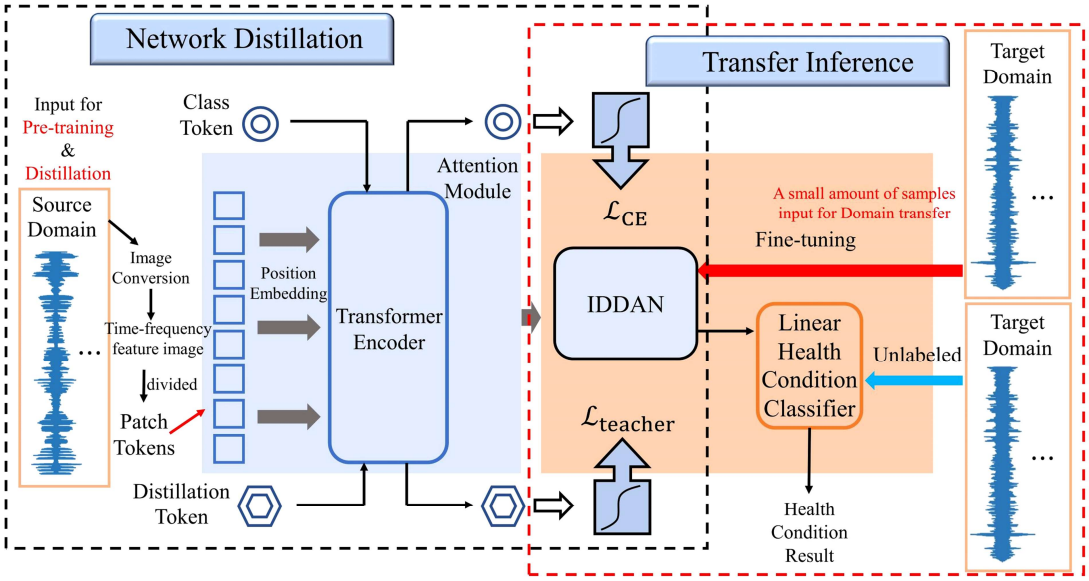


Fig. 3-6. The structure presentation of IDDAN (MSA block and FNN block are both in the Transformer encoder).

As shown in Fig. 3-6, the self-attention module includes a position embedding module, a transformer encoder (feature extraction in Fig.3-4), and a linear layer. The linear layer can be regarded as a condition classifier to classify the global features extracted by the self-attention mechanism in the transformer encoder.

The vibration signal of a certain length is first converted into the time-frequency map representation. For handling these 2D signal representations, the transformer encoder reshapes the image $\mathbf{x} \in \mathbb{R}^{H\times W\times C}$ into a sequence of flattened 2D patches $\mathbf{x}_p \in \mathbb{R}^{N\times(P^2\cdot C)}$, where (H, W) is the resolution of the signal representation, C is the number of channels, (P, P) is the resolution of each image patch. The effective input sequence length could be calculated through:

$$N = HW/P^2 \tag{3-8}$$

where N stands for the resulting number of patches. These patches representing machine health information have positions in the original time-series signal. Position embedding can achieve the effect of abstracting data in time series and represent relative or absolute position information in the input sequence. The implementation process of position embedding in this chapter is as follows:

$$\mathbf{z}_0 = \left[\mathbf{x}_{\text{class}};\ \mathbf{x}_p^1\mathbf{E};\ \mathbf{x}_p^2\mathbf{E};\ \cdots;\ \mathbf{x}_p^N\mathbf{E}\right] + \mathbf{E}_{pos} \tag{3-9}$$

where $\mathbf{E} \in \mathbb{R}^{(P^2\cdot C)\times D}$, $\mathbf{E}_{pos} \in \mathbb{R}^{(N+1)\times D}$, $\mathbf{E}$ is the patch embedding projection and is D the constant latent vector size through all of transformer layers. The patches with marked positional information need to perform a normalization operation before entering the multi-head self-attention (MSA) mechanism. This chapter adopts layer normalization (LN) [174]. The state at the output of the Transformer encoder $\mathbf{z}_L^0$ serves as the image representation $\mathbf{y} = \text{LN}(\mathbf{z}_L^0)$. Both during pre-training and fine-tuning, a classification head is attached to $\mathbf{z}_L^0$.

$$\mathbf{z'}_\ell = \text{MSA}\big(\text{LN}(\mathbf{z}_{\ell-1})\big) + \mathbf{z}_{\ell-1} \tag{3-10}$$

where $\ell = 1, \cdots, L$. MSA is an extension of standard self-attention (SA), which runs SA operations k times (The calculation principle has shown in Section 3.2.1).

$$\text{MSA}(\mathbf{z}) = [\text{SA}_1(z);\ \text{SA}_2(z);\ \cdots;\ \text{SA}_k(z)]\mathbf{U}_{msa} \tag{3-11}$$

For calculating MSA in this chapter, we set $\text{SA}(\mathbf{z}) = A\mathbf{v}$ and $\mathbf{U}_{msa} \in \mathbb{R}^{k\cdot D_h\times D}$,where $A = \text{softmax}\big(\mathbf{q}\mathbf{k}^\intercal/\sqrt{D_h}\big)$ ($A \in \mathbb{R}^{N\times N}$, $D_h = D/k$ and $[\mathbf{q}, \mathbf{k}, \mathbf{v}] = \mathbf{z}\mathbf{U}_{sa}$, $\mathbf{U}_{sa} \in \mathbb{R}^{D\times 3D_h}$). LN is employed before both MSA block and FFN block, and residual connections after every block. Our FFN block contains two layers: one hidden layer at pre-training time and by a single linear layer at fine-tuning time:

$$FFN(x) = \max(0, xW_1 + b_1)\,W_2 + b_2 \tag{3-12}$$

where two different parameters W_1 and W_2 from layer to layer are used, and a gaussian error linear unit (GELU) activation function is applied between layers [175]:

$$GELU(x) = x \cdot \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \approx 0.5x\left(1 + \tanh\left[\sqrt{2/\pi}\,(x + 0.044715x^3)\right]\right) \quad (3\text{-}13)$$

Both the class and the distillation embeddings of IDDAN are associated with linear condition classifier. The final health condition prediction result is determined by the addition of the *softmax* [176]outputs of the two routes:

$$Class_{predict} = Linear[softmax(O_{ce}) + softmax(O_{de})] \quad (3\text{-}14)$$

where O_{ce} and O_{de} are the output of the two classifiers respectively.

3.3.3 Backbone Network

Backbone network's distillation process is carried out simultaneously with pre-training. In the output of the *softmax* layer, in addition to positive examples, negative labels also carry a lot of information. This training method of KD makes each sample bring more information to student model than the traditional training method.

Firstly, in order to improve the recognition accuracy of health conditions, the teacher model is chosen as VGG-16. It is a strong feature extractor and classifier. In this distillation strategy, the hard decisions of the teacher model $y_{\mathrm{t}} = \mathrm{argmax}_c Z_{\mathrm{t}}(c)$ (Z_{t} is the logits of the teacher model) are true labels. The applied hard-label distillation objective should be defined as:

$$\mathcal{L}_{global}^{HardDistill} = \frac{1}{2}\mathcal{L}_{\mathrm{CE}}(\psi(Z_{\mathrm{s}}), y) + \frac{1}{2}\mathcal{L}_{\mathrm{CE}}(\psi(Z_{\mathrm{s}}), y_{\mathrm{t}}) \quad (3\text{-}15)$$

where $\mathcal{L}_{\mathrm{CE}}$ represents the cross-entropy, and Z_{s} is the logits of the student model (IDDAN). For the image represents of machine signal, it is possible that hard labels will change according to the DA (Section 3.1 in this paper). In this hard-label distillation loss function, the teacher prediction y_{t} has the same role with the true label y.

Fig. 3-6 demonstrates that the distillation token, the class token, and patch tokens interact through the self-attention layers in the transformer encoder. The output of the distillation token

is the hard-label predicted by the teacher model. In this method, the class token $\mathbf{w}_{class}$ and the distillation token $\mathbf{w}_{distill}$ are trained by back-propagation algorithm:

$$\mathbf{w}(m+1) = \mathbf{w}(m) - \eta \frac{\partial J(\mathbf{w}(m))}{\partial \mathbf{w}(m)} \tag{3-16}$$

$J(\mathbf{w})$ represents the training error at any instance, m denotes the number of iterations, and $\eta > 0$ is the pre-set learning rate before training. This paper employs the gradient-based parameter optimizer AdamW [177] (a modified optimizer of the original Adam), replacing L2 regularization of Adam with weight decay.

3.3.4 Transfer Inference Objectives

The transfer inference stage of IDDAN includes fine-tuning and condition classification. Fine-tuning is a pivotal step to infer the pre-trained model to the target task through the transfer learning method.

The backbone network can be divided into feature extractors and classifiers. The feature extractor extracts the low-level features of the image. In the pre-training stage, the proposed vibration signal data augmentation strategy can obtain large-scale pre-training samples of the source domain, and the pre-model trained with large-scale data has a higher generalization ability to extract the underlying features. Therefore, during the transfer process, the bottom layer weights are frozen, and the high layer weights are opened. In this paper, the FFN layers and linear classifier layer in the pre-trained IDDAN will be updated with parameters in fine-tuning (has been demonstrated in Fig. 3-4). This is because the previous self-attention layers and layer normalization layer are used to obtain a general representation from the image, and the latter FFN layer and linear classifier are more relevant to downstream special fault diagnosis tasks.

3.4 Experimental Results

To verify the approach proposed in section 3.3, this chapter uses three professionally measured bearing datasets. The below three datasets could be consider to collected from EV motors.

3.4.1 Experiment Data and Description

1) Dataset A: KAt-DataCenter bearing dataset [154] contains a variety of faults to perform fault diagnosis experiments. This dataset focuses on not only artificial bearing damages but also real damages. It could prove better than the proposed approach in this paper is competent for different kinds of bearing fault diagnosis. The tested motor is a 425W permanent magnet synchronous motor (PMSM) which has the nominal torque $T = 1.35\,Nm$, the nominal speed $n = 3000\,rpm$, the nominal current $I = 2.3\,A$ and the number of pole pair $p = 4$.
2) Dataset B: The IMS bearing dataset is measured by provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati [177]. Recorded vibration signals include normal condition, rolling element fault, inner race fault, outer race fault. Each data describes a test-to-failure experiment and consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals. Each file consists of 20,480 points with the sampling rate set at 20 kHz and collected by NI DAQ Card 6062E.
3) Dataset C: Experiments of the CWRU bearing dataset were conducted using a 2 hp Reliance electric motor, and acceleration data were measured at locations near to and remote from the motor bearings [178]. Motor bearings were seeded with faults using electro-discharge machining. Faults ranging from 0.007 inches in diameter to 0.040 inches in diameter were introduced separately at the rolling element fault, inner race fault, and outer race fault. Vibration data was collected at 12,000 samples per second, and data was also collected at 48,000 samples per second for drive end bearing faults. Speed and horsepower data were collected using the torque transducer.

These three bearing datasets include vibration signals while they are obtained from different machines and different operation conditions. Their detailed information is displayed in Table 3-1.

TABLE 3-1 Detailed Information of Various Experiment Bearing Datasets

Dataset names	Bearings	Conditions	Speed (rpm)	Load conditions
KAt	PMSM bearing	Normal	1500	0.1Nm, 0.7Nm
		Inner ring fault	1500	0.1Nm, 0.7Nm
		Outer ring fault	1500	0.1Nm, 0.7Nm
IMS	Rexnord ZA-2115 bearing	Normal	2000	6000 lbs
		Inner ring fault	2000	6000 lbs

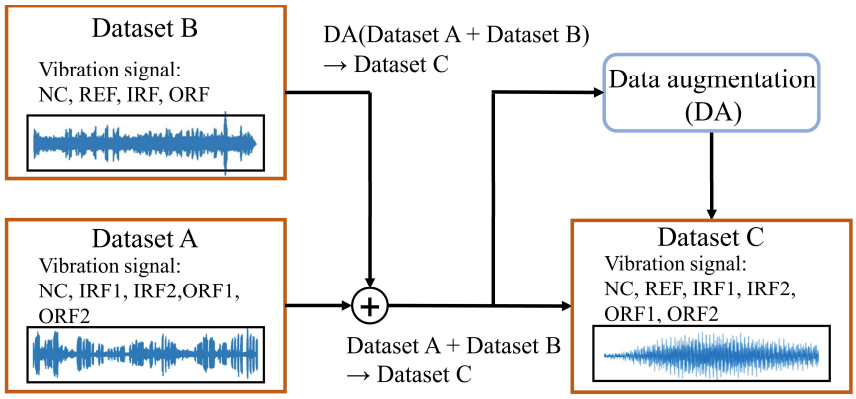
		Rolling element fault	2000	6000 lbs
		Outer ring fault	2000	6000 lbs
CWRU	Motor bearing	Normal	1750	0, 1,2,3 HP
		Inner ring fault	1750	0, 1,2,3 HP
		Rolling element fault	1750	0, 1,2,3 HP
		Outer ring fault	1750	0, 1,2,3 HP

3.4.2 Transfer Multiple Fault Diagnosis of the IDDAN

a) Experiments Setting

The planned transfer diagnosis experiment is shown in Fig. 7, which includes the use of each dataset and the partition of training data and test data. In this experiment, all used data are bearing vibration signals collected from different machines or devices. Therefore, this experiment put the target on testing the ability and performance of inferencing trained IDDAN to a new machine. It can be found that the dataset A and B are mixed as pretraining data. This is because the dataset A does not include vibration signal samples of rolling element fault. In No. 2 experiments, samples of two damage levels are also not included in the dataset B. No.1 experiment tests the accuracy of the proposed method with standard data amount (2000) which come from two different datasets, while No. 2 experiment tests with a larger amount of data (40000). The No.1 and No. 2 can be considered to evaluate the performance of DA.

In this section, the percentage of the dataset means that the data is picked evenly from each fault and each working condition in the dataset. 20% of the target domain dataset is used in this section as the fine-tuning dataset.



NC: normal condition REF: rolling element fault IRF: inner ring fault ORF: outer ring fault

No.	IDDTN transfer experiments	Training data	Test data
1	Dataset A + Dataset B → Dataset C	100% labeled Dataset A + 100% labeled Dataset B + 20% labeled Dataset C	50% unlabeled Dataset C
2	**DA**(Dataset A + Dataset B) → Dataset C	DA(100% labeled Dataset A + 100% labeled Dataset B) + 20% labeled Dataset C	50% unlabeled Dataset C

Fig. 3-7. Diagram of different transfer diagnosis experiments.

b) Validation

We evaluate the proposed IDDAN through two designed experiments shown in Fig. 3-7. In the first experiment, dataset A plus B refers to the source domain dataset and C refers to the target domain dataset. In the second experiment, A plus B after DA process turn into the source domain data, and C also refers to the target domain dataset. In each experiment, the training data includes all the labeled samples from the source domain dataset and the fine-turning data uses 20% labeled data from the target domain dataset. Then, we randomly take 50% of unlabeled samples as the test data. The test dataset is not contained in the training set.

The detailed information of training parameters is demonstrated as follows. The CPU and GPU devices for training are I9-12900K and RTX 3080Ti respectively. The epoch for pre-training and fine-tuning is set to be 500 and 150 respectively. In addition, the patch size is set as 32, the number of attention heads is 6. This step is trained using the AdamW optimizer with a learning rate 3×10^{-4}. In the fine-tuning step, the dimension of the MLP block is set as 2048, the resolution of target domain time-frequency images increases to 256 from 224. The training loss of experiments is plotted in Fig. 3-8. As shown in Fig. 3-8, the initial loss in the fine-tuning stage is significantly lower than in the pre-training stage.

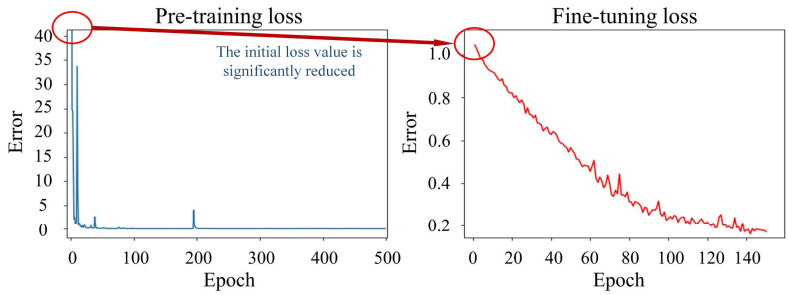


Fig. 3-8. The pre-training loss (500 epoch) and the fine-tuning loss (150 epoch) carve.

The No.1 experiment is repeated ten times with those pre-set parameters. In each experiment, all accuracies of the transfer diagnosis are over 82% and the average accuracy is around 85%. For obtaining a baseline testing accuracy, we set a control transfer test using the CNN model to condition recognition. In this CNN module, the number of convolutional layers is 5, and the size of the convolution kernel is set to be 3. The 5 pooling layers are followed by each convolutional layer separately and the size of the pooling kernel is 2. The CNN model is also pre-trained and fine-tuned using the same data samples. According to the same experiments times with IDDAN, the average accuracy is around 89% and the lowest accuracy is 85%. It means that the proposed IDDAN method can effectively diagnose the normal condition and three faults of bearing, but the accuracy of the CNN-based model is slightly higher than IDDAN with a small number of pre-training samples.

3.4.3 Fault Diagnosis Analysis of The Proposed Method

a) Effect analysis of data augmentation

Due to the difference between the self-attention mechanism and the convolutional network, self-attention-based network structures benefit more from large-scale data [172]. The experiments of this part take the same settings as Fig. 3-7 and the data augmentation strategy in Section III-A is applied in the experiment. We use MATLAB (signal processing toolbox) to perform the proposed data augmentation method on all vibration signals in the source domain dataset, followed by the signal-to-image conversion. The samples used in the No. 2 experiment were expanded from 2000 to 40000, reaching 20 times the original pre-training samples.

In this section, the No.2 experiments in Fig. 3-7 should be repeated several times. In these experiments, the amount of data used in the fine-tuning stage as a percentage of the total target domain data (P_T) is kept as 20%. It can be found that all accuracies of the transfer diagnosis

are over 92% and the average accuracy is around 95%. To provide visual insights into the effects of features transferring from the source datasets and target dataset, we use the *t*-distributed stochastic neighbour embedding (*t*-SNE) technique to map the high-dimensional features into a 2-D space. Fig. 3-9 demonstrates the feature recognition ability of the last layers before *softmax* layer when we completed the fine-tuning step.

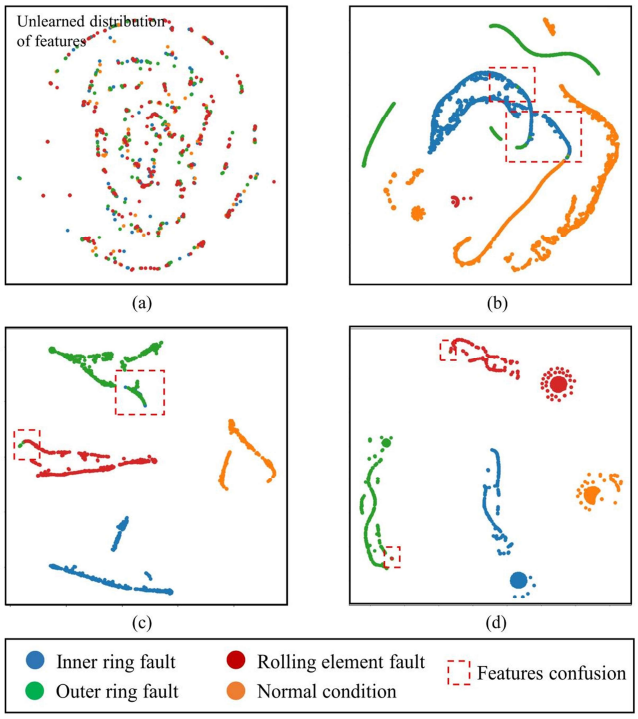


Fig. 3-9. t-SNE visualization of features after fine-tuning (P_T = 20%). (a) Before training. (b) Small amount of pre-training samples. (c) CNN method with large amount of pre-training samples. (d) IDDAN with large amount of pre-training samples.

b) FLOPs and parameters analysis of KD

When deploying the model in a device such as an electric vehicle (EV), the floating point operations (FLOPs) of the method must be considered. FLOPs describe the computing power required by the deployed device, and the number of parameters describes the required memory size.

In this paper, the KD strategy is used to further reduce the model parameters and required FLOPs, reducing the burden on equipment and memory. Fig. 3-10 illustrates FLOPs of IDDAN, which compares the number of model parameters and FLOPs between IDDAN and common deep learning frameworks.

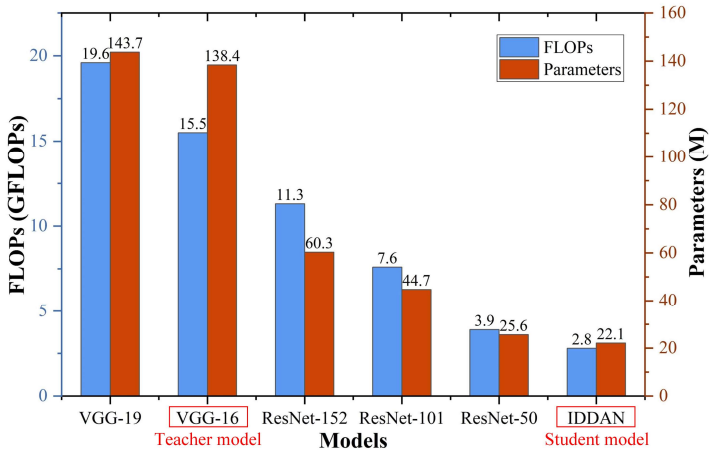


Fig. 3-10. t-SNE visualization of features after fine-tuning (P_T = 20%).

As shown in Fig. 3-10, the IDDAN achieves a relatively minimum requirement for FOLPs and parameter quantities compared to some commonly used frameworks. In detail, the required FLOPs of IDDAN are 2.8GFLOPs and the number of parameters is 22.1M, which is obviously less than the teacher model VGG-16.

3.4.4 Test Results Comparison

a) Multiple methods comparison

The dataset of bearing vibration signals under various working conditions is shown in Table 3-2. Therefore, a very harsh transfer learning test environment is formed, which can well detect the performance of the proposed method. However, the number of labeled samples provided by the target domain is also the key to performance. We divide nine grades according to P_T from less to more: 1%, 5%, 10%, 15%, 20%, 25%, 30%, 40% and 50%. The test data is still 50% unlabeled data of the target domain. It is necessary to compare the performance of the different methods. In this comparison, the historically outperforming classical algorithms support vector machine (SVM) and CNN are replaced with IDDAN, and short-time Fourier transform (STFT) and Hilbert-Huang transform (HHT) will be replaced with continuous wavelet transform (CWT). Data augmentation (DA) is also added as a controllable condition. Results of all tests are summarized in Table 3-2, which shows that the proposed method can achieve the best performance if $P_T \geq 10\%$.

TABLE 3-2 Fault Diagnosis Results of Various Methods

Methods \ P_T	1%	5%	10%	15%	20%	25%	30%	40%	50%
STFT+SVM	69.8%	73.9%	75.5%	77.2%	78.8%	82.7%	83.2%	87.3%	88.6%
STFT+CNN	67.9%	71.7%	78.4%	79.7%	85.3%	90.9%	91.0%	91.5%	92.4%
STFT+IDDAN	66.4%	66.8%	77.6%	78.1%	80.7%	84.4%	86.6%	93.2%	92.6%
HHT+SVM	**71.6%**	74.0%	80.9%	82.5%	80.4%	83.2%	81.5%	84.1%	89.8%
HHT+CNN	66.9%	75.5%	80.2%	84.6%	87.9%	91.3%	91.9%	90.7%	92.7%
HHT+IDDAN	60.1%	64.9%	77.1%	79.4%	84.2%	85.3%	90.6%	92.5%	93.6%
CWT+SVM	71.0%	79.4%	82.2%	83.7%	86.1%	87.8%	88.9%	90.5%	90.4%
CWT+CNN	69.9%	75.9%	80.3%	82.2%	89.5%	92.3%	92.2%	93.2%	93.9%
CWT+CNN+DA	71.4%	**79.3%**	86.7%	90.9%	94.1%	95.6%	95.7%	96.1%	96.2%
CWT+IDDAN	67.5%	69.2%	76.0%	80.5%	84.4%	87.9%	92.2%	92.7%	93.2%
CWT+IDDAN+DA	70.4%	85.5%	**87.2%**	**92.2%**	**95.9%**	**96.7%**	**97.3%**	**99.0%**	**99.5%**

b) Comparison with other methods

To demonstrate the performance of the proposed IDDAN, three different existing bearing diagnosis methods are used for comparison. Table 3-3 shows the comparison of the diagnostic results of other methods collected from the paper with the proposed IDDAN in transfer fault diagnosis experiments (P_T=50%). All methods are tested on CWRU dataset.

TABLE 3-3 Fault Diagnosis Results Compared with Other Existing Methods (P_T=50%)

Methods	Transfer method	Target domain dataset	Average accuracy
CNN	Fine-tuning	CWRU	93.9%
CNN-DA	Fine-tuning	CWRU	96.2%
DDC [179]	Maximum mean discrepancy	CWRU	78.2%
DCTLN [163]	Multiple domain adaptation	CWRU	86.8%
DANN [180]	Domain adversarial	CWRU	80.9%

IDDAN	Fine-tuning	CWRU	**93.2%**
IDDAN-DA	Fine-tuning	CWRU	**99.5%**

The results demonstrate that fine-tune-based methods outperform all compared methods. The accuracy of the CNN model is better than IDDAN in experiments without DA. However, IDDAN overtakes the CNN model after going through DA supporting. According to application conditions, it could be divided into two categories:

1) There is no any labeled data collected from the target domain machine. In this condition, domain adaptation-based methods are widely used for solving transfer learning problems. The most advanced research is based on the maximum mean discrepancy (MMD) of data samples. For example, the deep domain confusion (DDC) is to add an adaptation layer and an MMD module to the traditional CNN structure [179]. The domain adversarial training of neural networks (DANN) is the use of deep neural networks capable of domain discrimination [180]. The deep convolutional transfer learning networks (DCTLN) add two domain adaptation losses to the CNN loss function to optimize the MMD distance between the source domain and target domain [163]. However, this method will not reach the highest accuracy when diagnosing faults in the target domain.
2) There are a few labeled data collected from the target domain machine. Fine-tuning is currently widely used in computer vision to further adapt to the target domain after pre-training. And the accuracy of the model after fine-tuning is often related to the amount of pre-trained data. The application scenario is that the target domain can only provide a small amount of label data. At the same time, the self-attention mechanism included in IDDAN is proved to be more dependent on large-scale pre-training samples.

3.4.5 Multi-Level Fault Detection and Results

For further testing the ability of the proposed diagnostic method to identify fault features, we set two damage levels of inner ring fault and outer ring fault. In detail, the inner fault and the outer fault in CWRU bearing dataset are divided into 0.007 inches and 0.021 inches, while the inner fault and outer fault in KAt bearing dataset are also subdivided into two severity levels: within 2mm, between 2mm and 4.5mm.

The experiments are repeated several times with the proposed data augmentation process in the pre-training stage and the best experimental result is chosen to present. The result shows that

faults recognition accuracies of the CNN-based transfer method and IDDAN are around 84% and 92% (P_T =50%). The t-SNE features visualizations are demonstrated in Fig. 3-11. According to the information from sections 3.4.3, 3.4.4 and 3.4.5, we can further compare and analyse the effectiveness of IDDAN and other methods. We can observe the following points:

1) Compared with the classical CNN based transfer learning method, IDDAN achieves higher classification accuracy when given enough pre-training data. This means that transfer learning-based diagnosis accuracy has been further improved. Its purpose is to accurately classify various faults of the target machine when the labeled data of the target machine is not enough to directly train the diagnosis model.

2) When classifying the inter ring fault and outer ring fault into two damage levels (inter ring fault level 1, inter ring fault level 2, outer ring fault level 1 and outer ring fault level 2), the classification accuracy of the transfer learning-based diagnosis model decreases significantly. This is due to the high similarity between features of different damage levels.

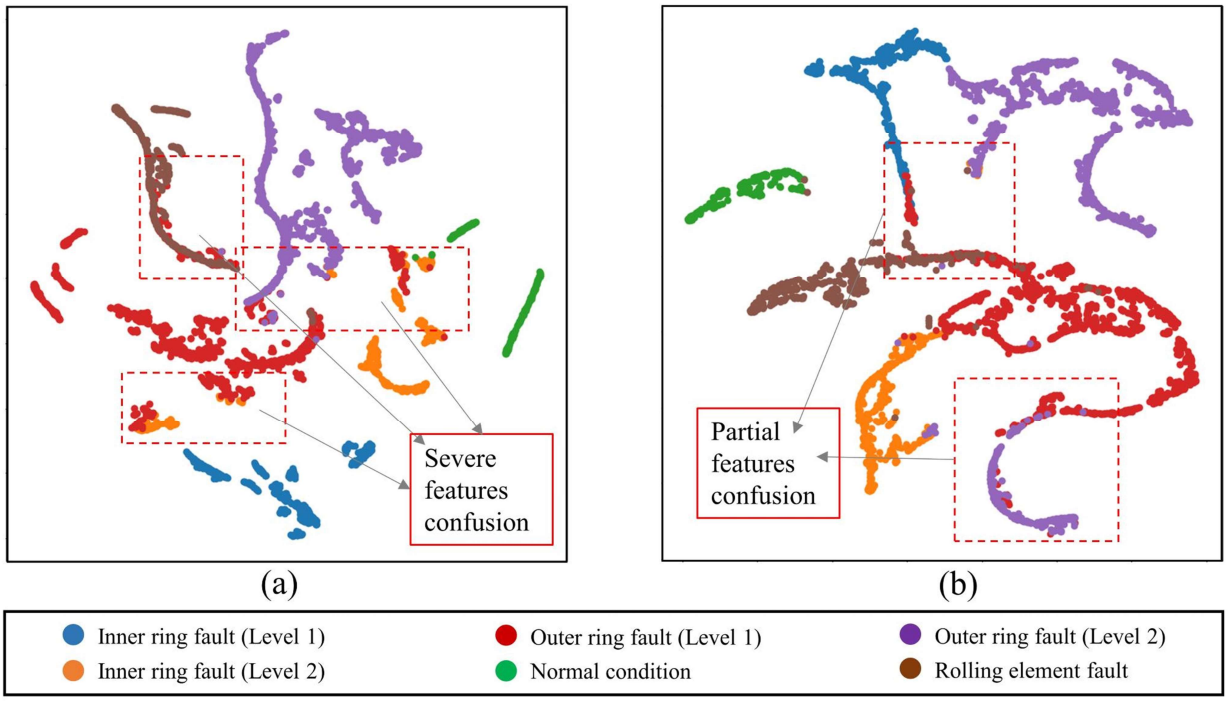


Fig. 3-11. t-SNE visualization of features after fine-tuning (P_T =50%). (a) CNN-based transfer method with large amount of pre-training samples. (b) IDDAN with large amount of pre-training samples.

As can be intuitively seen in Fig. 3-11, the classification model produces confusion between different fault types and levels. The proposed method has only smaller feature confusion part than CNN-based method when P_T = 50%.

3.5 Summary

This chapter proposes the self-attention mechanism in the field of online fault diagnosis of motor bearing with higher accuracy and proposed a new diagnosis framework based on IDDAN for solving the problem that it is hard to obtain enough labeled data to train a diagnosis model for a new target machine. Our experiment results present that the fine-tune-based transfer learning method could get better accuracy on the same dataset and the IDDAN has a better performance by pre-training using large-scale data. The mentioned DA method provides sufficient pre-training samples for IDDAN. Meanwhile, when IDDAN consumes large-scale data for pre-training, its diagnostic accuracy could surpass the CNN-based transfer learning model. The following points could conclude from this chapter:

1) The chapter proposes a self-attention mechanism-based intelligent fault diagnosis method IDDAN for deploying on new machines with a small number of labeled data by transfer learning.

2) The proposed DA method in Section 3.3.1 effectively expands the number of pre-training samples and has an excellent effect on IDDAN.

3) The IDDAN obtain a higher recognition accuracy of multiple bearing fault conditions with a large amount of pre-training data than the classic CNN-based method. When P_T=50%, The diagnosis accuracy of the proposed IDDAN method tested on CWRU dataset is 93.2%, the diagnosis accuracy of the proposed IDDAN-DA method tested on CWRU dataset is 99.5%, while the accuracy of the classic CNN-based method tested on CWRU dataset is 93.9%.

Chapter 4 Domain-Adversarial Adaptation Regression Model for IPMSM Permanent Magnet Temperature Estimation

This chapter aims to proposes a simulation-driven unsupervised transfer learning approach that leverages the domain-adversarial adaptation regression (DAAR) model to enable online and accurate estimation of PMT. Permanent magnet synchronous motors (PMSMs) are widely applied for industrial use as well as electric vehicles due to their remarkable electrical characteristics. However, PMSMs are vulnerable to damage from high temperatures, which can cause the demagnetization of the permanent magnets (PMs). To prevent demagnetization, accurate real-time estimation of the PM temperature (PMT) is crucial. Firstly, due to the difficulty in acquiring labeled PM data, lumped-parameter thermal network (LPTN)-based simulation provides a convenient way to obtain a large amount of simulation data. Secondly, this simulation dataset serves as the source domain data to train the proposed modified Transformer-based long sequence time-series regression (MT-LSTR) model, which generates coarse predicting labels (CP-labels). The CP-labels can be used to label the target domain dataset collected from a real interior PMSM (IPMSM). Finally, the DAAR model is trained with both feature-level and domain-level adaptations to obtain a well-trained online PMT estimator. In experiments, simulation data generated from MATLAB/Simulink is used for offline training, while the effectiveness of the proposed method is validated using the IPMSM temperature dataset provided by Paderborn University (PU).

4.1 Introduction

In recent years, permanent magnet synchronous motor (PMSM) is widely used in electric vehicles and other occasions due to its high torque and power densities along with a high degree of efficiency [181]. However, permanent magnets (PMs) are prone to irreversible demagnetization when overloaded under high thermal stress and must be monitored in various

industrial applications. Due to its complex rotating structure, the rotor is extremely difficult to monitor with embedded thermal sensors. Forcibly burying the sensor will lead to a sharp increase in production costs. These issues have led to the desire for accurate real-time estimation methods of PM temperature (PMT).

There are various physical model-based methods proposed and developed for estimating PMT, including lumped-parameter thermal networks (LPTNs)-based method [182], [183], flux observer-based method [184], [185], high-frequency (HF) signal injection-based method [186], [187], [188], signal fusion-based method [189], [190], [191]. Relatively speaking, LPTNs are one of the more successful and efficient approaches to approximate heat transfer within machines based on thermodynamic theory [192]. In conventional methods, LPTN is the representative of non-invasive method that does not require additional hardware and achieves high estimation accuracy for PMT with fewer model parameters. Paper [182] applies four nodes LPTN which model the stator yoke, stator winding, stator teeth, and the permanent magnets and achieves PMSM temperature estimation within 8°C error. To consider power losses of electromechanical power conversion of PMSM in more detail, [192] considered and introduced an extended iron loss model in LPTN to minify the error of temperature estimation. However, for a certain PMSM, the parameterization of its LPTN must rely on the knowledge intervention of domain experts. The geometry of the PMSM and the cold-blooded system will have a big impact on this. Two common PMT estimation methods, HF signal injection [193] [186] and flux observer [184], [185], are also used by industries. The principle of these two methods is to indirectly calculate the PMT value through the thermal properties of the electrical model parameters such as PM flux linkage or winding resistance. In addition, typical observers (Luenberger, Gopinath and Kalman, etc.) and direct thermal estimators can be merged and extended to new signal fusion method. Paper [190] combines Kalman filter and a state-space model to estimate PMT. And paper [191] combines a LPTN operating at low speeds and a Gopinath-style flux observer-based PMT observer operating at medium and high speeds. However, this type of methods is usually limited by high additional computational load and large application effort in the development stage.

Data-driven approaches often do not require domain expertise or specific drivetrain specifications for their topological designs (data collection and algorithm designing). Paper [194] attempts a data-driven approach, using residual deep recurrent and convolutional neural

network (RNN and CNN) to achieve highly dynamic temperature estimation of PMs. The authors of [194] trained a deep learning model using exhaustive data collected by testing, resulting in accurate predictions of PMT while avoiding expert thermal modeling. Paper [195] provides an empirical evaluation of several machine learning (ML) models, revealing the potential of simpler ML models in terms of regression accuracy, model size, and their data demand. In recent years, hybrid methods in between LPTNs and black-box ML models are currently advancing, which allow to incorporate general pre-knowledge regarding the underlying physical phenomena with a certain degree of interpretability [196], [197]. Furthermore, hybrid method generally has higher estimation accuracy than physical model-based methods and supervised black-box methods. For example, paper [197] proposes a thermal neural network (TNN) which consolidated knowledge in the form of LPTNs, and data-driven nonlinear function approximation with supervised ML. However, real-world data is an expensive and limited resource. In many cases, we do not have the required labeled data to train a ML model. When estimating different PMSM temperatures, it is difficult for us to obtain the corresponding rotor temperatures as training data. Therefore, this will be a significant hurdle for the deployment of deep learning-based methods in industry. Details of state-of-the-art methods on PMT estimation are summarized in Table 4-1.

TABLE 4-1 Summary of Up-to-data Existing Methods

Method type	Related works	Description	Deficiencies
LPTN-based	[182], [183]	Consisting of four thermal nodes is designed to model PMSM.	Requires some prior knowledge of model parameters.
Flux observer-based	[184], [185]	Estimating magnetization and PMT through fundamental wave flux observers.	Lower accuracy at low motor speed.
Neural network and LPTN-based hybrid method	[197], [196], [198]	An approach combining neural networks and PMSM thermal model.	Supervised by real PMT labels.
HF signal injection-based	[186], [187], [188]	HF signal is injected into the machine to estimate the HF resistance.	Causes extra loss and EMI problems.
Signal fusion-based	[190], [191]	Noninvasive method computing by PMSM steady-state equation.	High additional computational load at run time.

Supervised ML-based	[194], [195], [199], [200]	Forecasting PMT through a well-trained neural network model.	Training requires labeled PMT data.
Simulation and transfer learning-based	The proposed	Estimating PMT by transferring pre-trained knowledge from simulation data.	Training requires simulated PMT data.

To solve this problem, this paper proposes a simulation-driven and domain adaptation-based approach for real-time estimate PMT without measured PMT labels. Physical model-based simulation provides a solution for acquiring the datasets required for deep learning [201]. Domain adaptation techniques, as a type of transfer learning, can generalize the knowledge learned by model training without supervision [202]. It can not only perform model transfer in classification tasks, but also solve regression problems. For example, paper [203] proposes their deep transfer learning-based approach for online predicting remaining useful life (RUL) of rolling bearings. Up to now, deep learning model transfer between actual collected datasets has been successfully applied. Model transfer from simulated dataset to real dataset has only been applied once in the fault classification task [201]. The simulation-based domain adversarial adaptation regression (DAAR) scheme proposed in this paper can be used as a general method to build the deep learning estimation model for interior PMSMs (IPMSMs). In this chapter, the simulated IPMSM temperature data will be used as the source domain data in the domain adaptation framework. Experiments prove that the deep transfer learning-based PMT estimation model constructed by this method is effective and has high accuracy.

The main novelties and technical contributions of this paper can be summarized as follows:

1) Different from previous data-driven methods, the proposed approach can estimate PMT of target motor without labels from real measurement.
2) LPTN-based simulation model provides time-series IPMSM dataset under multi-working conditions for well training a modified Transformer-based long sequence time-series regression (MT-LSTR) model which provides coarse predicting labels (CP-labels) for feature adaptation and domain alignment processes.
3) This chapter proposes a novel DAAR method to real-time estimate PMT without labeled real data of target motor. CP-labels can be target domain labels of the DAAR model.

The rest of this chapter is presented as follows: In Section 4.2, necessary preliminaries are described synoptically. Section 4.3 describes the proposed approach based on PM thermal model simulation and domain regression adaptation technique. Section 4.4 presents the collected simulation dataset and the real experimental dataset, the Paderborn University (PU) IPMSM temperature dataset [194]. A detailed experimental validation is placed in Section 4.5 and results on the PU dataset [194] has been analysed. Section 4.6 is the conclusion of the whole chapter.

4.1.1 Description of Transfer Learning Problem

The essence of the model pre-training is transfer learning [171]. To clearly explain the proposed architecture in Section III, here it is necessary to introduce two types of domains in transfer learning: the source domain D_{s} and the target domain D_{t}. Both domains are composed of the feature space X and the probability distribution $P(X)$ of the data. The category spaces of the learning objectives of transfer learning in D_{s} and D_{t} are represented by Y_{s} and Y_{t}, respectively. When $D_{\mathrm{s}} \neq D_{\mathrm{t}}$, the data distribution before and after transfer is also different. Transfer learning can improve the performance of the target task learning function f_{t} when $D_{\mathrm{s}} \neq D_{\mathrm{t}}$ or $Y_{\mathrm{s}} \neq Y_{\mathrm{t}}$. The following Fig. 4-1 shows the principle of transfer learning to process unlabelled data task and what is different process between transfer learning and traditional neural networks. From the perspective of transfer learning, the data used for pre-training $X = \{x_1, x_2 \cdots x_n\}$ and the corresponding label space $Y_{\mathrm{s}} = \{y_{x_1}, y_{x_2} \cdots y_{x_n}\}$ compose source domain $D_{\mathrm{s}} = \{X, Y_{\mathrm{s}}\}$. The unlabelled data of the target machine is $X_{\mathrm{t}} = \{x_{\mathrm{t}1}, x_{\mathrm{t}2} \cdots x_{\mathrm{t}n}\}$, where $x_{\mathrm{t}n}$ is the data samples of the target task. The more adequate samples in D_{s}, the larger the category space, the stronger generalization ability, and the better performance of the model after pre-training transfer.

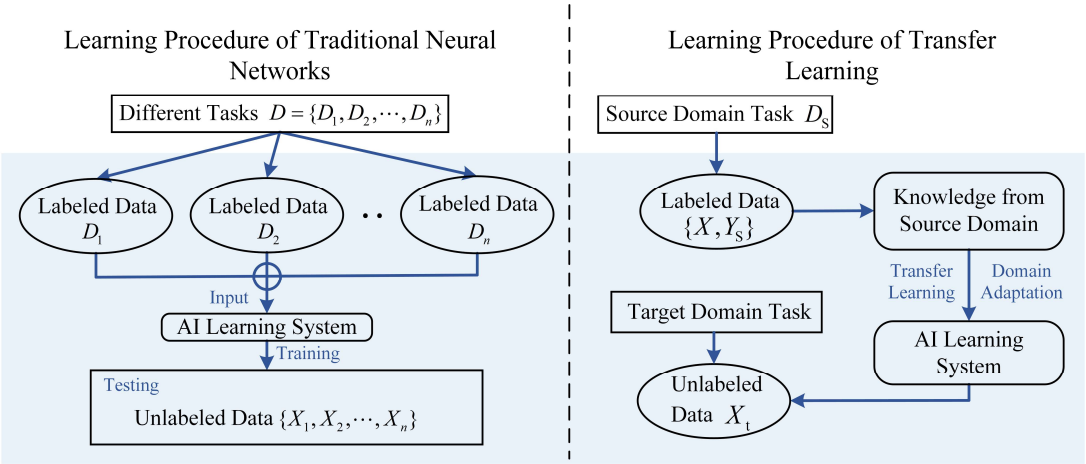


Fig. 4-1. The difference diagram of the learning procedure between traditional neural networks and unsupervised transfer learning.

4.1.2 Domain Adversarial Neural Network

Domain adversarial neural network (DANN) is a kind of unsupervised domain adaptation-based transfer learning, first proposed in [204]. DANN benefits from the idea of generative adversarial network (GAN) but unlike GAN because the domain adaptation problem does not require a generator but a feature extractor. As demonstrated in Fig. 4-1, a normal DANN consists of three parts: feature extractor $G_{\mathrm{f}}(\cdot\,;\theta_{\mathrm{f}})$, label predictor $G_{\mathrm{y}}(\cdot\,;\theta_{\mathrm{y}})$ and domain classifier $G_{\mathrm{d}}(\cdot\,;\theta_{\mathrm{d}})$. The goal of DANN is to extract features from the source and target domain separately so that the domain classifier cannot distinguish between the two. Therefore, the network loss of domain adaptation problem generally consists of two parts: training loss $\mathcal{L}_{\mathrm{y}}$ and domain discrimination loss $\mathcal{L}_{\mathrm{d}}$. As shown in Fig. 4-2, feature extractor and label predictor constitute a feed-forward neural network. During the training process, on the one hand, the network must continuously minimize the loss of the label predictor; on the other hand, the network must continuously minimize the loss function of the domain discriminator, which is connected with the feature extractor through a gradient reversal layer (GRL).

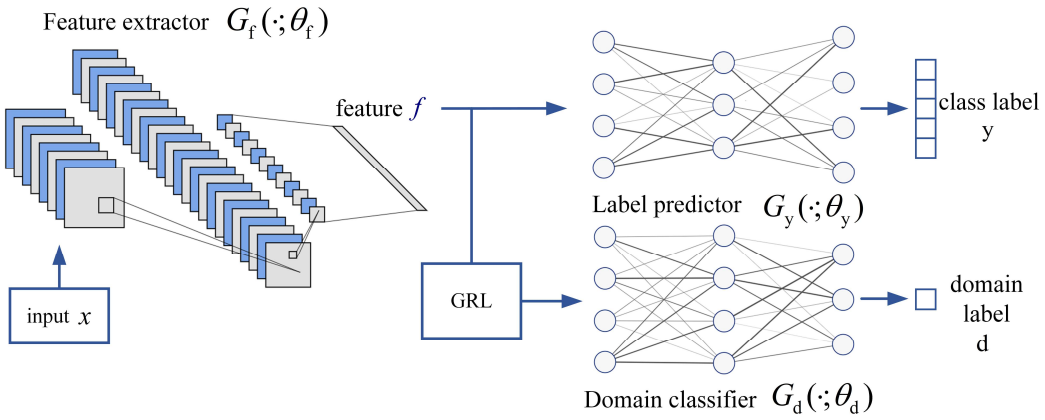


Fig. 4-2. Architecture diagram of the regular DANN.

4.2 The Proposed PMT Estimation Approach

In this section, a self-supervised simulation-driven domain regression adaptation method is proposed for real-time PMT estimation, as shown in Fig. 4-3. This proposed approach includes three parts: (1) Simulated PMT related signals acquirement. An LPTN-based thermal network is achieved by MATLAB/Simulink for collecting completed labeled temperature data; (2) Supervised model training using simulated data for learning information of source domain. A modified Transformer-based long sequence time-series regression (MT-LSTR) model is trained for generating coarse predicting labels (CP-labels) of feature adaptation and domain alignment processes; (3) The domain adversarial adaptation regression (DAAR) model which updates from classic DANN is used to perform the regression task of PMT estimation. It can extract domain-invariant feature information in source domain data based on CP-labels. Thus, the online real-time high-precision estimation model is obtained in the joint optimization guided by the multi-loss function. Fig. 4-3 illustrates the summarised architecture of the proposed approach. In this diagram, details of any part can be found in this section.

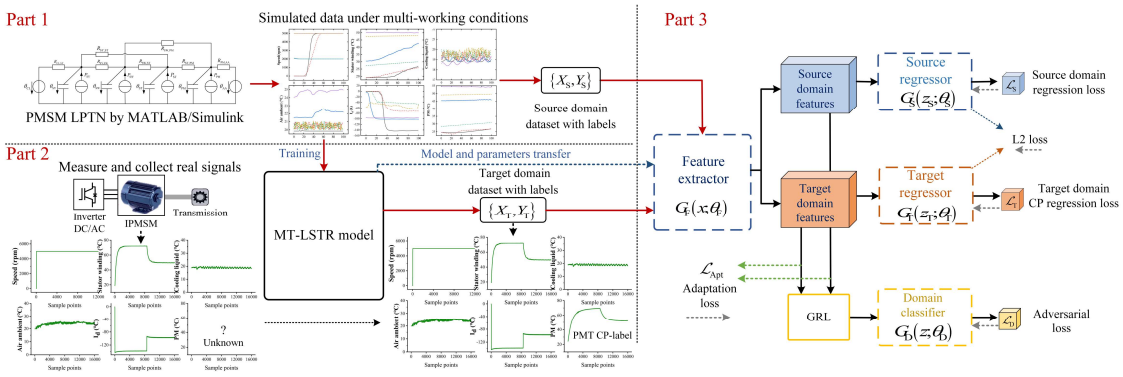


Fig. 4-3. Structure diagram of the whole proposed approach. (The approach includes three parts: simulation data acquisition part, MT-LSTF model part, DAAR part. In this figure, part 1 provides source domain dataset $\mathcal{D}_S$ to the training procedure of part 3, part 2 provides target domain dataset $\mathcal{D}_T$ to the training procedure of part 3).

4.2.1 IPMSM Modelling for PMT Simulation Data Acquisition

Electric machine is essentially a non-homogeneous thermal body, which could be expressed as a combination of an external thermodynamic environment and serval temperature nodes that connect each other by thermal resistances [205]. In this paper, we choose a LPTN-based IPMSM model to obtain all simulation data related to PMT. This paper employs the grey box model low-order LPTN of IPMSM that content a large number of nodes and machine's prior knowledge of material and geometry.

PMSM physical simulation model is vital for acquiring simulated PMT dataset. The low-order LPTN proposed by Wallscheid and Böcker is a physical model that effectively simulates the four most important components of IPMSM: stator yoke, stator winding, stator teeth and PMs [182]. Fig. 4 illustrates the structure of IPMSM' LPTN used in this section. In this paper, LPTN-based thermal network is constructed by MATLAB/Simulink for obtaining time-series simulation data.

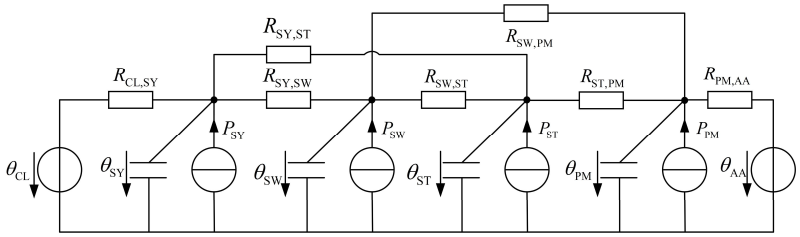


Fig. 4-4. Structure diagram of the low-order LPTN. [182] (**$\boldsymbol{\theta_{CL}}$ donates the cooling liquid temperature, $\boldsymbol{\theta_{SY}}$ donates the stator yoke temperature, $\boldsymbol{\theta_{SW}}$ donates the stator winding temperature, $\boldsymbol{\theta_{ST}}$ donates the stator teeth temperature, $\boldsymbol{\theta_{PM}}$ donates the PMT, $\boldsymbol{\theta_{AA}}$ donates the ambient air temperature, every $\boldsymbol{R}$ represents the thermal resistance between two nodes, and every $\boldsymbol{P}$ represents the inner heat generation within the system from different motor components (heat sources)).**

The inner heat generation within the IPMSM thermal system can be called as motor losses. The copper losses are equal to the stator winding losses, which can be calculated based on the actual current:

$$P_{v,\mathrm{Cu}} = P_{v,\mathrm{SW}} = 3 \cdot I^2 \cdot R_\mathrm{s}(\theta_\mathrm{SW}, n) \tag{4-1}$$

where R_s is the ohmic resistance. The difference between the copper losses $P_{v,\mathrm{SW}}$ and the overall power losses P_v are called residual losses, and assumed to be equal the the iron losses in the stator and rotor parts:

$$\tilde{P}_{v,\mathrm{Fe}} = \tilde{P}_{v,\mathrm{res}} = P_v\left(I, i_\mathrm{q}, n\right) - \tilde{P}_{v,\mathrm{SW}}\left(I, n, \theta_{\mathrm{SW},0}\right) = \tilde{P}_{v,\mathrm{PM}} + \tilde{P}_{v,\mathrm{SY}} + \tilde{P}_{v,\mathrm{ST}} \tag{4-2}$$

where $\tilde{P}_v$ denotes the equivalent copper losses in the different motor components of this system at reference temperature θ_0. Then the loss can be computed by:

$$P_{v,\mathrm{SY}} = \tilde{P}_{v,\mathrm{SY}} \cdot \left[1 + \alpha_\mathrm{Fe} \cdot \left(\theta_\mathrm{SY} - \theta_{\mathrm{SY},0}\right)\right] \tag{4-3}$$

$$P_{v,\mathrm{ST}} = \tilde{P}_{v,\mathrm{ST}} \cdot \left[1 + \alpha_\mathrm{Fe} \cdot \left(\theta_\mathrm{ST} - \theta_{\mathrm{ST},0}\right)\right] \tag{4-4}$$

$$P_{v,\mathrm{PM}} = \tilde{P}_{v,\mathrm{PM}} \cdot \left[1 + \alpha_\mathrm{Fe} \cdot \left(\theta_\mathrm{PM} - \theta_{\mathrm{PM},0}\right)\right] \tag{4-5}$$

$$\text{with } \alpha_\mathrm{Fe} \in \{\mathbb{R} | -1\%/\mathrm{K} \leq \alpha_\mathrm{Fe} \leq 0\}$$

where the temperature coefficient α_Fe is set as negative value, -1% is equal to -0.01. As content of the low-order LPTN of IPMSM, the seven thermal resistance of this model is different to be calculated. The three resistances $R_{\mathrm{SY,SW}}$, $R_{\mathrm{SW,ST}}$ and $R_{\mathrm{SY,ST}}$ associated with the stator can be calculated from the heat transfer Eq. (4-6) for the nodes inside the stator.

$$R_{i,j} = \frac{l_{i,j}}{\lambda_{i,j} \cdot A_{i,j}} \tag{4-6}$$

$$\text{with } i,j \in [\text{SY}, \text{SW}, \text{ST}] \text{ and } i \neq j$$

where $l_{i,j}$ is the effective conduction length from node i to node j, $\lambda_{i,j}$ is the thermal conductivity of material, A donates the effective cross section area. The value of the thermal resistance from the stator yoke to the cooling liquid jacket $R_{\text{CL,SY}}$ relies too much on the physical shape information, so it can only be roughly estimated by Eq. (4-7) in this LPTN model.

$$R_{\text{CL,SY}} = R_{\text{CL,SY,0}} \cdot \left[1 + \alpha_{\text{CL,SY}} \cdot \left(\theta_{\text{CL}} - \theta_{CL,0}\right)\right] \tag{4-7}$$

$$\text{with } \alpha_{\text{CL,SY}} \in \left\{\mathbb{R} \middle| -1\%/\text{K} \leq \alpha_{\text{CL,SY}} \leq 0\right\}$$

In above formula, coefficient $\alpha_{\text{CL,SY}}$ will be negative. The values of resistances $R_{\text{PM,ST}}$, $R_{\text{SW,PM}}$ and $R_{\text{PM,AA}}$ are defined to calculate model heat transfer:

$$R_{\text{PM},i}(n) = R_{\text{PM},i,0} \cdot e^{-\frac{n}{n_{\max}} \cdot \frac{1}{b_{\text{PM},i}}} + \alpha_{\text{PM},i} \tag{4-8}$$

$$\text{with } R_{\text{PM},i,0} \in \left\{\mathbb{R} \middle| 0 \leq R_{\text{PM},i,0} \leq R_{\text{PM},i,0,\max}\right\}$$

$$a_{\text{PM},i} \in \left\{\mathbb{R} \middle| 0 \leq a_{\text{PM},i} \leq a_{\text{PM},i,\max}\right\}$$

$$b_{\text{PM},i} \in \left\{\mathbb{R} \middle| 0 \leq b_{\text{PM},i} \leq b_{\text{PM},i,\max}\right\}$$

$$\text{with } i \in [\text{SW}, \text{ST}, \text{A}]$$

The temperature simulation data of PMs are obtained by MATLAB/Simulink simulation model based on the low-order LPTN. The simulated dataset is for the training of MT-LSTR model shown as Part 2 of Fig. 4-3, and also for training the DAAR model in part 3 of Fig. 4-3.

4.2.2 MT-LSTR Model for CP-labels Generating

In this section, MT-LSTR mode is built for transferring prior knowledge to the online PMT estimating function. The whole structure of this model is illustrated in Fig. 4-5. The MT-LSTR model can recursively predict the sequence of the online temperature through learning from offline simulation data. Then we compute the CP PMT label of online tested IPMSMs using this well-trained MT-LSTR model.

In this model, we originally employ modified Transformer-based LSTR model based on Informer model [206]. The whole structure of MT-LSTR model is shown in Fig. 4-5.

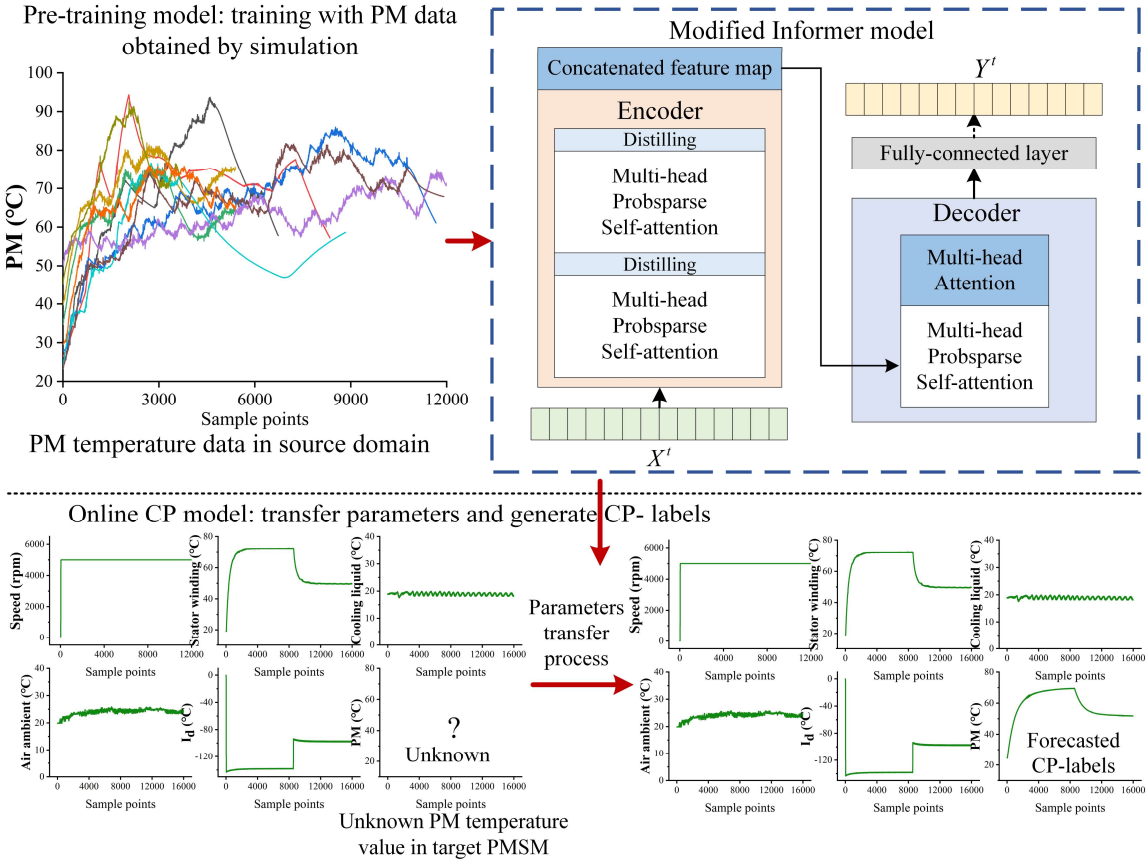


Fig. 4-5. Structure diagram of MT-LSTR and CP-labels generating. (The parameters transfer process means that the CP-labels of real dataset should be forecasted by the model trained by simulated data) (Source domain data is from simulation (both reference

signals and label) and CP-labels provide semi-supervised information for the training of DAAR model) (Detailed presentation of the part 2 in Fig. 4-3).

The encoder in Fig. 4-5 uses multi-head probsparse self-attention to give higher weight to the dominant feature from the j to $j+1$ layers, and the activation function uses exponential linear unit (ELU):

$$x_{j+1}^{t} = \mathrm{MaxPool}\left(\mathrm{ELU}\left(\mathrm{Conv1d}\left(\left[x_j^t\right]\right)\right)\right) \tag{4-9}$$

In order to enable the MT-LSTR model to further extract prior knowledge from offline simulation data and prevent outliers in training data from causing gradient explosion, the off-line training of MT-LSTR model combines mean square error (MSE) and mean absolute error (MAE) as the loss function used for optimization parameters of Informer decoder:

$$L_\delta\big(y, f(x)\big) = \begin{cases} \frac{1}{2}\big(y - f(x)\big)^2 & \text{, when } |y - f(x)| \le \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2 & \text{, when } |y - f(x)| > \delta \end{cases} \tag{4-10}$$

where δ is a hyperparameter which decides how to deal with outliers. L_δ can ensure that the model updates parameters at a fast speed when $|y - f(x)| > \delta$ and ensure that the model can obtain the global optimal value more accurately when $|y - f(x)| \le \delta$. Loss function (4-10) should be minimized by Adam gradient descent optimizer [207] to seek a group of well-optimized network weights W_x.

MT-LSTR should transfer off-line trained weight parameters from source domain to target domain to online generate CP-labels. The value of predicted PMT is computing by decoder and FC layers:

$$x_{\mathrm{de}}^{t} = Concat(x_{\mathrm{token}}^{t}, x_0^t), x_{\mathrm{token}}^{t} \in R^{L_{\mathrm{token}} \times d_{\mathrm{model}}} \tag{4-11}$$

$$Y = W \cdot X, d = d_{\mathrm{out}} \tag{4-12}$$

where $Concat$ is a mathematical operator, L_{token} represents the length of the token, d_{model} is the dimension of the model, W is the weight matrix.

4.2.3 Domain Adversarial Adaptation Regression (DAAR) Model with Multi-Working Condition for PMT Transfer Estimation

In many practical cases, the speed and torque set during the source domain data acquired through the simulation process are highly random as well as the verification data, which means that the source domain and target domain data have the characteristics of strong distribution complexity. The source domain and target domain data are designed to simulate the complex and changeable real working conditions in the actual application of the IPMSM. If we use the CP-labels value of the target domain obtained from the MT-LSTR model directly as the estimated value of PMT, it is not accurate enough, which is determined by the cumulative prediction error caused by the different distribution of the data in the two domains. Since the true labels are unavailable at all, CP-labels are still valuable for the domain adaptation regression, which can provide necessary target domain supervision information that classic DANN does not have during training. This means that CP-labels can provide data distribution alignment direction and domain-invariance features for DANN's domain adaptation task. Compared with traditional DANN, CP-labels can benefit to establish a more precise domain adaptation prediction model.

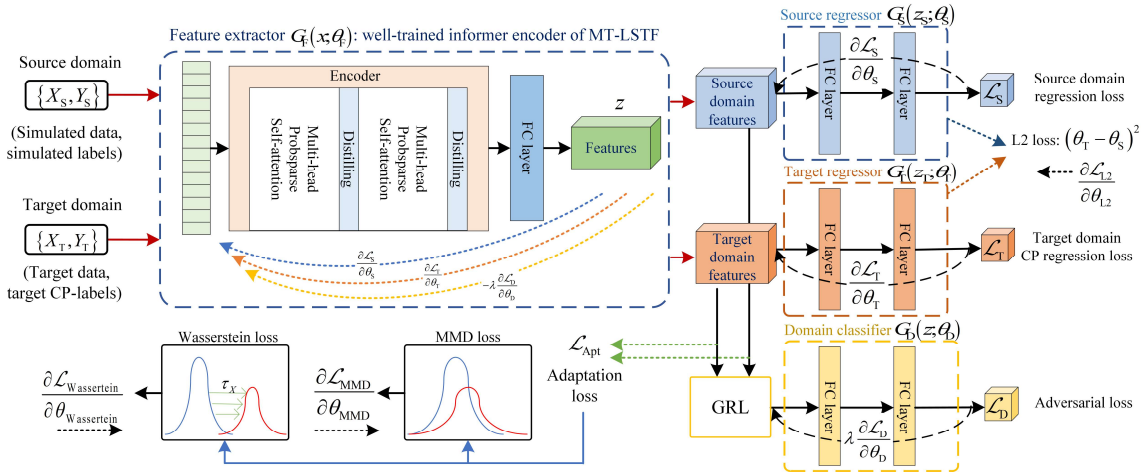


Fig. 4-6. Architecture of the DAAR model for PMT high-precision estimation.(The DAAR model includes a feature extractor $G_F(x;\theta_F)$, a source domain regressor $G_S(z_S;\theta_S)$ with source domain regression loss $\mathcal{L}_S(\theta_F,\theta_S)$, a target domain $G_T(z_T;\theta_T)$ with target domain

regression loss $\mathcal{L}_{\mathrm{T}}(\theta_{\mathrm{F}}, \theta_{\mathrm{T}})$ and $G_{\mathrm{D}}(z; \theta_{\mathrm{D}})$ with adversarial loss $\mathcal{L}_{\mathrm{D}}(\theta_{\mathrm{F}}, \theta_{\mathrm{D}})$ and a softmax layer) (Detailed presentation of the part 3 in Fig. 4-3).

The architecture in Fig. 4-6 is our DAAR model. This similar structure was first proposed in [201] and used for DANN-based classification task and in this paper for regression forecasting task. Compared with classic DANN model, DAAR model has two different parts named source domain recursive predictor $G_{\mathrm{S}}(z_{\mathrm{S}}; \theta_{\mathrm{S}})$, target domain recursive predictor $G_{\mathrm{T}}(z_{\mathrm{T}}; \theta_{\mathrm{T}})$ and two similar parts named feature extractor $G_{\mathrm{F}}(x; \theta_{\mathrm{F}})$and domain classifier $G_{\mathrm{D}}(z; \theta_{\mathrm{D}})$. θ_{S}, θ_{T} and θ_{F} donate the parameters of G_{S}, G_{T} and G_{D}, respectively. In this model, we use the pre-trained Informer encoder in MT-LSTR as G_{F}, which is enough to extract deep features from both source domain and target domain. This means using the transfer learning method to directly move the parameters trained in Fig. 4-3.Part 2 as the initial parameters of G_{F}. This strategy can better increase the speed of model training and converge to the optimal result faster. G_{S} and G_{T} consists of two fully-connected (FC) layers (Eq. (4-13)) and one output point. G_{S} and G_{T} aim to forecast PMT value and calculate $\mathcal{L}_{\mathrm{S}}(\theta_{\mathrm{F}}, \theta_{\mathrm{S}})$ and $\mathcal{L}_{\mathrm{T}}(\theta_{\mathrm{F}}, \theta_{\mathrm{T}})$ assisted with source labels Y_{S} and target domain CP-labels Y_{T}, respectively. G_{D} includes two FC layers and one softmax layer, which has an adversarial loss $\mathcal{L}_{\mathrm{D}}(\theta_{\mathrm{F}}, \theta_{\mathrm{D}})$. G_{S}, G_{T} and G_{D} all adopt the cross-entropy (CE) loss function which is mainly used to determine how close the actual output is to the expected output. Equations of FC layers and softmax layer are shown in (4-13) and (4-14), in which W is the weight matrix, z is the input vector, K donates the number of classes in the multi-class classifier, e is the natural constant.

$$Y = W \cdot X \tag{4-13}$$

$$softmax(z^i) = e^{z^i} \Big/ \sum_{j=1}^{K} e^{z^i} \tag{4-14}$$

In the source regressor, the input of $G_{\mathrm{S}}(z_{\mathrm{S}}^i; \theta_{\mathrm{S}})$: $\mathbb{R}^{\mathrm{D}} \rightarrow [0,1]$ is $z_{\mathrm{S}}^i = G_{\mathrm{F}}(x_{\mathrm{S}}^i; \theta_{\mathrm{F}})$. G_{F} has the function of extracting a D-dimensional feature vector z_{S}^i. $\mathcal{L}_{\mathrm{S}}^i(\theta_{\mathrm{F}}, \theta_{\mathrm{S}})$ donates the loss function of G_{S} and could be defined as:

$$\mathcal{L}_{\mathrm{S}}^i(\theta_{\mathrm{F}}, \theta_{\mathrm{S}}) = \mathrm{CE}\big(G_{\mathrm{S}}\big(G_{\mathrm{F}}(x_{\mathrm{S}}^i; \theta_{\mathrm{F}}); \theta_{\mathrm{S}}\big), y_{\mathrm{S}}^i\big) \tag{4-15}$$

where $\left(x_{\mathrm{S}}^{i}, y_{\mathrm{S}}^{i}\right)$ is the i-th given example of source domain dataset $\left\{x_{\mathrm{S}}^{i}, y_{\mathrm{S}}^{i}\right\}_{i=1}^{m}$. In the target regressor, the input of $G_{\mathrm{T}}\left(z_{\mathrm{T}}^{i}; \theta_{\mathrm{T}}\right)$: $\mathbb{R}^{\mathrm{D}} \rightarrow [0,1]$ is $z_{\mathrm{T}}^{i} = G_{\mathrm{F}}\left(x_{\mathrm{T}}^{i}; \theta_{\mathrm{F}}\right)$. The loss function of G_{T} can be represented as:

$$\mathcal{L}_{\mathrm{T}}^{i}(\theta_{\mathrm{F}}, \theta_{\mathrm{T}}) = \mathrm{CE}\left(G_{\mathrm{T}}\left(G_{\mathrm{F}}\left(x_{\mathrm{T}}^{i}; \theta_{\mathrm{F}}\right); \theta_{\mathrm{T}}\right), y_{\mathrm{T}}^{i}\right) \tag{4-16}$$

where $\left(x_{\mathrm{T}}^{i}, y_{\mathrm{T}}^{i}\right)$ is the i-th given example of target domain dataset $\left\{x_{\mathrm{T}}^{i}, y_{\mathrm{T}}^{i}\right\}_{i=1}^{n}$. In the domain classifier, a domain classifier $G_{\mathrm{D}}\left(z^{i}; \theta_{\mathrm{D}}\right)$ learns a logistic regressor G_{D}: $\mathbb{R}^{\mathrm{D}} \rightarrow [0,1]$, in which $z = G_{\mathrm{F}}\left(x^{i}; \theta_{\mathrm{F}}\right)$. $\mathcal{L}_{\mathrm{D}}^{i}(\theta_{\mathrm{F}}, \theta_{\mathrm{D}})$ donates the loss function of G_{D} and could be defined as:

$$\mathcal{L}_{\mathrm{D}}^{i}(\theta_{\mathrm{F}}, \theta_{\mathrm{D}}) = \mathrm{CE}\left(G_{\mathrm{D}}\left(G_{\mathrm{F}}\left(x^{i}; \theta_{\mathrm{F}}\right); \theta_{\mathrm{D}}\right), d^{i}\right) \tag{4-17}$$

where d^{i} represents the binary domain label of the i-th example, which is used to indicate whether this sample belongs to the source domain or the target domain.

In domain level, we separately introduce L2 loss $\mathcal{L}_{\mathrm{L2}}(\theta_{\mathrm{T}}, \theta_{\mathrm{S}})$ to represent the difference between source and target domain model parameters. In feature level, $\mathcal{L}_{\mathrm{Apt}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right)$ is designed to reduce the distance between the features of two domain, maximize domain confusion and make DAAR models more discriminative and domain invariant. $\mathcal{L}_{\mathrm{Apt}}$ can be obtained by summing maximum mean discrepancy (MMD) loss [202] $\mathcal{L}_{\mathrm{MMD}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right)$ and 1-D Wasserstein distance-based loss $\mathcal{L}_{\mathrm{Wasserstein}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right)$ [208].

$$\mathcal{L}_{\mathrm{L2}}(\theta_{\mathrm{T}}, \theta_{\mathrm{S}}) = |\theta_{\mathrm{T}} - \theta_{\mathrm{S}}|^{2} \tag{4-18}$$

$$\mathcal{L}_{\mathrm{Apt}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right) = \mathcal{L}_{\mathrm{MMD}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right) + \mathcal{L}_{\mathrm{Wasserstein}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right) \tag{4-19}$$

$$\begin{aligned} \mathcal{L}_{\mathrm{MMD}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right) &= \left\| \frac{1}{m}\sum_{i=1}^{m} \phi\left(z_{\mathrm{S}}^{i}\right) - \frac{1}{n}\sum_{i=1}^{n} \phi\left(z_{\mathrm{T}}^{i}\right) \right\|_{\mathcal{H}}^{2} \\ \mathcal{L}_{\mathrm{Wasserstei}}\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right) &= \inf_{\pi \in \Gamma(P_{\mathrm{S}}, P_{\mathrm{T}})} \int_{\mathbb{R}\times\mathbb{R}} \left|z_{\mathrm{S}}^{i} - z_{\mathrm{T}}^{i}\right| \mathrm{d}\pi\left(z_{\mathrm{S}}^{i}, z_{\mathrm{T}}^{i}\right) \end{aligned} \tag{4-20}$$

where m and n are the number of samples in source domain and target domain. $\phi(\cdot)$ is a mapping function, $\Gamma(P_{\mathrm{S}}, P_{\mathrm{T}})$ represents the set of probability distributions on $\mathbb{R} \times \mathbb{R}$. Therefore,

the total loss of the DAAR model could be represented as:

$$\begin{aligned}\mathcal{L}_{\mathrm{DAAR}} &= \frac{1}{m}\sum_{i}^{m}\mathcal{L}_{\mathrm{S}}^{i}(\theta_{\mathrm{F}},\theta_{\mathrm{S}}) + \frac{1}{n}\sum_{i}^{n}\mathcal{L}_{\mathrm{T}}^{i}(\theta_{\mathrm{F}},\theta_{\mathrm{T}}) - \mu\mathcal{L}_{\mathrm{D}}^{i}(\theta_{\mathrm{F}},\theta_{\mathrm{D}}) + \lambda\mathcal{L}_{\mathrm{L2}}(\theta_{\mathrm{T}},\theta_{\mathrm{S}}) + \gamma\mathcal{L}_{\mathrm{Apt}}\left(z_{\mathrm{S}}^{i},z_{\mathrm{T}}^{i}\right) \\ &= \frac{1}{m}\sum_{i}^{m}\mathcal{L}_{\mathrm{S}}^{i}(\theta_{\mathrm{F}},\theta_{\mathrm{S}}) + \frac{1}{n}\sum_{i}^{n}\mathcal{L}_{\mathrm{T}}^{i}(\theta_{\mathrm{F}},\theta_{\mathrm{T}}) \\ &\quad + \mu\left(\frac{1}{m}\sum_{i}^{m}\mathcal{L}_{\mathrm{D}}^{i}(\theta_{\mathrm{F}},\theta_{\mathrm{D}}) + \frac{1}{n}\sum_{i}^{n}\mathcal{L}_{\mathrm{D}}^{i}(\theta_{\mathrm{F}},\theta_{\mathrm{D}})\right) + \lambda\mathcal{L}_{\mathrm{L2}}(\theta_{\mathrm{T}},\theta_{\mathrm{S}}) + \gamma\mathcal{L}_{\mathrm{MMD}}\left(z_{\mathrm{S}}^{i},z_{\mathrm{T}}^{i}\right) \\ &\quad + \gamma\mathcal{L}_{\mathrm{Wasserstein}}\left(z_{\mathrm{S}}^{i},z_{\mathrm{T}}^{i}\right)\end{aligned} \tag{4-21}$$

4.2.4 Implement Procedure of DAAR

In the training process, the parameters in DAAR model will be updated with each iteration by using optimizer Adam to obtain the optimal parameters. The pseudocode of the whole procedure of the proposed approach is shown in Algorithm 1.

Algorithm 1: DAAR model training with target domain CP-labels

Input: Labeled source domain dataset $\mathcal{D}_{\mathrm{S}} = \left\{x_{\mathrm{S}}^{i}, y_{\mathrm{S}}^{i}\right\}_{i=1}^{m}$ obtained from IPMSM thermal simulation, target domain dataset CP-labels $\mathcal{D}_{\mathrm{T}} = \left\{x_{\mathrm{T}}^{i}, y_{\mathrm{T}}^{i}\right\}_{i=1}^{n}$, required hyperparameters μ, λ and γ.

Output: The forecasted PMT value y_{T} and PM estimation accuracy of testing data $\mathcal{D}_{\mathrm{test}}$.

1 Transfer and keep parameters of encoder of MT-LSTR model as initial parameters θ_{F} of G_{F}. Initialize parameters θ_{S}, θ_{T} and θ_{D} of G_{S}, G_{T} and G_{D};

2 **for** *each training iteration* **do**

3 Forward propagation through G_{F}, G_{S} and G_{D};

4 Input $\mathcal{D}_{\mathrm{S}}$ into G_{F} to extract source domain features z_{S} with parameters θ_{F};

5 Compute $\mathcal{L}_{\mathrm{S}}^{i}\left(G_{\mathrm{S}}, y_{\mathrm{S}}^{i}\right)$ and $\mathcal{L}_{\mathrm{D}}^{i}\left(G_{\mathrm{D}}, d^{i}\right)$ by using Eq. (4-15) and Eq. (4-17) with the parameters θ_{S} and θ_{D};

6 Forward propagation through G_{S} and G_{D};

7 Input $\mathcal{D}_{\mathrm{T}}$ into G_{F} to extract target domain features z_{T} with parameters θ_{F};

8 Compute $\mathcal{L}_{\mathrm{T}}^{i}\left(G_{\mathrm{T}}, y_{\mathrm{T}}^{i}\right)$ and $\mathcal{L}_{\mathrm{D}}^{i}\left(G_{\mathrm{D}}, d^{i}\right)$ by using Eq. (4-16) and Eq. (4-17) with the parameters θ_{T} and θ_{D}.

9 Compute $\mathcal{L}_{\mathrm{L2}}$ with two domain regressor parameters θ_{S} and θ_{T};

10 Compute $\mathcal{L}_{\mathrm{Apt}}$ by adding two feature level losses $\mathcal{L}_{\mathrm{MMD}}$ and $\mathcal{L}_{\mathrm{Wasserstein}}$ (Eq. (4-20));

11 Compute the total loss $\mathcal{L}_{\mathrm{DAAR}}$ by using Eq. (4-21).

Backpropagation with Adam optimizer to minimize $\mathcal{L}_{\mathrm{DAAR}}$.
Update the parameters θ_{F}, θ_{S}, θ_{T} and θ_{D};
Obtain a group of optimal parameters $\theta_{\mathrm{F}}{}^{*}$, $\theta_{\mathrm{S}}{}^{*}$, $\theta_{\mathrm{T}}{}^{*}$ and $\theta_{\mathrm{D}}{}^{*}$ of DAAR model;
end
Evaluate the model with testing data $\mathcal{D}_{\mathrm{test}}$ and compute the final forecasted PMT value y_{T}.

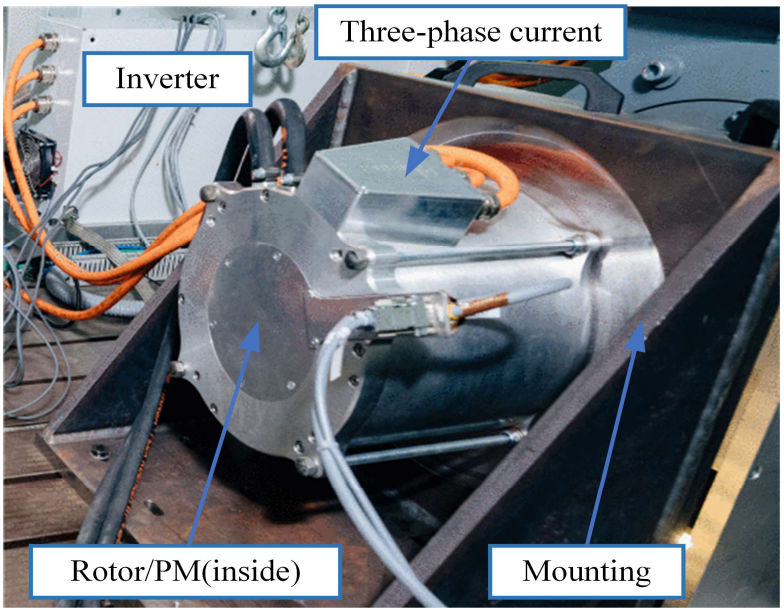
4.3 Experimental Datasets

4.3.1 Experimental Dataset Description

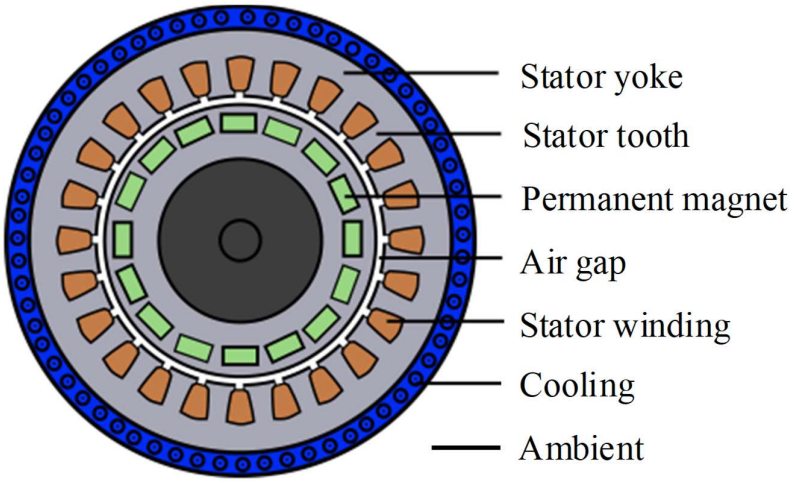
The PMT dataset comprises several sensor data collected from a IPMSM deployed on a test bench (Fig. 4-7(a)), which were collected by the LEA department at PU [194]. Fig. 4-7(b) shows the cross-section of this IPMSM. The whole dataset includes 55 drive circles/profiles and totally 185 hours, in which each drive circle contains several hours sampled data. In validation experiments, we retain the data of 10 drive circles in the PU dataset as test data, and all the rest of the data is used for the training of DAAR model. All signals are sampled at 2Hz, which include coolant liquid temperature θ_{CL}, stator winding temperature θ_{SW}, stator tooth temperature θ_{ST}, motor speed v, PMT θ_{PM}, stator yoke temperature θ_{SY}, currents in d/q-coordinates $I_{\mathrm{d}}/I_{\mathrm{q}}$, voltages in d/q-coordinates $U_{\mathrm{d}}/U_{\mathrm{q}}$, ambient air temperature $\theta_{\mathrm{AA}}$and torque τ. All data are collected by using a standard control strategy. According to the work condition, the data set consists of multiple measurement sessions. Most driving cycles denote random walks in the speed-torque-plane in order to imitate PMSM complex work conditions in real-world driving to a more accurate degree than just constant excitations and ramp-ups and -downs would. Furthermore, to simplify the complexity of model training and reduce the complexity of this method in practical applications, we use heatmap to visualize the correlation matrix to find which features are most suitable as model training data.

As shown in Fig. 4-7(c), features which have the absolute value of correlation coefficient between θ_{PM} and themselves$|\rho| \geqslant 0.45$ are selected as the features required for training and estimation of this method. Among them, the correlation coefficient between θ_{SW}, θ_{ST}, and θ_{SY} are all $\rho \geqslant 0.8$. Considering that too high or too low correlation will have some negative impact on training so that we only keep θ_{SW} in these three stator temperatures. In this paper, five

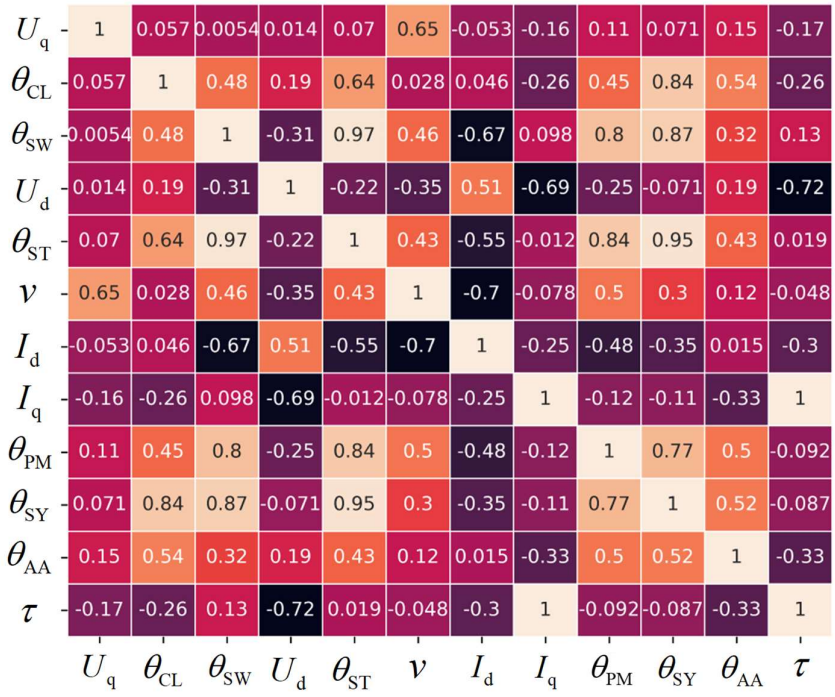
reference signals for PMT estimation are chosen as v, θ_{SW}, θ_{CL}, θ_{AA}, I_{d}.



(a)



(b)



(c)

Fig. 4-7. (a) Test bench used in this paper with a three-phase automotive traction IPMSM for PU temperature dataset [194]. (b) Schematic cross-section of the IPMSM used in the test bench [197]. (c) The correlation coefficient visualization of all features using Python heatmap function.

4.3.2 Simulated Time-series PMT Related Signals

The motor is excited by designed driving cycles that specify a reference motor speed and a reference torque. Simulated datasets are collected by many pseudo-random combinations of speed and torque. In the case of extremely sufficient simulation data, we can approximately consider that this dataset contains almost all working conditions of the PMSM used in the real world. And the signal length of the simulation signals is set equal to the length of the real ones. In this paper, we consider a simulated motor of the same type slightly different from the experimental motor, to reflect the proposed domain adaptation model can be used in cases where two domain datasets have certain differences. The type of PMSM for both PU and simulation dataset are interior. For IPMSM of PU, the power rating is 52kW, the voltage is 177V, the current rating is 283A, the torque rating is 250 N·m and the cooling type is water-

glycol; For IPMSM of simulation, the power rating is 23.4kW, the voltage is 325V, the current rating is 80A, the torque rating is 40 N·m and the cooling type is simulated liquid cooling. This procedure should generate simulated signals of different working conditions, all of which are time-series that present the feature signals of v, θ_{SW}, θ_{CL}, θ_{AA}, Id and label signal θ_{PM}.

In order to visualize the difference and distance between simulated data and real data, we choose three short real drive circles to show real and simulated time-series curves in Fig.4-8. For example, in an acceleration drive circle, rotor speed increases from zero to 5000 rpm and θ_{SW} and θ_{PM} rise constantly; In a ramp-up period, rotor speed is decreasing to a lower value, with θ_{SW} and θ_{PM} rise laxly. In a constant speed period, θ_{SW} and θ_{PM} are basically maintained at an upper limit, and θ_{CL} is slowly rising. In all working conditions, θ_{AA} is kept constant at 20℃. For simulating complex drive circles, speed and torque are controlled to change at the speed-torque-plane. To be similar to the working conditions in PU dataset, the speed is changed between 0 to 6000 rpm, and the torque is changed between -300 to 300 N·m.

All data analysis and experiments are conducted on computing device equipped with AMD Ryzen 9 5900X processor, 32 GB memory and NVIDIA GeForce RTX 3080Ti.

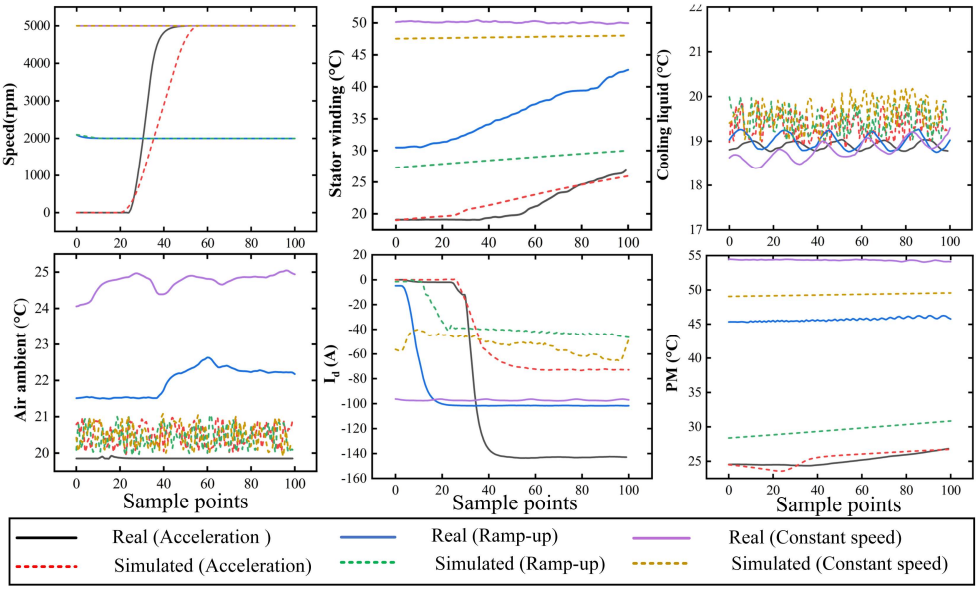


Fig. 4-8. Comparison of simulation data and experimental real data in the time domain under three short drive circles.

4.4 Verification Result

This section describes the validation results based on the proposed PMT estimation approach in Section 4.2 and experimental datasets in Section 4.3.

4.4.1 Performance Index

The result of the proposed method is evaluated using two quantitative indexes: MAE and root mean square error (RMSE), which are represented by Eq. (4-22) and Eq. (4-23), respectively. These evaluation metrics provide a clear indication of the accuracy of the PMT estimation task. The MSE and MAE measures assess the level of deviation between the estimated PMT values and the ground truth. Together, these indexes provide a comprehensive assessment of the accuracy and fidelity of the proposed method in estimating PMT values.

$$\mathrm{MAE} = \frac{1}{L_\mathrm{s}} \sum_{n=1}^{L_\mathrm{s}} |y^n - \hat{y}^n| \tag{4-22}$$

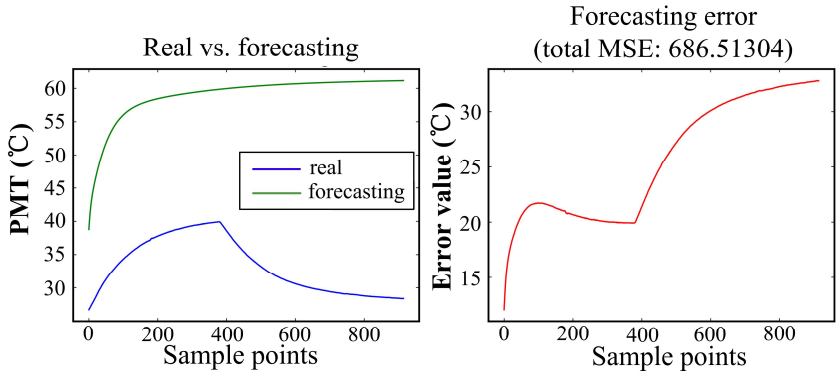
$$\mathrm{RMSE} = \sqrt{\frac{1}{L_\mathrm{s}} \sum_{n=1}^{L_\mathrm{s}} (y^n - \hat{y}^n)^2} \tag{4-23}$$

where y^n and $\hat{y}^n$ are the true value and estimated value of the DPT waveform amplitude, n is the nth sample point, and L_s is the length of samples. These indexes are used to evaluate performance of methods under different working conditions in Section 4.4.4.

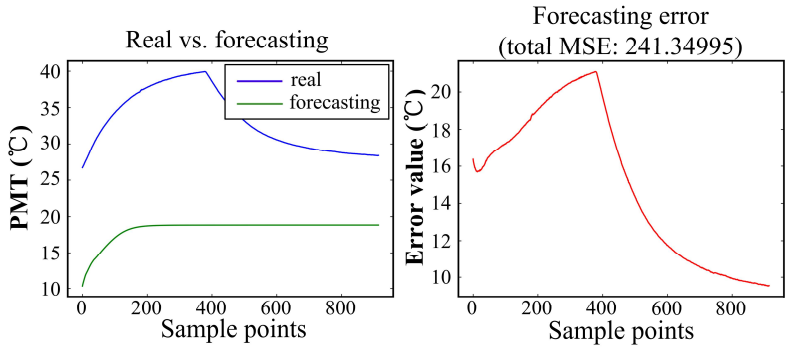
4.4.2 Results of CP-labels Generation

In this part, we want firstly to discuss the reason to choose MT-LSTR model to do the forecasting work. For validating the performance of proposed method, we try to obtain CP-labels by using three classic regression model with drive circle 17 of PU dataset: support vector regression (SVR) [209], random forest regression (RFR) [200], long short-term memory

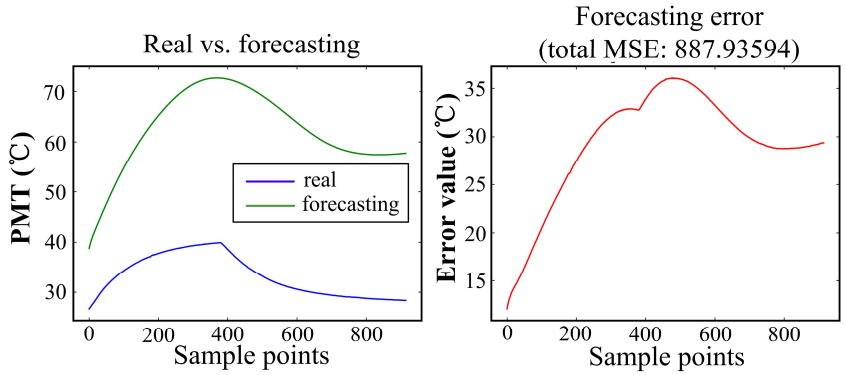
(LSTM) network [210]. SVR estimator choose radial basis function (RBF) kernel. For further parameters, regularization parameter set as 1, kernel coefficient for RBF set as $1/n_{\text{features}}$. For the RFR estimator, the number of trees set as 20, the number of features to consider when looking for the best split is n_{features}, the optimizer set as MSE. For LSTM algorithm, the network includes two LSTM layers, three FC layers and two dropout layers with dropout rate 0.2. The learning rate sets as 0.01, learning rate decay is 0.0005, and the optimizer chooses as Adam. For MT-LSTR model, the encoder includes one attention block, two one-dimensional convolutional layers and two layer-normalization layers. The activation function is ReLU. The kernel size of convolutional layer is 1, the rate of dropout layers is 0.1. The training epoch of all four algorithms is 100. Fig. 4-9 provides four examples of CP-labels generating of each comparative algorithms to show the capacity of MT-LSTR.



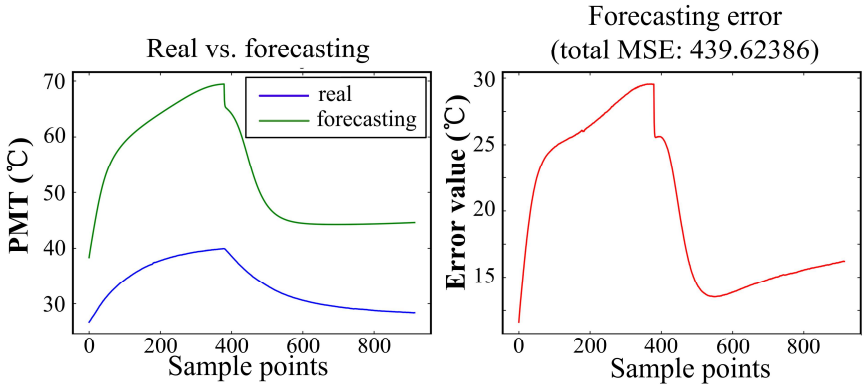
(a)



(b)



(c)



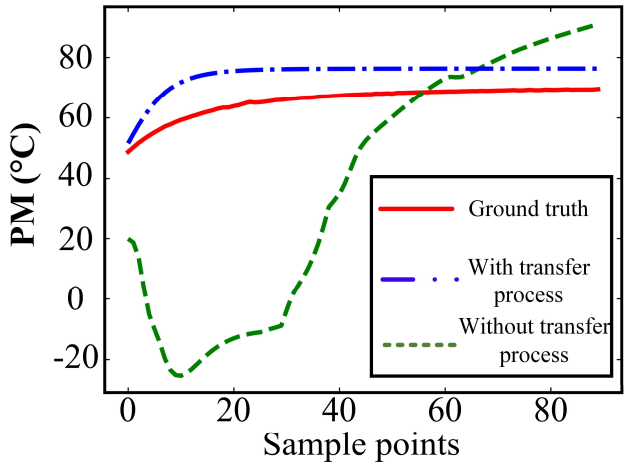
(d)

Fig. 4-9. Evaluation of forecasting performance and error by comparing algorithms (Smoothed result for drive circle 17) (a) RFR, (b) SVR, (c) LSTM, (d) MT-LSTR (proposed).

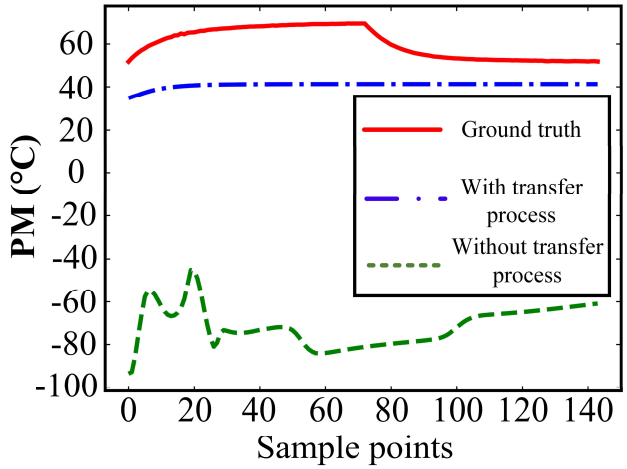
Fig. 4-9(a) presents the forecasted CP-labels by using RFR. The overall trend between real data and forecasting data is different, but the error value is much higher in the condition of PM cool down. Fig. 4-9(b) shows there are significant errors and deviations between the forecasted data and actual data trends. In Fig. 4-9(c), the overall forecasted trend is roughly similar. As Fig. 4-9(d) shown, the trend error and absolute error are both the lowest among the four subfigures. In general, it can be seen from Fig.4-9(d) that the overall trend between the CP-labels and real data remains consistent. We are confident that the forecasted results can provide valuable supervised information for the further DAAR step.

4.4.3 Results of PMT Estimation

As introduced in Section 4.3, the encoder of MT-LSTR is transferred to feature extractor of DAAR model to accelerate model training. The transfer process of both the model and parameters maximizes the training speed of the DAAR source and target regressors. To visualize the effect of the transfer process, Fig. 4-10 illustrates the performance of the regressors with the transfer process, as well as without the transfer process. Source regressor is tested on the selected data segment from source domain dataset. Target regressor is tested on PU dataset [14]. It should be noticed that the feature extractor without transfer process will be given a set of initial parameters during parameter initialization RandomUniform.



(a)

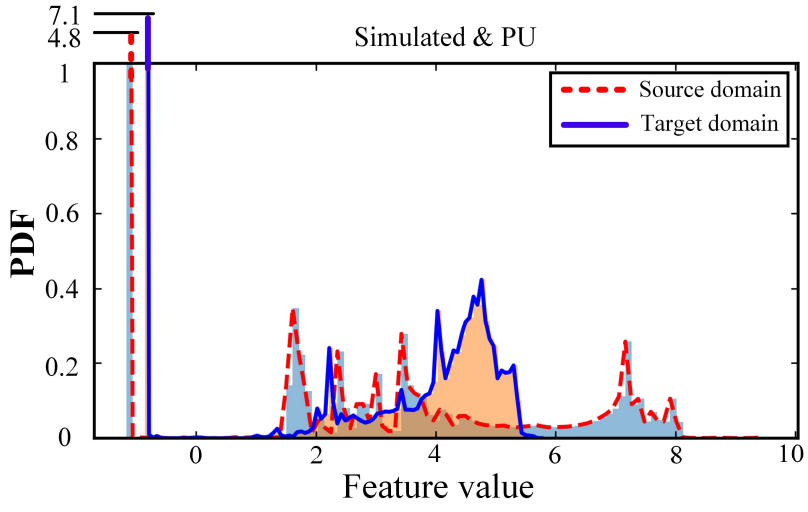


(b)

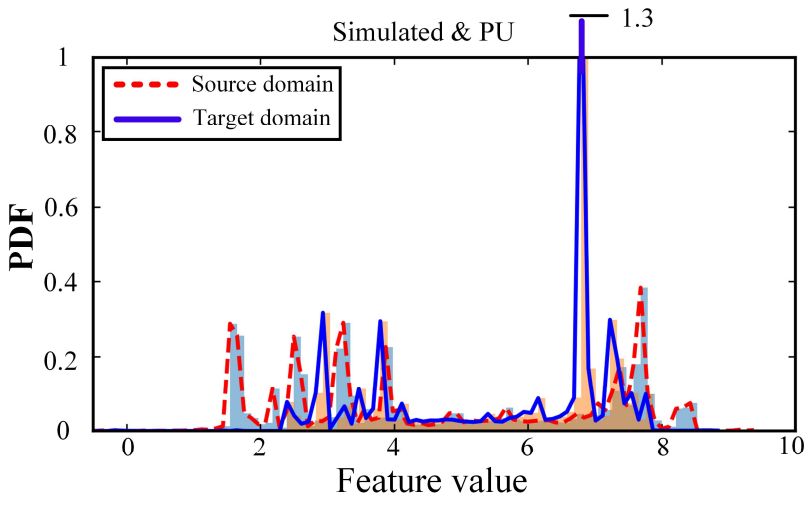
Fig. 4-10. The performance of the regressors with and without transfer process (random initialization parameters) (Before DAAR training): (a) Source regressor, (b) Target regressor.

Following the DAAR training, we provide Fig.4-11(a) to demonstrate the probability density function distribution (PDFD) of the extracted features, and Fig.4-11(b) to present a baseline model. The baseline model is a CNN model that predicts the target domain directly after training with source domain data, and thus does not undergo the process of domain adaptation transfer learning.

In order to demonstrate the effect of the DAAR model, we use the t-distributed stochastic neighbour embedding (t-SNE) dimensionality reduction technology to map the features extracted from the source domain and target domain data to the two-dimensional (2D) feature space to further observe the model effect before and after training. Fig. 4-12 shows the features of the last layer of G_{F}. After DAAR training, the data distribution of the source domain and the target domain in this space are significantly close. This means that the trained network has learned domain-invariant features and can be competent for the PMT estimation task of the target domain.

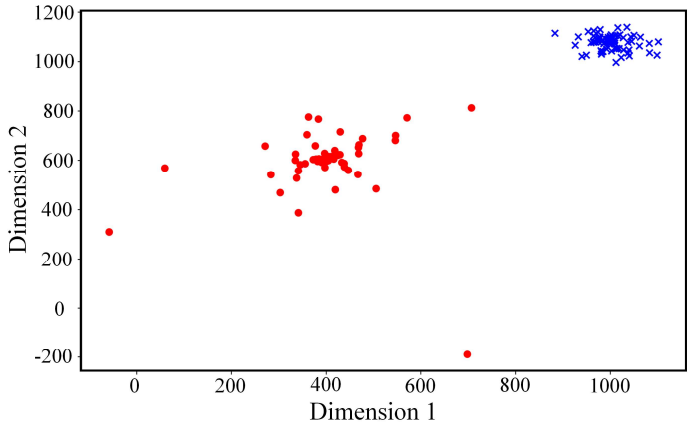


(a)

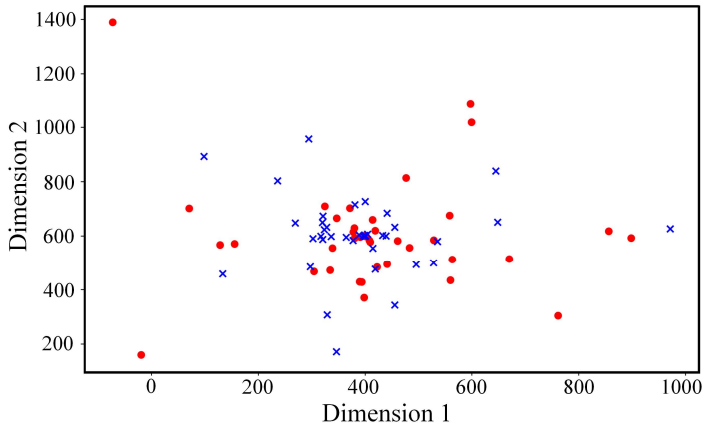


(b)

Fig. 4-11. The PDFD of source domain data and target domain data. (a) Baseline CNN, (b) DAAR(proposed).



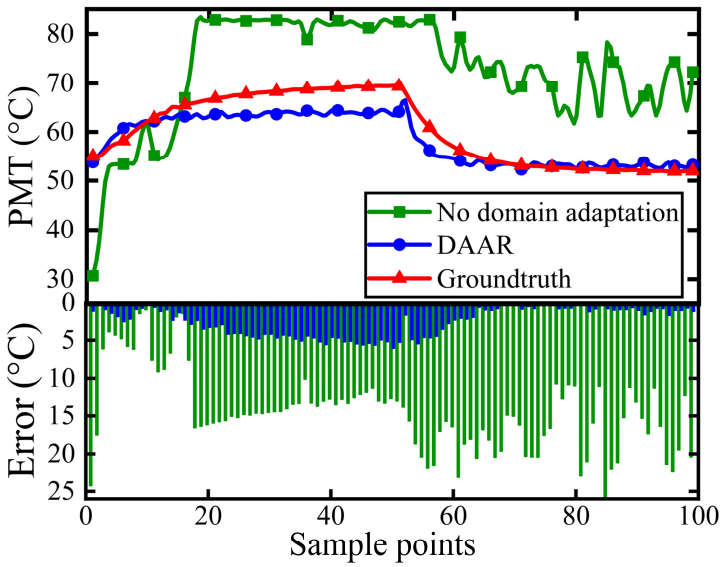
(a)



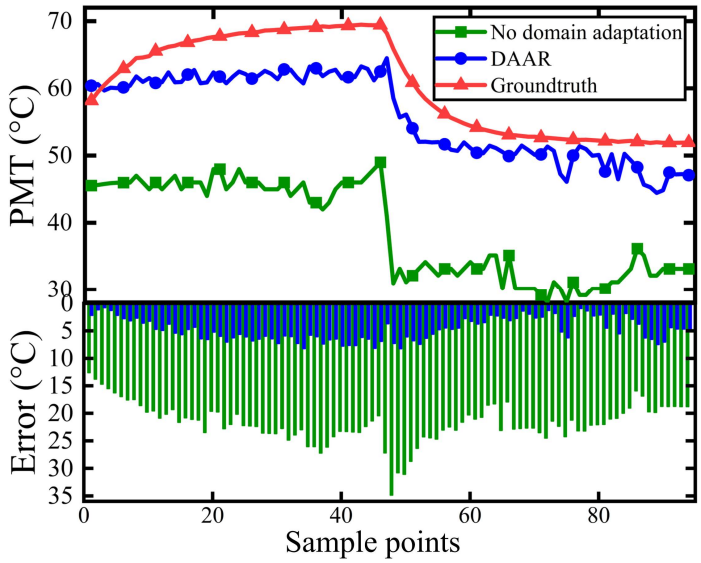
(b)

Fig. 4-12. The feature space of source domain data and target domain data (a) before DAAR training and (b) after DAAR training.

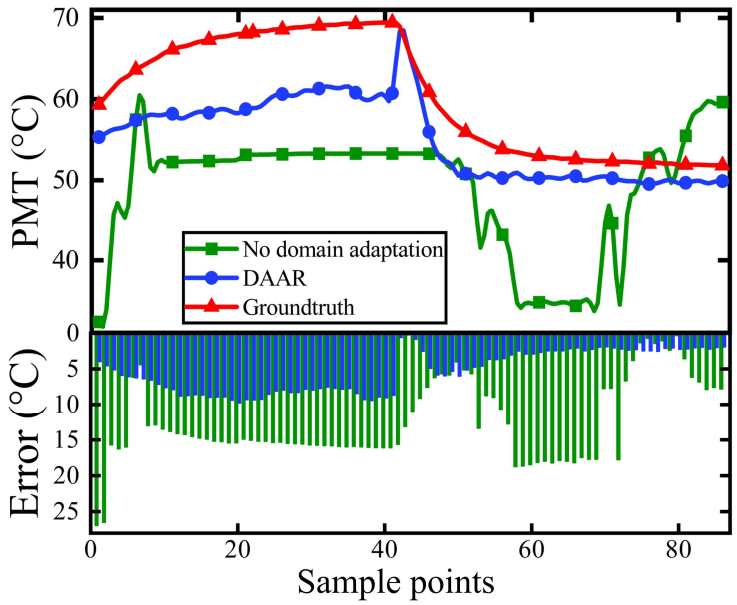
To demonstrate the effectiveness of the method in practical applications, Fig. 4-13 displays the results of PMT estimation on four specific profiles (69, 72, 73 and 76) from the PU dataset that are not included in the training data. In these drive circles, the torque ranges from -300 to 300N·m, the motor speed ranges from 0-6000rpm. As each profile contains a large number of data points, we applied a downsampling preprocessing prior to the estimation to remain 100 points for each test profile. According to the results of four profiles in Fig. 4-13, the maximum estimation error of the DAAR model is 9.79° C. The average error is 3.24° C for Profile 69, 4.13° C for Profile 72, 4.98° C for Profile 73 and 4.74° C for Profile 76. For comparison, the maximum error of the no domain adaptation CNN model is 34.86° C. The average error is 14.27° C for Profile 69, 19.58° C for Profile 72, 12.35° C for Profile 73, 20.77° C for Profile 76. These results demonstrate that the proposed method significantly improves the PMT estimation accuracy.


PMT (°C)
Error (°C)
Sample points
No domain adaptation
DAAR
Groundtruth

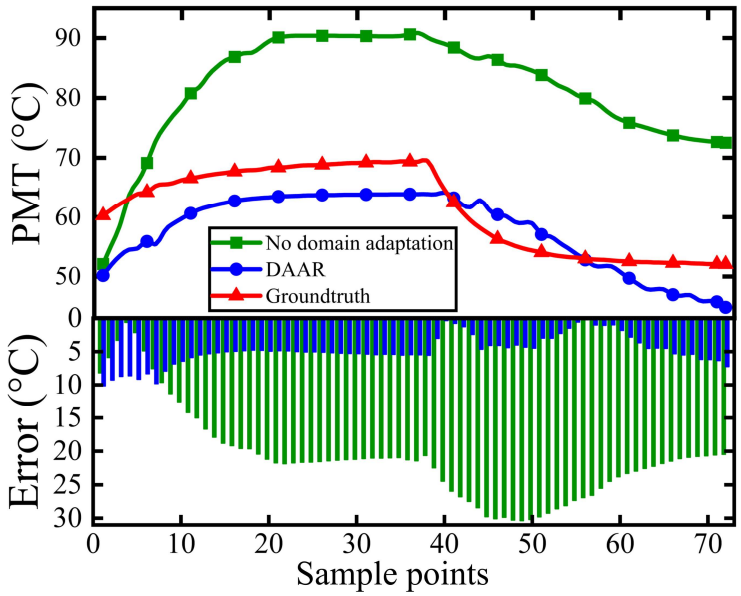
(a)


PMT (°C)
Error (°C)
Sample points
No domain adaptation
DAAR
Groundtruth

(b)


PMT (°C)
Error (°C)
Sample points
No domain adaptation
DAAR
Groundtruth

(c)


PMT (°C)
Error (°C)
Sample points
No domain adaptation
DAAR
Groundtruth

(d)

Fig. 4-13. The PMT estimation results (estimated PMT value and absolute error value) on four specific profiles from PU dataset (The torque ranges from -300 to 300N·m, the motor speed ranges from 0-6000rpm in profile 69, 72, 73 and 76).

4.4.4 Results of Comparative Experiments

In this section, comparative analysis includes two parts. In the first part, we introduce 6 popular regression algorithms to conduct a comprehensive comparison of CP-labels generation process. Each algorithm is trained using the entire simulation dataset and is required to generate CP-labels for the same piece of test data. Table 4-2 compares computing efficiency of many other methods for PMT prediction and CP-labels generation. In Table 4-2, L represents the number of training examples, k represents the number of decision trees, m donates the number of features and d is the depth of decision tree. As shown in Table 4-2, the offline training time of the MT-LSTR is slightly longer than classic machine learning and deep learning algorithms and the real-time inference time is also slightly longer than theirs. However, compared with the classic Transformer algorithm, the performance of the MT-LSTR is better. Although, compared with popular machine learning algorithms, the inference time and fitting time are longer, but can produce better results. It is cost-effective to exchange the offline learning time for the prediction performance.

TABLE 4-2 Model Computing Efficiency Comparison in Training of CP-Label Generating Process

Model	Time complexity	Memory complexity	Offline training time (s)	Real-time CP-labels generating (s)
SVR [209]	$\mathcal{O}(L^3)$	$\mathcal{O}(L^2)$	6.97×10^3	4.11×10^{-1}
RFR [200]	$\mathcal{O}(kmL \log L)$	$\mathcal{O}(kd)$	6.53×10^3	3.64×10^{-1}
LSTM [210]	$\mathcal{O}(L)$	$\mathcal{O}(L)$	1.22×10^4	7.58×10^{-1}
Transformer	$\mathcal{O}(L^2)$	$\mathcal{O}(L^2)$	2.43×10^4	8.91×10^{-1}

MT-LSTR (proposed)	$\mathcal{O}(L\log L)$	$\mathcal{O}(L\log L)$	1.83×10^{4}	8.78×10^{-1}

In the following section, we totally summaries 9 other artificial intelligence-based methods for PMT estimation, which are evaluated in terms of estimation accuracy (introduced in Table 4-3). Method 1 utilizes SVR model to extract features from signals [195]. Method 2 and 4 use a LSTM layer of 16 neurons added with the output of a 16-neuron FNN layer applied on the current time step as residual connection and two convolution blocks with a FNN, respectively [194]. Method 3 uses FNN to obtain past temperature and operating conditions to predict temperature [199]. Method 5 uses the MT-LSTR model proposed in this paper. Since there is no transfer learning-based related literature for PMSM PMT estimation, we design Method 6-8 to compare the performance of the same model (in method 2-4) before and after adding the transfer learning framework.

TABLE 4-3 Details of 10 PMT Estimation Methods for Comparation

Method	Type of method	Implementation
1 [195]	Machine learning without transfer learning	SVR
2 [194]	Deep learning without transfer learning	LSTM
3 [199]		FNN
4 [194]		TCN
5 (ours)		MT-LSTR in Section 4-3
6	Deep model with unsupervised transfer learning	LSTM with domain adaptation
7		FNN with domain adaptation
8		TCN with domain adaptation

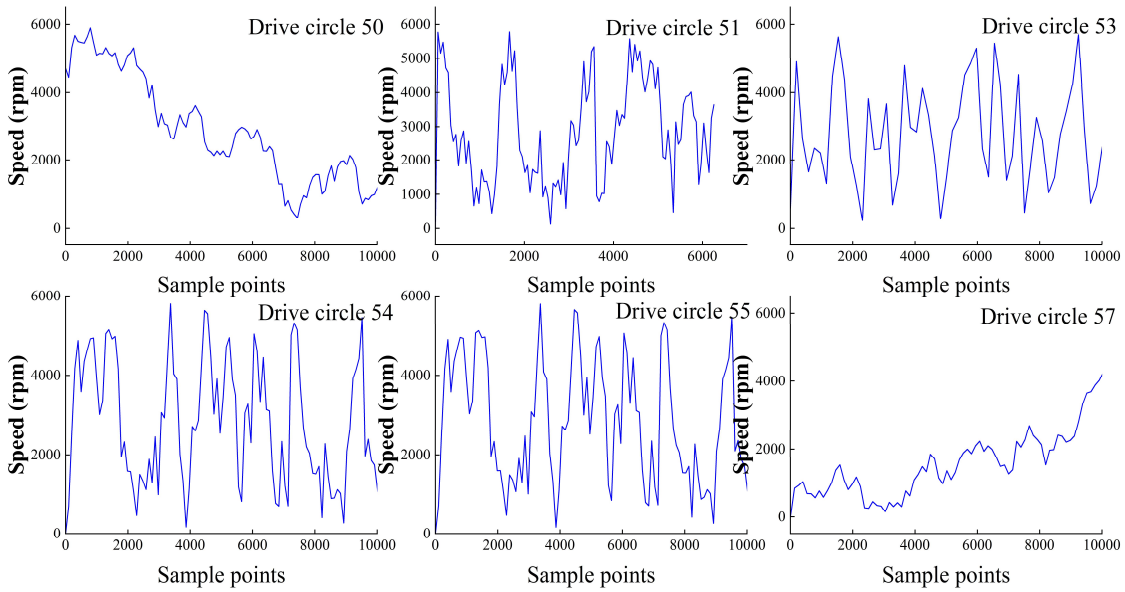
Proposed	Deep model with semi-supervised transfer learning	MT-LSTR with DAAR

However, the method without transfer learning (Methods 1-4) cannot successfully learn knowledge from target domain. Method 5 is the part 2 of the proposed approach. For the transfer learning methods with domain adaptation (Methods 6-8), these methods can obtain target domain knowledge through classic unsupervised domain adaptation [204]. We select 6 drive circles from the PU dataset as the test dataset to evaluate performance under complicated working conditions. Due to the high sampling frequency of the dataset, we again applied a downsampling preprocessing prior to the estimation to remain 100 points for each test circle. Table 4-4 presents MAE and RMSE values of these methods. Fig. 4-14(a)-(b) show the detailed change of motor speed and torque under 6 working conditions; Fig. 4-14(c)-(h) demonstrate PMT estimation results of methods introduced in Table 4-4.

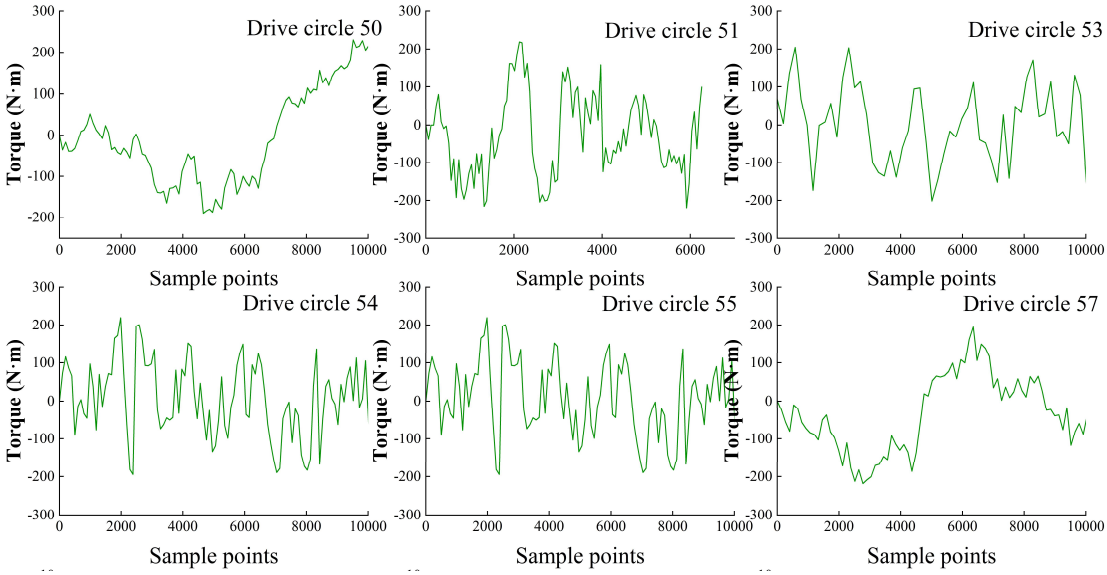
TABLE 4-4 Performance of Methods for Comparison Under 6 Complex Working Conditions

Drive circle id	50		51		53		54		55		57	
Indexes	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	44.587	50.391	115.109	117.091	46.138	46.912	56.927	58.092	80.362	81.097	107.455	110.471
2	26.446	29.745	99.529	103.962	30.477	32.443	109.197	110.295	36.021	38.334	17.357	23.642
3	55.932	60.242	46.333	47.886	34.312	38.218	32.510	40.535	41.562	33.812	82.296	83.125
4	45.405	48.365	38.216	40.929	46.465	47.101	62.732	66.989	45.159	46.421	112.007	124.205
5	25.466	29.843	29.014	29.805	9.695	12.524	70.515	76.223	51.758	49.396	25.928	28.021
6	11.488	14.086	14.995	18.994	6.114	8.434	12.232	13.874	14.218	16.802	16.078	20.848

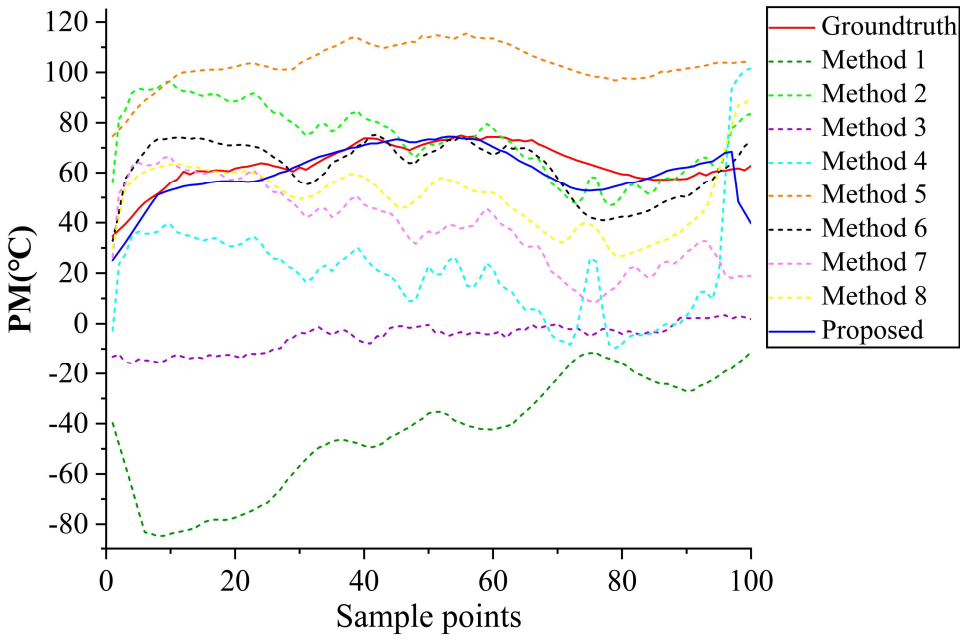
7	12.896	15.786	17.979	21.143	6.512	8.675	15.463	12.111	17.864	20.963	47.061	65.339
8	23.968	26.962	9.813	12.798	11.956	14.372	16.470	20.190	24.067	25.538	14.924	18.372
Proposed	**4.111**	**6.161**	**6.464**	**9.257**	**5.82**	**7.314**	**8.656**	**10.958**	**4.299**	**6.262**	**7.030**	**8.957**



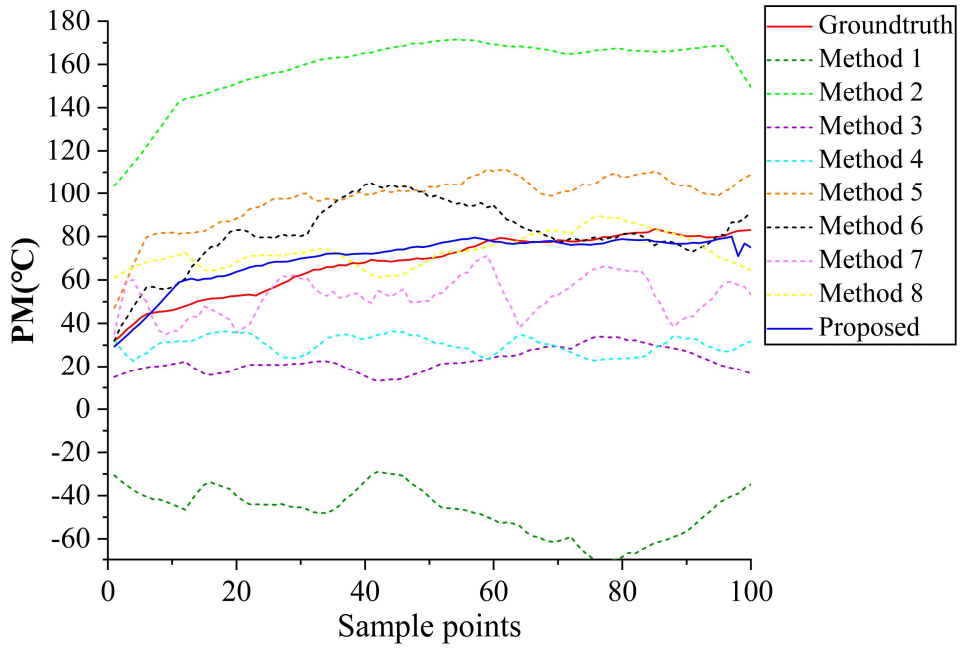
(a) Speed conditions



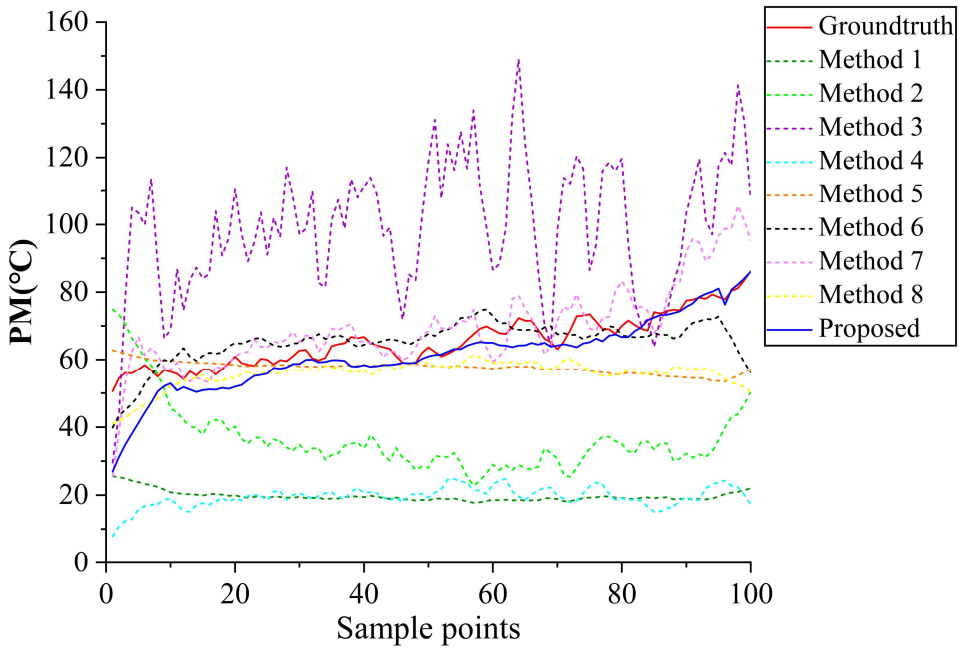
(b) Torque conditions


PM(°C)
Sample points
Groundtruth
Method 1
Method 2
Method 3
Method 4
Method 5
Method 6
Method 7
Method 8
Proposed

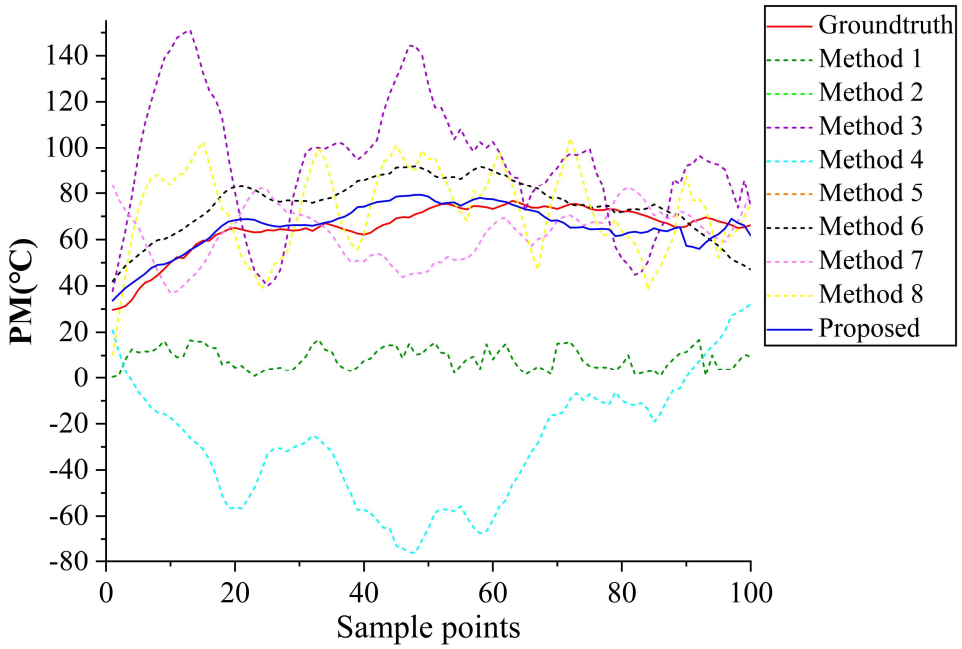
(c)


PM(°C)
Sample points
Groundtruth
Method 1
Method 2
Method 3
Method 4
Method 5
Method 6
Method 7
Method 8
Proposed

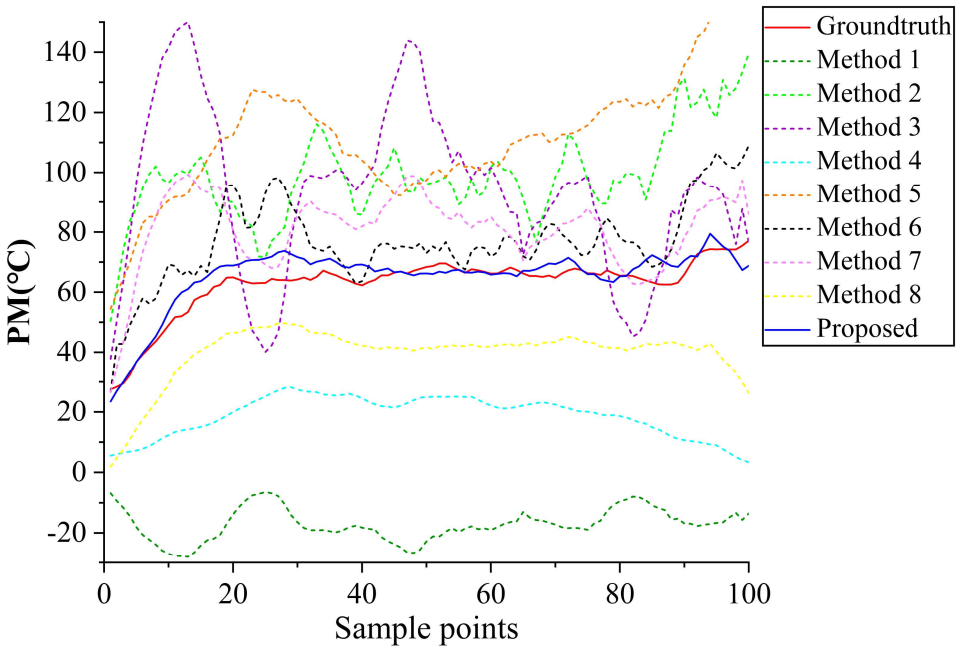
(d)


PM(°C)
Sample points
Groundtruth
Method 1
Method 2
Method 3
Method 4
Method 5
Method 6
Method 7
Method 8
Proposed

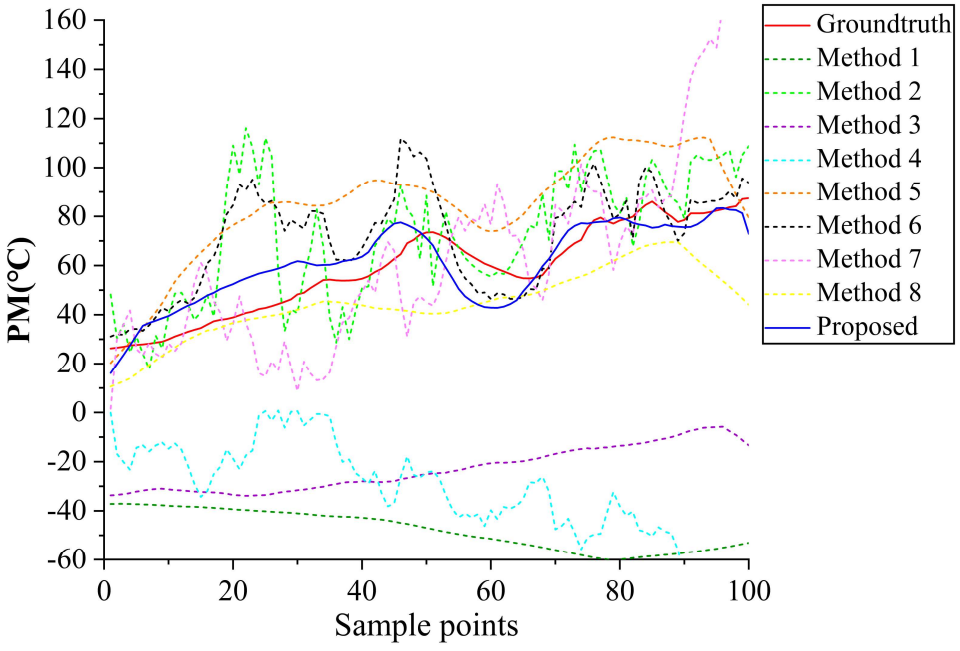
(e)


PM(°C)
Sample points
Groundtruth
Method 1
Method 2
Method 3
Method 4
Method 5
Method 6
Method 7
Method 8
Proposed

(f)



(g)



(h)

Fig. 4-14. The PMT estimation results of the 8 methods on 6 test drive circles with their detailed working conditions. ((a) The speed change of 6 drive circles (b) The torque

change of 6 drive circles (c) Drive circle 50 (d) Drive circle 51 (e) Drive circle 53 (f) Drive circle 54 (g) Drive circle 55 (h) Drive circle 57).

4.4.5 Discussion

a) Error analysis

Firstly, the training set in the source domain is obtained by simulation. This training dataset should contain samples should be collected under various working conditions as much as possible. Secondly, there are certain differences between the simulation model and the real PMSM, making the trained MT-LSTR model better suited for simulation data. In practical applications, all aspects of the PMSM system cannot be exactly the same, so it is necessary to generalize the estimation model from the source domain to the target domain through a method based on domain adaptation. Therefore, in this paper, we consider a simulated motor of the same type but with slight differences compared to the experimental motor, to reflect the proposed domain adaptation model can be used in cases where two domain datasets have certain differences. Finally, the limitations of semi-supervised transfer learning result in DAAR model that can only adapt to the target domain as efficiently as possible, thus affecting accuracy.

b) Estimation accuracy optimization and improvement

The training set of the MT-LSTR should encompass as many samples as possible and generalize as many operating conditions as possible. Such training will prevent overfitting, resulting in a model with enhanced generalization capabilities.

If conditions permit, the parameters, operating system and control method of the simulated motor and the experimental motor should strive to align as closely as possible. When targeting specific PMSMs in practical applications, enhancing the quality of source domain data will effectively improve the accuracy of the final estimate. For example, source domain data is acquired directly from a PMSM model developed through multi-physics finite element analysis (FEA) using tools like Motor-CAD or ANSYS to enhance authenticity.

4.5 Summary

This chapter proposes a novel DAAR model with an MT-LSTR CP-labels generator for precisely estimating PMT of IPMSM. Through the above experimental analysis, the following conclusions can be drawn:

1) The proposed method can effectively reduce the PDF discrepancy between source and target domains. During experimentation, the offline training time of MT-LSTR is 1.83×10^4 seconds, and real-time CP-labels generating time is 8.78×10^{-1} seconds.
2) By verifying on 4 PU PMSM subdatasets, the DAAR model demonstrated a maximum estimation error of 9.79° C, and average estimation errors of 3.24° C, 4.13° C, 4.98° C and 4.74° C across four different test profiles.
3) In comparative experiments on 6 complex drive circles of PU dataset, the minimum estimation MAE and RMSE are 4.111 and 6.161 while the maximum MAE and RMSE are 8.656 and 10.958.

Since this chapter focuses on IPMSM, we will attempt to acquire source domain data directly from the surface-mounted PMSM (SPMSM). The performance of this approach on the SPMSM will be investigated in future work.

Chapter 5 IGBT Module DPT Efficiency Enhancement Via Multimodal Fusion Networks and Graph Convolution Networks

IGBT dynamic electrical characteristics are of great significance in practical application and are usually obtained through double pulse test (DPT). However, DPTs of IGBTs under various working conditions is time-consuming and laborious. Traditional estimation methods are based on detailed physical parameters and complex formula calculations, making deployment difficult. This paper proposes a novel DPT efficiency enhancement method based on graph convolution network (GCN) and feature fusion technology, which can estimate and supplement switching transient waveforms of all working conditions. Thereby, dynamic electrical characteristics of the IGBT are obtained by estimated waveforms of DPT. This chapter proposes a multimodal attention fusion network (MAFN) to capture and fuse the features of switching transient waveforms between different positions thereby improving the expressive power and performance of the model. On the other hand, this method first proposes to utilize GCN to embed DPT data under multiple working conditions into graph structure, which can use the graph structure information to fuse the features of spatially correlated working conditions data to obtain reliable estimation result. The method is verified to be effective and accurate on real dataset collected on two batches of IGBTs.

5.1 Introduction

Insulated-gate bipolar transistors (IGBTs) are now widely used in a variety of medium-power or high-power PE converters [211], such as full-bridge inverters [212], [213] and boost converters [214]. IGBT is one of the key components in the PE system, and its performance directly affects the efficiency, response time, switching frequency, power density, reliability and stability of the system. Therefore, design engineers, researchers and manufacturers of PE systems need to fully verify the dynamic behavior of the IGBT and estimate its power loss to

judge whether the corresponding requirements are satisfied. However, the parameters given in the specification are measured under specific conditions. The external parameters in practical applications are personalized and often vary, so some of these parameters cannot be used directly. As shown in Fig. 5-1, any wrong choice or improper use may directly or indirectly cause the failure of the IGBT, resulting in serious consequences. To evaluate the behavior of the switch of the IGBT and observe parameters, the most effective method is the double pulse test (DPT).

The DPT method is the standard method for characterizing and evaluating the dynamic electrical performance of IGBTs, which is achieved with an inductive load and a power supply [215]. The inductor is used to replicate circuit conditions in a converter design. The power supply is used to provide voltage to the inductor. An arbitrary function generator is used to output pulses that triggers the gate of the IGBT and thus turns it on to start conduction of current. The parameters describing the behavior of the switch of the IGBT mainly include the turn on delay time $t$don, the current rise time $t$ri, the turn off delay time t_{doff}, the current fall time t_{fi}, the turn on loss E_{on} and the turn on loss E_{off} [216]. Due to some limitations of DPT in practical application, some estimation methods are proposed to analyze and evaluate the electrical characteristics of IGBT.

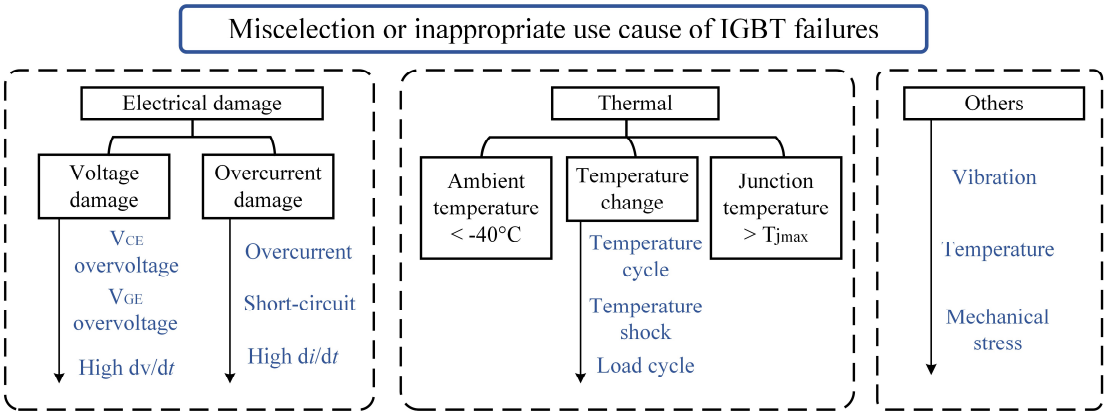


Fig. 5-1. Significance of IGBT DPT indicated by cause of failures.

Conventionally, there are three commonly used methods to obtain the static and dynamic characteristics of IGBT: Circuit simulation [217], [218], behavioral model [219] and electrothermal model [220], [221] and [222]. Device-level circuit simulation is based on IGBT physical model such as Kraus's model [223], Hefner's [224] and Wang's [217]. For example,

the simulation program with integrated circuit emphasis (SPICE) is a popular IGBT circuit simulation tool that can improve the accuracy of the simulation by using more physics knowledge of devices [225], [226]. However, these methods always require detailed physical information about the IGBT model and circuit topology. In addition, these methods bring computational burden while pursuing higher accuracy. The features of the behavioral model are ignoring device physics to a certain extent, more convenience, and fast simulation speed [227]. Recent paper [219] proposes an FPGA-based IGBT behavioral model to simulate IGBT fast transients in PE circuits without any iterative solving algorithm. However, this type of method cannot represent the switching transients without information of parasitic parameters and diode reverse recovery. Although the electrothermal model can use electrical and thermal couplings to solve the heat-flow calculation and consider the temperature effect, adding multi-dimensional thermal fields and considering packaging characteristics in the simulating estimation process will greatly increase the complexity of the model.

For accurate estimation of dynamic characteristics of IGBTs, researchers prefer to directly perform curve fitting on DPT results under specific working conditions, and then estimate switching transient waveforms based on measurement results and data sheets. For example, the electrical transient model (ETM) has proposed to simulate static and dynamic behaviors of IGBT-diode switching cell in order to determine semiconductor losses by using specially developed algebraic equations [228], [229]. However, this type of method ignores the effect of temperature on the device and relies heavily on mathematical formulas and parameter extraction. Until recently, [215] added temperature as a parameter to the original ETM to determine the losses of the IGBT diode during the simulation. However, this method can still not get rid of the dependence on many physical parameters. These characteristics of related methods are summarized in Table 5-1.

TABLE 5-1 Superiority and Weakness of IGBT DPT Estimation Methods

Type of methods	Related papers	Superiorities	Weaknesses
Circuit simulation	[217], [218]	Relatively high accuracy Strong interpretability	Relying on physics information of devices High computational difficulty

Behavioral model	[219]	Ignoring device physics More convenient Fast simulation speed	Relying on parasitic parameters and diode reverse recovery
Electrothermal model	[220], [221], [222]	Considering temperature effect	High complexity High computing requirement
ETM	[228], [229]	Accuracy limitation Easy access to parameters	Involving complex mathematical formulas
Deep learning-based	Ours	Ignoring device physics High accuracy Transferable model for other devices Fast getting estimation result	Requiring training data

With the development of artificial intelligence techniques, deep learning-based algorithms are increasingly used in industrial applications. Essentially, measured DPT waveforms are various time-series signals, which have correlations with input parameters such as gate voltage, temperature, and load current.

In this article, GCN with multimodal attention fusion network (MAFN), referred to as MAFGCN, are proposed to estimate IGBT DPT waveforms. A novel MAFN is first proposed to capture the relationship between different positions in input signals, thereby improving the expression ability of the fusion feature. The GCN-based module is used to estimate switching transient waveforms of the IGBT diode, obtain the dynamic behavior of the switching unit under various working conditions, and determine the switching loss of the IGBT at the same time. The highlight of this method is that through DPT data of a small number of working conditions, DPT results of all working conditions can be quickly estimated, and the model can be extended to other types of IGBTs. The number of required DPTs can be greatly reduced, saving labor costs and time. The complete algorithm and model are implemented in Python and validated with experimental data obtained from a DPT rig.

There are three technical contributions in this paper:

1) This chapter proposes a GCN-based approach to obtain IGBT dynamic characteristics through estimating DPT waveforms, which embeds DPT data under each working conditions into spatial graph structure.
2) MAFN is proposed to provide strong impressive features to GCN module so that graph convolutional layers can capture core feature of DPT waveforms successfully.
3) For generalizing the well-trained model to other devices, a specific transfer strategy is introduced.

The rest of the sections structured as following. Section 5.2 presents the applicability analysis of graph structure embedding in DPT. Section 5.3 detailly describes the proposed method. Section 5.4 introduces the designed experiment platform and data measurement. Experiment validations and results are reported in Section 5.5. Finally, Section 5.6 concludes this chapter.

5.2 Applicability Analysis of Graph Structure Embedding of DPT Data

5.2.1 Designed DPT Procedure

In our DPT experiment, the upper device is the freewheeling diode (FWD), the lower-side IGBT is the device under test (DUT). The diagram of the applied DPT circuit is shown in Fig. 5-2. A DPT is a tool that enables a power switch to be turned on and off at different current levels as also shown in Fig.5-2. By adjusting the switching times T1, T2 and T3, the turn-on and turn-off waveforms of DUT can be controlled and measured over the full range of operating conditions. For the second pulse, it is important to build up current in the complementary device or diode so that when the switch turns on, the effects of any reverse recovery current can be evaluated. A load inductor limits the rate of di/dt.

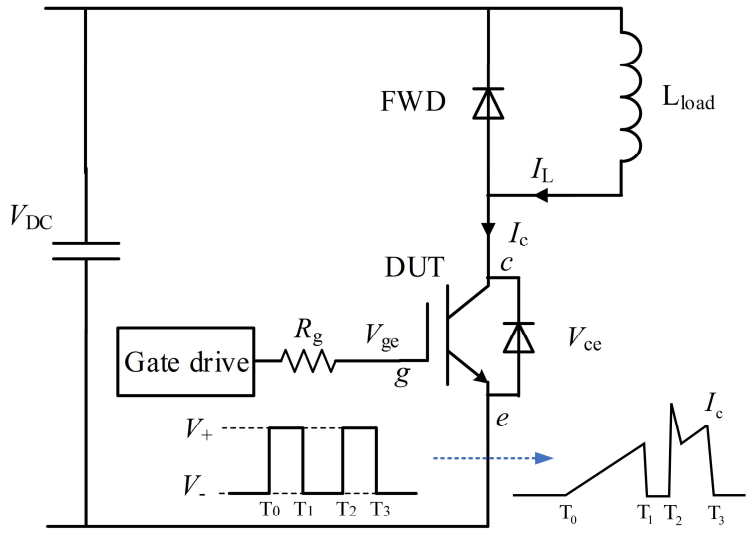


Fig. 5-2. The circuit diagram of DPT for deploying the experimental platform.

Two consecutive pulses generated by the pulse function generator are transmitted to DUT. By adjusting the first pulse duration, the switching transient can be captured at the desired current level at the end of the first pulse and at the beginning of the second pulse. VDC is the dc-link voltage, Lload is the load inductance, Vge is the gate voltage, Vce is the collector voltage, Ic is the collector current, and I_L is the load current. In the gate drive, the positive voltage V+ =15V, the negative voltage V- = -10V, the gate resistance is Rg=2.5Ω.

During T_0 to T_1, the first pulse is applied to IGBT and the IGBT is turned on. A constant voltage VDC is added to the load inductance L_{load} which makes the current through it increases in linear:

$$I_c = \frac{V_{DC} \cdot T}{L_{load}} \tag{5-1}$$

According to equation (5-1), IL which is also Ic of IGBT, depends on VDC, L_{load} and pulse duration time T. It is feasible to set the current by the method of controlling T with fixed VDC and L_{load}.

During T_1 to T_2, IGBT is turned off, but the current flowing through L_{load} cannot be suddenly changed. So, the current will commutate to FWD of counterpart arm. The duration time from T_2 to T_3 of turning-off is not of importance during test condition. Therefore, in this experiment,

we set it between 10μs and 20μs to meet the experimental requirements. After the reverse recovery process, the current flow into IGBT through L_{load} again like period T_0 to T_1. At T_3, the IGBT will be turned off again. Based on the P_{on} and P_{off}, the switching loss E_{on} in turn-on transient and E_{off} in turn-off transient are calculated by oscilloscope as:

$$E_{on/off} = P_{on/off}(t_2) - P_{on/off}(t_1) = \int_{t_1}^{t_2} V_{ce} I_c dt \tag{5-2}$$

From switching transient waveforms generated by DPT, we can also get information of IGBTs about turn-on/turn-off delay ($t_{d(on)}/t_{d(off)}$), rise time (t_r), fall time (t_f), t_{on}/t_{off} (turn-on/turn-off time), dv/dt, and di/dt, etc.

5.2.2 External Influencing and Limitation Factors

When using the data-driven method, we hope that the changes in the data are only related to the working conditions. In this case, the model obtained after training is more accurate and general. In this part, the main external influencing factors of IGBT DPT switching transient waveforms are discussed.

The first influencing factor is the measurement error. In fact, this is unavoidable in experiments due to the limitation of equipment precision, uncertainty and random error, etc.

Secondly, using different DC busbars (VDC) also affects the switching transient voltage waveform. Different structures of the DC bus can lead to changes in the parasitic inductance L_S. At a high current change rate di/dt, the induced potential of L_S is superimposed on the IGBT, which will affect the waveform of V_{ce}:

$$V_{ce} = -L_S \frac{di}{dt} + V_{DC} \tag{5-3}$$

That is manifested that high L_S at the IGBT turn-on transient will result in low E_{on}, and the turn-on time will be extended; too high L_S at the turn-off transient will cause V_{ce} to overshoot the rated voltage of the IGBT, cause overvoltage breakdown fault, and reduce the tail current.

Furthermore, different FWD of the circuit will affect turn-on transient overcurrent peak I_{peak} and E_{on}. During the turn-on transient of the IGBT, the reverse recovery characteristic (RRC)

of FWD can lead to I_{peak} in the collector current, which affects the turn-on rate and turn-on loss. In addition to being affected by di/dt, RRC is also related to the stored charge Q_{rr} of the diode. I_{peak} can be expressed as

$$I_{\mathrm{peak}} = I_{\mathrm{L}} + \sqrt{2\frac{di}{dt}\frac{Q_{\mathrm{rr}}}{SF}} \tag{5-4}$$

where SF is the reverse recovery softness factor of FWD [230]. Q_{rr} is determined by the forward conduction current I_{L} before reverse recovery and the remaining carrier lifetime of the diode itself τ_{H}:

$$Q_{\mathrm{rr}} \approx I_{\mathrm{L}}\tau_{\mathrm{H}} \tag{5-5}$$

Under a fixed I_{L}, different Q_{rr} of the diode will affect I_{peak}, resulting in a change in E_{on}.

However, we should notice that the same users or same manufacturers usually tend to use the same DC busbars and FWD. Therefore, external influencing factors have little impact on the stability and accuracy of the data-driven method proposed in this paper. In other words, the proposed model is personalized for a specific practical application from the training step.

5.2.3 Graph Structure Embedding for DPT Estimation Problem

This paper proposes to embed DPT data for all working conditions in a labeled graph of GCN. GCN belongs to the category of deep learning, which learns and trains based on graph structure data. GCNs enable prediction and inference tasks at both the node level and the graph level by effectively capturing the relationship between the topological structure of graph data and node features [231].

It can be seen from the experiment that the DPT transient waveforms under various working conditions are correlated, and the degree of correlation is different. We propose to treat each working condition of DPT as a node of the graph structure. The DPT switching transient waveforms of unmeasured nodes could be accurately estimated through the correlation between each node, so as to obtain the complete IGBT dynamic electrical characteristics. Fig. 5-3 demonstrates the principle and strategy of graph structure embedding for DPT estimation

problem.

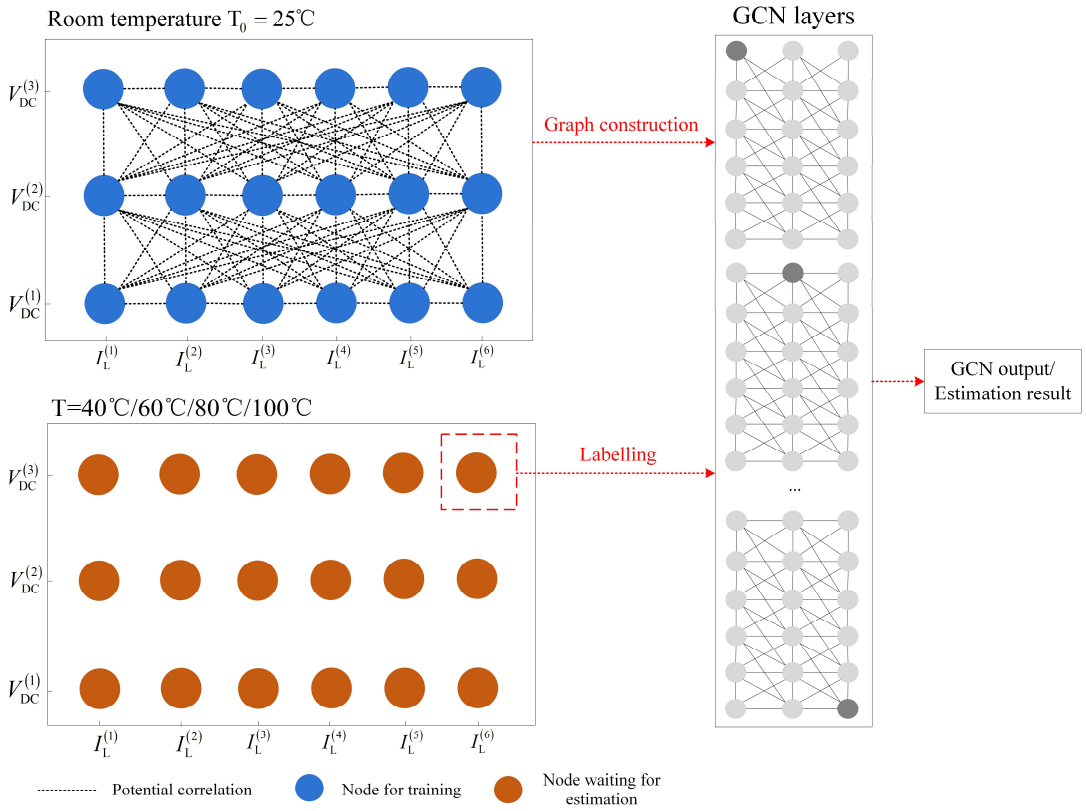


Fig. 5-3. The principle of graph structure embedding for DPT estimation problem.

In DPT data measurement, $V_{\mathrm{DC}} = \left\{V_{\mathrm{DC}}^{(1)}, V_{\mathrm{DC}}^{(2)}, V_{\mathrm{DC}}^{(3)}\right\}$, $I_{\mathrm{L}} = \left\{I_{\mathrm{L}}^{(1)}, I_{\mathrm{L}}^{(2)}, I_{\mathrm{L}}^{(3)}, I_{\mathrm{L}}^{(4)}, I_{\mathrm{L}}^{(5)}, I_{\mathrm{L}}^{(6)}\right\}$. Therefore, total measured working conditions $C_{\mathrm{w}} = \left\{\left(V_{\mathrm{DC}}^{(i)}, I_{\mathrm{L}}^{(j)} \mid 1 \le i \le 3, 1 \le j \le 6\right)\right\}$, in which $i, j \in \mathbb{N}^{+}$. As shown in Fig. 5-3, the number of C_{w} is 18. The DPT dataset measured under $\mathrm{T_c^{(0)}}$ is the training data, and the DPT dataset measured under case temperature $\mathrm{T_c} = 40℃, 60℃, 80℃, 100℃$ are labels. That means the well-trained model can estimate DPT waveforms under any $\mathrm{T_c}$.

5.3 The Proposed Methodology

The proposed method MAFGCN consists of three main parts: graph construction module, MAFN block, GCN-based module and optional graph transfer module: 1) In order to capture

more comprehensive DPT waveform information and enhance the overall performance of the model, the input data needs to be input to the proposed MAFN for feature extraction and fusion. The overall input data is denoted by X; 2) SCC-based graph construction block conveys an adjacency matrix A_{ij} to GCN module; 3) Spatial GCN-based module is trained by output of the MAFN and A_{ij} to integrate the information of graph nodes to provide spatial dependence for the DPT waveform estimation; 4) The graph transfer block can generalize the model so that it can be used to predict DPT results of other batch IGBT modules and even other brand IGBT modules. Fig. 5-4 illustrates the complete architecture of the proposed method.

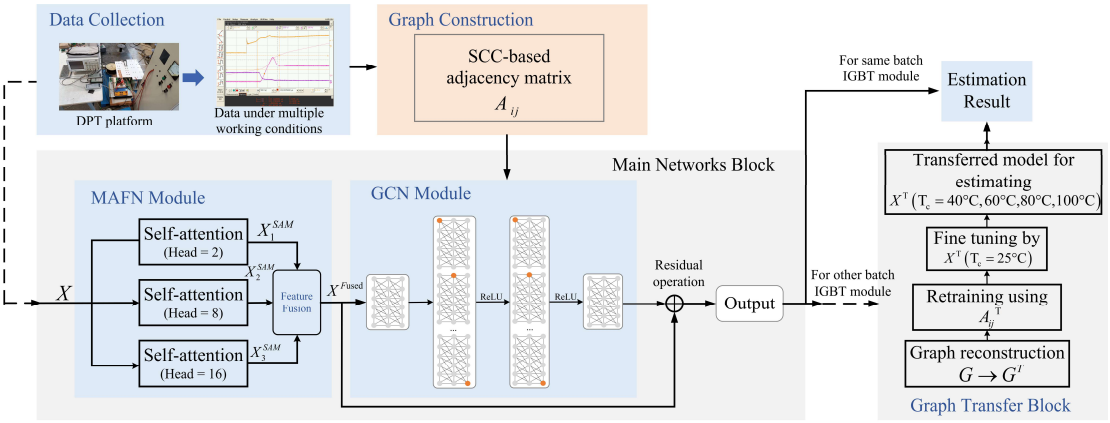


Fig. 5-4. Structure of the proposed (MAFGCN) method (the GCMCN method includes a MAFN module, a graph construction block, a GCN-based spatial module and an optional graph transfer block, in which a residual operation between the input and output of the of GCN module for improving the effect of feature fusion and preventing the gradient disappearance problem during training).

5.3.1 Data Input and MAFN Block

Appropriate input variables and their input forms should be selected from the data acquisition module to ensure the accuracy and reliability of DPT waveform estimation method. The total input data collected by this method from the DPT system is:

$$X = \left\{X_{V_{ge}}, X_{V_{ce}}, X_{I_c}, X_{P_{on}}, X_{P_{off}}\right\} \tag{5-6}$$

These input variables are the core waveforms in DPT and the final estimated target waveforms. For each training of the model, one type of data in X should be as input data to the next step.

Dataset X contains a total of 5 variables, which means that the model needs to be completely trained 5 times to obtain a complete DPT estimation result.

Attention mechanism can solve the defects that RNN cannot perform parallel computing and CNN cannot capture the long-distance relationships in sequence [232]. The designed MAFN as a multimodal feature fuser consists of three self-attention modules with different number of attention head as shown in Fig. 5-5. Fewer attention heads provide more stable and consistent feature representations, while more attention heads can capture more fine-grained relations and patterns of the input sequence. Selected dataset will firstly be input into MAFN module to get most expressive feature representations as the input of GCN module. The details are as follows.

Self-attention (SA) as a special attention mechanism. This module uses multi-head self-attention (MSA) mechanism network. Input X needs to go through positional encoding process. For the hth head, the weights of SA can be calculated by

$$\mathrm{SA}(Q_h, K_h, V_h) = \mathrm{softmax}\left(\frac{Q_h {K_h}^{\mathrm{T}}}{\sqrt{d_k}}\right) V_h \tag{5-7}$$

where $Q_h = XW_q^h$, $K_h = XW_k^h$, $V_h = XW_v^h$. W_q^h, W_k^h and W_v^h are linear transformation matrix of Q_h, K_h and V_h, respectively. Weights of SA of each head are weighted and spliced to obtain the representation of MSA:

$$\mathrm{MSA}(Q, K, V) = \mathrm{Concat}(head_1, \cdots, head_h) W^{\mathrm{O}} \tag{5-8}$$

$$head_h = \mathrm{SA}(Q_h, K_h, V_h) \tag{5-9}$$

where W^{O} donates the output weight matrix of MSA. The output of MSA (X_i^{MSA}) needs to go through a feed-forward network (FFN) as the final representation X^{SAM}. FFN includes two fully-connected (FC) layers:

$$\mathrm{FFN} = \mathrm{ReLU}(W^{\mathrm{FFN}} X_i^{\mathrm{MSA}} + b) \tag{5-10}$$

where $\mathrm{ReLU}(\cdot)$ is the activation function, W^{FFN} is weight matrix, b is the bias. Outputs of three self-attention modules X_i^{SAM} should be fused following strategy:

$$X^{\text{Fused}} = \sum_{i=1}^{3} w_i^{\text{Fusion}} X_i^{\text{SAM}} \tag{5-11}$$

where w_i^{Fusion} donates the weight of the ith self-attention module. X^{Fused}, as the most expressive feature representation of DPT sequence, will be the input of GCN-based module.

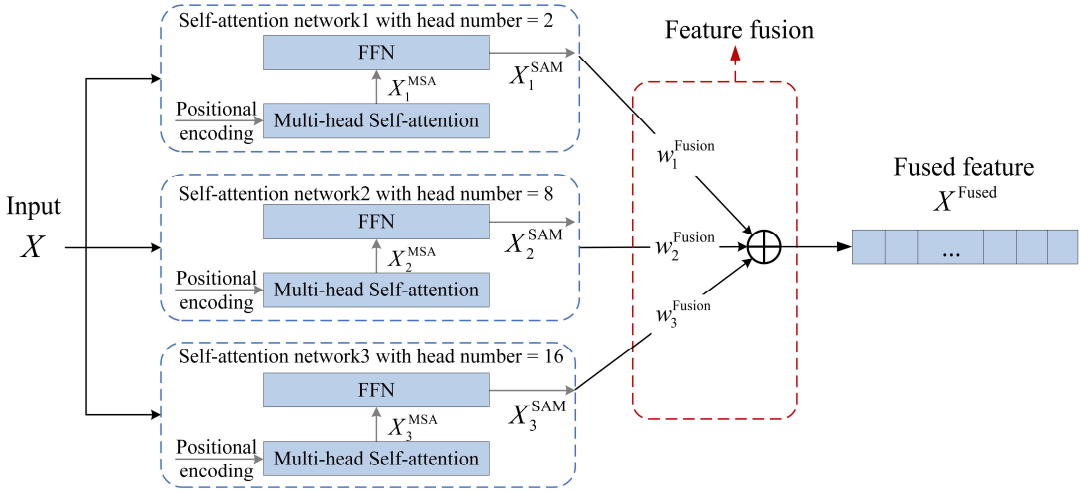


Fig. 5-5. Detailed structure of the MAFN block.

5.3.2 Graph Construction Module

As described in Section 5.2.3, the whole DPT dataset can be embedded into a graph structure with multiple nodes. In this paper, this method applies the Spearman correlation coefficient technique to calculate and visualize the spatial correlation between subdatasets under each working condition, which is constrained by linear relationships so as to does not lose nonlinearity information.

According to the training data, we should firstly calculate the SCC between each two nodes. Given two input sequences $D_1 = (x_1, x_2, \cdots, x_{l_s})$ and $D_2 = (y_1, y_2, \cdots, y_{l_s})$, the Spearman correlation coefficient ρ can be calculated by

$$\rho = 1 - \frac{6\sum_{i=1}^{l_s} d_i^2}{l_s(l_s^2 - 1)} \tag{5-12}$$

where l_s is the length of the input sequence, $d_i = R_x - R_y$ is the difference between ranks of

variables D_1 and D_2. The calculation principle of R_x and R_y can refer to [233], which will not be explained in detail here.

This method uses ρ as the similarity index between nodes to build an adjacency matrix. In GCN, the graph can be referred as $G = (P, E)$, in which $P = \{p_1, p_2, p_3, \cdots, p_N\}$ is the set of N nodes and E is the edge set. In the set E, element $e_{ij} \in E$ is the edge between node p_i and node p_j. To presents the correlations between nodes, the adjacency matrix can be identified as $A \subset \mathbb{R}^{N \times N}$, and elements a_{ij} in A can be computed by

$$a_{ij} = \begin{cases} w_{ij}, & if(p_i, p_j) \in E \\ 0, & otherwise \end{cases} \tag{5-13}$$

where $w_{ij} \in [0,1]$ is the degree of relevance between node p_i and p_j. In this method, $w_{ij} = \rho_{ij}$. Due to $\rho_{ij} \in [-1,1]$, further normalization process is not required here.

To reduce the number of edges in the labeled graph and reduce computational cost, we preserve strong correlations between nodes and eliminate weak correlations. Therefore, A_{ij} can be updated as

$$a_{ij} = \begin{cases} a_{ij}, & if\, a_{ij} < \eta \\ 0, & otherwise \end{cases} \tag{5-14}$$

where η is a preset threshold so that any a_{ij} less than η will be reset to 0. If the device has sufficient computing power, this step can be ignored.

5.3.3 Spatial GCN-based Estimation Module

Instead of applying regular convolutional and recurrent networks, this method formulates the problem on graphs and build the model with complete convolutional structures, which enable much faster training speed with fewer parameters [234].

After embedding nodes according to the strategy in Section II.C and building an adjacency matrix [235], GCN is used to integrate the information of neighbor nodes into the target node to provide spatial dependence for the DPT waveform estimation task. The calculation process of graph convolution is:

$$Z^{(k+1)} = \mathrm{ReLU}\left(L^{norm} Z^{(k)} W^{(k)}\right) \tag{5-15}$$

$$\tilde{A} = A + I_N \tag{5-16}$$

$$L^{norm} = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} \tag{5-17}$$

where $Z^{(k)}$ is the input of the $(k+1)$th graph convolution layer, $W^{(k)}$ is the layer-specific trainable weight matrix of the $(k+1)$th layer, $\mathrm{ReLU}(\cdot)$ is the activation function, L^{norm} is the normalized Laplace matrix, D is the degree matrix of node, I_N is the identity matrix. In this paper, the number of GCN layer is set as 2. The output layer is a FNN layer. The structure of the spatial GCN module is shown in Fig. 5-4.

5.3.4 Graph Transfer Block

For generalizing the model to estimate other IGBTs with large parameter differences and less data, the transfer learning strategy is proposed. However, this transfer learning operation is not necessary if testing IGBTs from the same batch or similar batches. There are two main steps of this block: graph reconstruction and fine tuning. Fig. 5-6 shows the detailed flowchart of this graph transfer block.

In the graph reconstruction step, the graph structure of the target IGBT $G^{\mathrm{T}} = (P^{\mathrm{T}}, E^{\mathrm{T}})$ should be redefined, in which $P^{\mathrm{T}} = \{p_1^{\mathrm{T}}, p_2^{\mathrm{T}}, p_3^{\mathrm{T}}, \cdots, p_N^{\mathrm{T}}\}$, E^{T} is the edge set. The adjacency matrix A_{ij}^{T} should be recalculated according to G^{T}:

$$A_{ij}^{\mathrm{T}} = \begin{cases} w_{ij}^{\mathrm{T}}, & if\left(p_i^{\mathrm{T}}, p_j^{\mathrm{T}}\right) \in E^{\mathrm{T}} \\ 0, & otherwise \end{cases} \tag{5-18}$$

where $w_{ij}^{\mathrm{T}} = \rho_{ij}^{\mathrm{T}}$ is degree parameters. To obtain a pre-trained model with G^{T}, the whole should be retrained using X. Then the retrained model is fine-tuned by the DPT dataset of the target IGBT $X^{\mathrm{T}} = \left\{X_{V_{\mathrm{ge}}}^{\mathrm{T}}, X_{V_{\mathrm{ce}}}^{\mathrm{T}}, X_{I_{\mathrm{c}}}^{\mathrm{T}}, X_{P_{\mathrm{on}}}^{\mathrm{T}}, X_{P_{\mathrm{off}}}^{\mathrm{T}}\right\}$. We should note it is allowed if the amount of data in $X^T \ll X$.

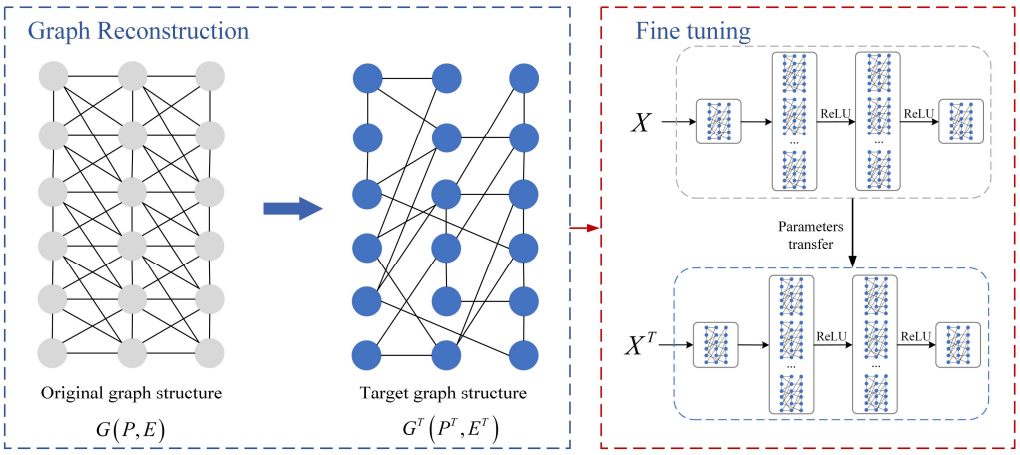


Fig. 5-6. Detailed flowchart of the graph transfer learning block.

5.3.5 Implement Procedure of MAFGCN

The implementation process of the proposed MAFGCN is divided into training stage and verification stage. The entire algorithm updates parameters with each iteration. The pseudocode of the whole procedure of MAFGCN is shown in Algorithm 1.

Algorithm 1: The proposed MAFGCN with multimodal feature fusion for DPT estimation.

Input: The DPT dataset $X = \left\{X_{V_{\mathrm{ge}}}, X_{V_{\mathrm{ce}}}, X_{I_{\mathrm{c}}}, X_P\right\}$ obtained from real DPT experiment that is divided into training set X_{train}, validating set X_{val} and testing set X_{test}. The DPT dataset from similar IGBT module $X^{\mathrm{T}} = \left\{X^{\mathrm{T}}_{V_{\mathrm{ge}}}, X^{\mathrm{T}}_{V_{\mathrm{ce}}}, X^{\mathrm{T}}_{I_{\mathrm{c}}}, X^{\mathrm{T}}_{P}\right\}$.

Output: The estimated PDT results at X_{est} and $X^{\mathrm{T}}_{\mathrm{est}}$ and estimation accuracy of testing data.

In the training and validating stages of MAFGCN:

1 Collect dataset from real IGBT modules and obtain X and X^{T}. Initialize parameters of the whole MAFGCN;

2 Calculate A_{ij} (Eq. (5-13)) through Eq. (5-12);

3 Compute and obtain updated A_{ij} through Eq. (5-14);

Training the MAFN module:

4 Input dataset X_{train} into three MSA networks to extract features X_i^{MSA};

5 Forward propagation through three MSA networks with parameters Q_h, K_h, V_h and W^{O};

6 Calculate X_i^{SAM} by using Eq. (5-10);

Forward propagation through FFNs with parameters W^{FFN} and b;

7 Calculate X^{Fused} by using Eq. (5-11);

Training the GCN module with residual operation:

8 Input MAFN result X^{Fused} into GCN;

9 Calculate GCN output by using Eq. (5-15)-(5-17);

10 Forward propagation through $Z^{(k)}$ with parameters $W^{(k)}$;

Compute the loss function $\mathcal{L}$ and update parameters of the entire model;

Backpropagation with RMSprop optimizer to minimize $\mathcal{L}$;

Update the parameters of the entire MAFGCN model;

Compute the performance indexes of the model on X_{val};

In the graph transfer block:

11 Calculate A_{ij}^{T} (Eq. (5-18)) through Eq. (5-12);

12 Retraining MAFGCN with the input $X^{\mathrm{T}}(\mathrm{T_c}^{(0)})$;

13 Fine-tune by the dataset $X^{\mathrm{T}}(\mathrm{T_c}^{(0)})$ and update the parameters of the entire MAFGCN model;

In the testing stage:

14 Evaluate the model with testing data X_{test} and X^{T} under $\mathrm{T_c}^{(1)}, \mathrm{T_c}^{(2)}, \mathrm{T_c}^{(3)}, \mathrm{T_c}^{(4)}$, and compute accuracy of estimated PDT results.

5.4 Setup of DPT Platform and Data Acquisition

This section will provide information about the setup of the DPT platform. The experimental and validation datasets are collected from this described DPT bench.

In order to collect accurate and reliable data, we have carried out a series of real DPT experiments. Based on the circuit in Fig. 5-2 in Section 5.2, a DPT platform was designed and implemented for characterizing IGBTs and diodes. As shown in Fig. 5-7, the DPT measurement device is composed of high voltage power supply, capacitor bank, laminated DC busbar, load inductances, DPT and drive circuit, measuring devices (voltage probe and current probe), oscilloscope and pulse generator.

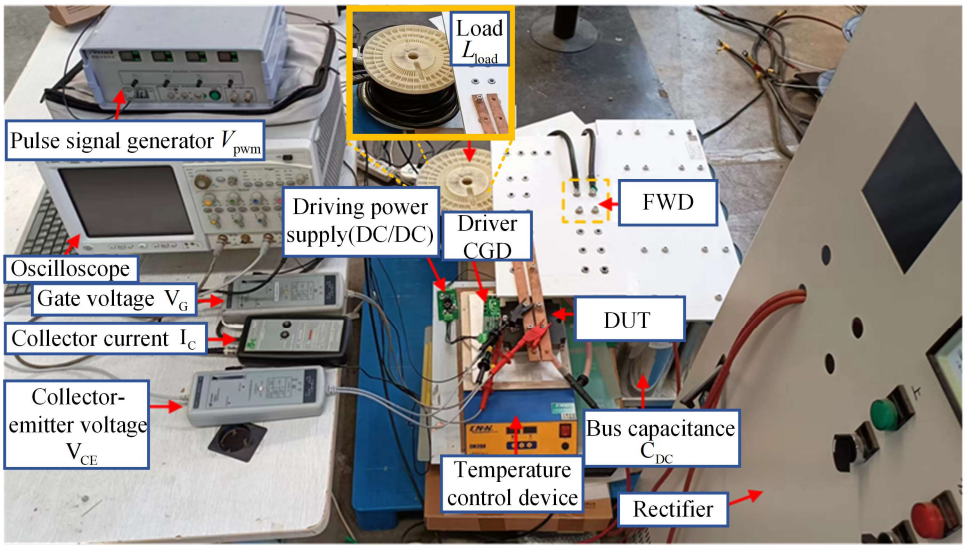


Fig. 5-7. Setup of the DPT measurement platform.

The main DUTs are DIM1200FSS12-A000 and Dynex DIM1200FSS12-A076 from Dynex semiconductor. Both of two IGBTs are single switch 1.2kV, *n*-channel enhancement mode, IGBT module. This IGBT has a wide reverse bias safe operating area (RBSOA) plus 10μs short circuit withstand. The main applications of these IGBTs are high power inverter and motor controller. The parameters of the tested IGBTs are summarized in Table 5-2.

TABLE 5-2 Parameters of IGBT Methods

Batch	DIM1200FSS12-A000	DIM1200FSS12-A076
V_{CES}	1.2kV	1.2kV
$V_{CE(sat)}$ (typ)	2.2V	2.2V
I_C (max)	1200A	1200A
$I_{C(PK)}$ (max)	2400A	2400A

The variables of test conditions are T_c, V_{DC}, I_L and turn-on/turn-off transient:

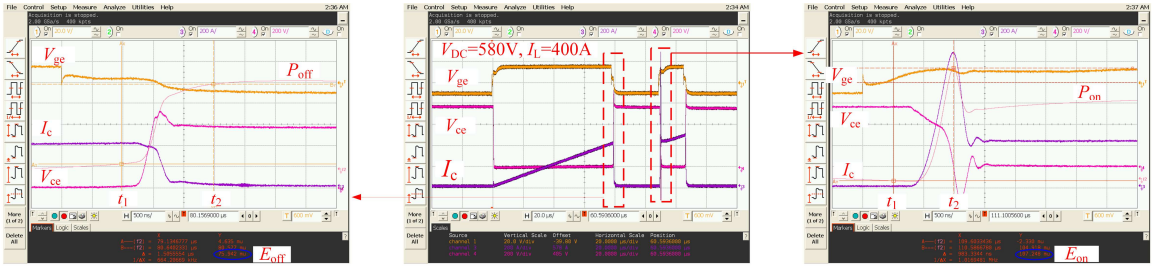
T_c: 25℃, 40℃, 60℃, 80℃ and 100℃;

V_{DC}: 200V, 400V and 600V;

I_L: 50A, 200A, 400A, 600A, 800A and 1000A;

Transient condition: turn-on, turn-off.

where the number of total test conditions is 180. Each IGBT was subjected to 10 double-pulse tests for each test condition. Measurement data of each batch includes 1800 groups, and a total of 4800 groups of data are obtained. Test time *t*, Vge, Vce, Ic, Pon and Poff are included in each group data. We apply four channels oscilloscope to observe and output waveforms of Vge, Vce, Ic, Pon and Poff. Waveforms of channels 1,3 and 4 are respectively represented as Vge, Ic and Vce as shown in Fig. 5-8. There are examples of collected waveforms on oscilloscope with the working condition of Tc=25℃, V_{DC}=580V, Ic =400A.



(a) T_c=25 °C, V_{DC}=580V, I_c =400A and in turn-off transient (b) The complete DPT waveforms at T_c=25 °C, V_{DC}=580V, I_c =400A (c) T_c=25 °C, V_{DC}=580V, I_c =400A and in turn-on transient

Fig. 5-8. Typical example waveforms of the measured and collected DPT result on oscilloscope at working condition Tc=25℃, VDC=580V, Ic =400A ((a) in turn-off transient. (b) the complete DPT waveforms (c) in turn-on transient.

5.5 Experimental Result

This section describes the experimental validations based on the proposed DPT estimation approach in Section 5.3 and the collected dataset in Section 5.4.

5.5.1 Performance Criteria

The proposed method is evaluated using three quantitative indexes: mean absolute error (MAE), root mean square error (RMSE) and Logcosh error (LE), which are represented by Eq. (5-19), Eq. (5-20), and Eq. (5-21), respectively. These evaluation metrics provide a clear indication of the accuracy of the DPT waveform estimation task. The MSE and MAE measures assess the level of deviation between the estimated waveform and the ground truth, while the CS captures the similarity between the two waveforms in terms of their shape and direction. Together, these indexes provide a comprehensive assessment of the accuracy and fidelity of the proposed method in estimating the DPT waveform.

$$\mathrm{MAE} = \frac{1}{L_\mathrm{s}} \sum_{n=1}^{L_\mathrm{s}} |y^n - \hat{y}^n| \tag{5-19}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{L_\mathrm{s}} \sum_{n=1}^{L_\mathrm{s}} (y^n - \hat{y}^n)^2} \tag{5-20}$$

$$\mathrm{LE} = \frac{1}{L_\mathrm{s}} \sum_{n=1}^{L_\mathrm{s}} \left(\frac{e^{\hat{y}^n - y^n} + e^{-(\hat{y}^n - y^n)}}{2} \right) \tag{5-21}$$

where y^n and $\hat{y}^n$ are the true value and estimated value of the DPT waveform amplitude, n is the nth sample point, and L_s is the length of DPT signals.

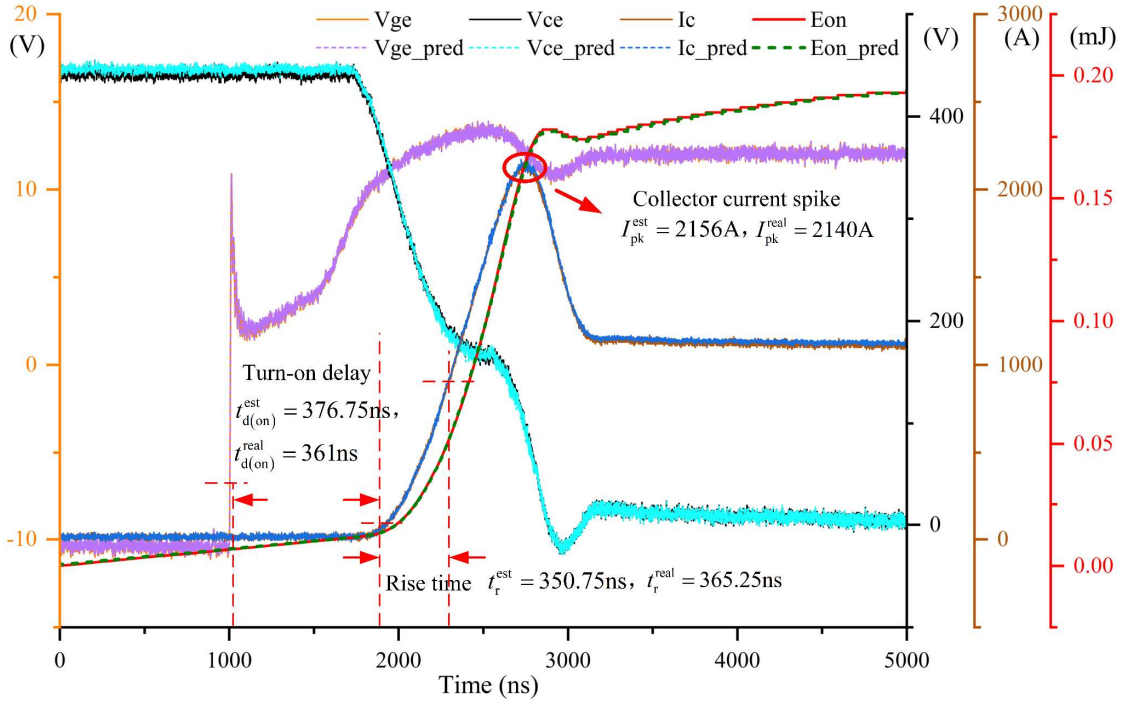
5.5.2 Experiments Settings and Result

Firstly, we should number the data collected from the two types of IGBTs in Section 5.4. The 1800 groups collected from DIM1200FSS12-A000 are referred as Dataset A; the 1800 groups collected from DIM1200FSS12-A760 are referred as Dataset B. This method can estimate DPT results under 144 different working conditions (except 36 kinds under 25℃). The use condition of these datasets is listed in Table 5-3.

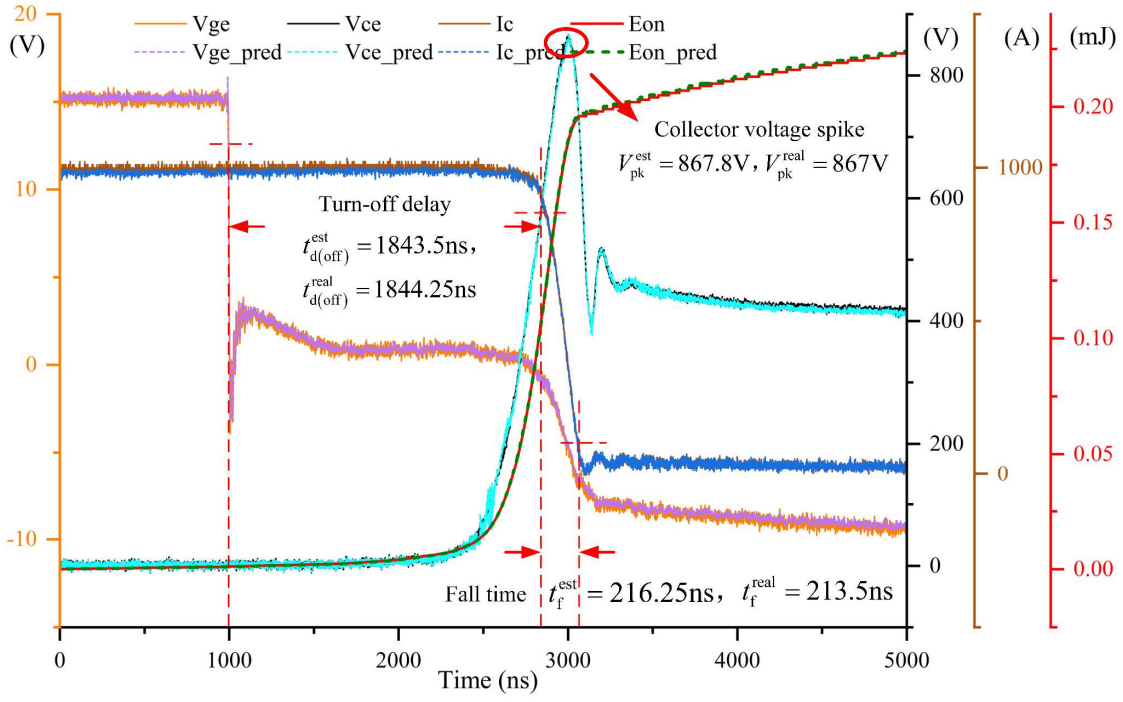
TABLE 5-3 The Use of Two Measured DPT Datasets

Dataset	Data division	Detailed use
A	90% in T_c = 25ºC, 40ºC, 60ºC, 80ºC and 100ºC	Training and validating data
	10% in T_c = 40ºC, 60ºC, 80ºC and 100ºC	Testing data
B	90% in T_c = 25 ºC	Graph fine-tuning data
	10% in T_c = 40ºC, 60ºC, 80ºC and 100ºC	Testing data

Due to limited space, turn-on and turn-off transient detailed experimental result under one working condition (Tc=100℃, VDC=400℃, Ic=1000℃) in Dataset A and one (Tc=80℃, V_{DC}=400V, Ic=600A) in Dataset B are illustrated in Fig, 5-9 and Fig, 5-10 respectively. Batch size is set as 64 and training epoch is 100. The RMSprop optimizer is chosen with initial learning rate 1×10^{-5}. All data are normalized before model training.



(a) Turn-on transient



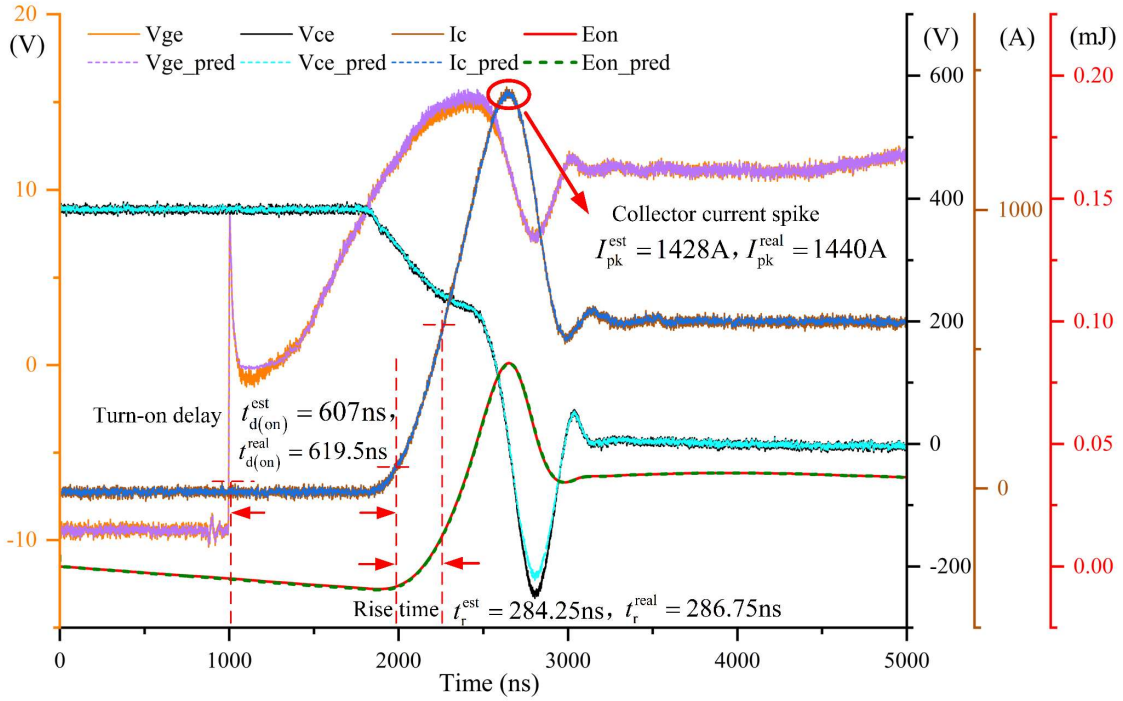
(b) Turn-off transient

Fig. 5-9. Turn-on and turn-off transient waveforms of IGBT DIM1200FSS12-A000 DPT estimating results at the working condition Tc=100℃, VDC=400V, Ic=1000A ((a) turn-on transient, (b) turn-off transient).

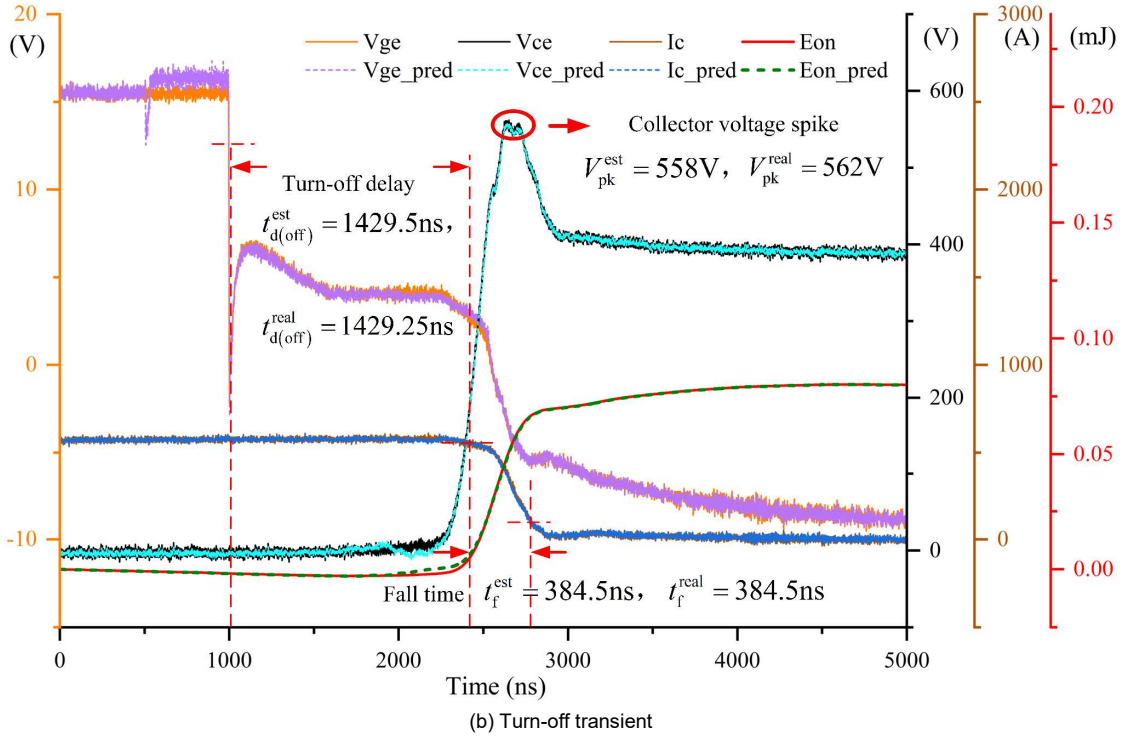
Fig. 5-9(a) and (b) demonstrate the performance of the proposed method on turn-on transient and turn-off transient, respectively. It can be found that when using the data test of DIM1200FSS12-A000, the DPT waveform estimation is highly accurate.

5.5.3 Graph Transfer Estimation Result

After the graph transfer operation, we performed a complete DPT result estimation verification for the IGBT batch DIM1200FSS12-A076.



(a) Turn-on transient



(b) Turn-off transient

Fig. 5-10. Turn-on and turn-off transient waveforms of IGBT DIM1200FSS12-A076 DPT estimating results at the working condition Tc=80℃, VDC=400V, Ic=600A ((a) turn-on transient, (b) turn-off transient).

As shown in Fig. 5-10, the transferred model can correctly estimate the DPT results of DIM1200FSS12-A076. In order to make the experimental results more reliable, we conducted DPT result estimation experiments under all working conditions. The experimental performance of the proposed MAFGCN method tested on 40℃, 60℃, 80℃ and 100℃ (average value of 36 working conditions at each Tc) is summarized in Table 5-4. However, MAFGCN is the first deep learning-based method for IGBT DPT waveform estimation so that we cannot compare with similar methods in other literature.

TABLE 5-4 Performance of The Proposed Method Tested on Two Batches of IGBTs (Average Value of 36 Working Conditions at Each Tc)

T_c	40℃			60℃			80℃			100℃		
Batch	DIM1200FSS12-A000											
Indexes	MAE	RMSE	LE	MAE	RMSE	LE	MAE	RMSE	LE	MAE	RMSE	LE
V_{ge}	0.0379	0.0953	1.0019	0.0354	0.1085	1.0086	0.0547	0.0912	1.0296	0.0482	0.0713	1.0095

V_{ce}	1.2684	1.9497	2.456e+11	1.5983	3.0613	5.654e+20	1.2659	2.1545	8.699e+14	1.3973	2.3164	8.869e+16
I_c	2.5474	6.0122	5.654e+21	3.5645	6.6128	4.894e+27	3.8973	7.5486	4.263e+32	6.7681	9.2125	8.308e+38
E_{on}/ E_{off}	0.0002	0.0003	1.0000	0.0002	0.0004	1.0000	0.0002	0.0004	1.0000	0.0003	0.0004	1.0000
Batch	DIM1200FSS12-A076											
Indexes	MAE	RMSE	LE	MAE	RMSE	LE	MAE	RMSE	LE	MAE	RMSE	LE
V_{ge}	0.1598	0.2959	3.0489	0.0568	0.1387	1.0214	0.1935	0.2825	1.0688	0.1652	0.2544	1.0311
V_{ce}	0.2565	0.8975	9.671e+2	1.3112	2.8485	8.699e+17	0.1992	1.0231	7.626e+3	1.8469	4.0545	4.549e+19
I_c	1.9857	11.9497	1.620e+35	2.4963	10.9862	1.620e+31	2.7917	13.3959	1.620e+37	2.2489	11.2167	2.816e+33
E_{on}/ E_{off}	0.0002	0.0004	1.0000	0.0003	0.0005	1.0000	0.0004	0.0007	1.0000	0.0004	0.0008	1.0001

5.6 Summary

This chapter proposes a novel DPT waveform result estimating method, referred to as MAFGCN. The MAFGAN comprehensively considers the correlation between each working condition to build a spatial graph structure and embeds DPT waveform data. This method also designs MAFN block connected to GCN to ensure that the model can capture excellent temporal feature representations and preform high accuracy. Through the above experimental analysis, the following conclusions can be drawn:

1) This MAFGAN based on deep learning and graph neural network is accurate for IGBT DPT estimation. These reliable estimates can be used to calculate the dynamic electrical characteristics of the IGBT module.
2) The estimation of the test datasets of the same and similar batches of IGBTs has high accuracy (as shown in Table 5-4).
3) Through graph transfer process, the well-trained model can be used to estimate other batch IGBTs. For IGBT batches with certain differences, the proposed method can still accurately estimate the correct DPT results.

In the feature work, we will try to use the proposed MAFGCN with graph migration method to estimate other brands of IGBT module (e.g., Infineon) DPT results at all working condition

with only DPT data at room temperature. In the case of the same graph structure, the proposed method has great potential.

Chapter 6 Conclusions and Future Work

6.1 Conclusions

As an eco-friendly transportation method, EVs have been witnessing an exponential growth. However, it must be acknowledged that the technologies concerning EVs are not entirely mature and there are still many problems that need in-depth explorations. Notably, the crash safety of EVs is attracting significant attention from not only the scholars and manufacturers but also the ordinary people now, making it a valuable topic for research. Reliability problems of EV powertrains may occur in any power electronic (PE) component and mechanical part, both sudden and cumulative. These faults in different locations and degrees will continuously threaten the life of drivers and pedestrians, bringing irreparable consequences. Therefore, monitoring and predicting the real-time health status of EV powertrain is a high-priority, arduous and challenging task. Through using the developed AI-supported techniques, contributing to improving safety of EV powertrains. The main work and contributions of this study can be summarized as follows:

In Chapter 2, four aspects about AI-supported EV powertrain safety improvement topics are reviewed. With the rapid development of AI, support methods that are developed based on this technology can further enhance the electric powertrain condition monitoring and fault diagnosis. This study presents a comprehensive review of this, and it can be learned from this paper that AI-based data-driven method can avoid the need for accurate physical models of the system, and AI can also help build accurate system models. A large number of experiments and case studies have proved that both the AI data-driven method and the AI-supported model-based method have high accuracy and performance, and some advanced AI technologies can achieve functions and effects that cannot be achieved by conventional methods. Therefore, the application of AI in PE has important practical significance. Firstly, this chapter focuses on feature engineering and three main components of EV powertrain and summarizes the recent research on EV. The motivation, advantages and challenges of using AI methods in PE problems are explained. Secondly, this chapter carries out the quantitative comparative analysis on condition monitoring and fault diagnosis methods in specific fields. In each application field, specific practical case studies on EV and quantitative comparative analysis with traditional

methods are provided. Thirdly, based on the characteristics of EV, this review analyses the feasibility, advantages and disadvantages of each method in the practical application of EV.

In Chapter 3, the self-attention mechanism in the field of online fault diagnosis of motor bearing with higher accuracy and a new diagnosis framework based on IDDAN for solving the problem that it is hard to obtain enough labeled data to train a diagnosis model for a new target machine are proposed. Our experiment results present that the fine-tune-based transfer learning method could get better accuracy on the same dataset and the IDDAN has a better performance by pre-training using large-scale data. The mentioned DA method provides sufficient pre-training samples for IDDAN. Meanwhile, when IDDAN consumes large-scale data for pre-training, its diagnostic accuracy could surpass the CNN-based transfer learning model. Firstly, this chapter proposes a self-attention mechanism-based intelligent fault diagnosis method IDDAN for deploying on new machines with a small number of labeled data by transfer learning. Secondly, the proposed DA method in Section 3.3.1 effectively expands the number of pre-training samples and has an excellent effect on IDDAN. Thirdly, the IDDAN obtain a higher recognition accuracy of multiple bearing fault conditions with a large amount of pre-training data than the classic CNN-based method.

In Chapter 4, a novel DAAR model with an MT-LSTR CP-labels generator for precisely estimating PMT of IPMSM is proposed. Through the above experimental analysis, the proposed method can effectively reduce the PDF discrepancy between source and target domains. During experimentation, the offline training time of MT-LSTR is 1.83×10^4 seconds, and real-time CP-labels generating time is 8.78×10^{-1} seconds. By verifying on 4 PU PMSM subdatasets, the DAAR model demonstrated a maximum estimation error of 9.79° C, and average estimation errors of 3.24° C, 4.13° C, 4.98° C and 4.74° C across four different test profiles. In comparative experiments on 6 complex drive circles of PU dataset, the minimum estimation MAE and RMSE are 4.111 and 6.161 while the maximum MAE and RMSE are 8.656 and 10.958. Since this chapter focuses on IPMSM, we will attempt to acquire source domain data directly from the surface-mounted PMSM (SPMSM). The performance of this approach on the SPMSM will be investigated in future work.

In Chapter 5, This chapter proposes a novel DPT waveform result estimating method, referred to as MAFGCN. The MAFGAN comprehensively considers the correlation between each working condition to build a spatial graph structure and embeds DPT waveform data. This

method also designs MAFN block connected to GCN to ensure that the model can capture excellent temporal feature representations and preform high accuracy. Experimental analysis has proved that the MAFGAN based on deep learning and graph neural network is accurate for IGBT DPT estimation. These reliable estimates can be used to calculate the dynamic electrical characteristics of the IGBT module. Secondly, the estimation of the test datasets of the same and similar batches of IGBTs has high accuracy (as shown in Table 5-4). Finally, through graph transfer process, the well-trained model can be used to estimate other batch IGBTs. For IGBT batches with certain differences, the proposed method can still accurately estimate the correct DPT results.

All in all, in terms of the topic about AI-based powertrain safety improvement applications, this research combines both theoretical analysis and simulation/experimental validation to investigate and develop novel AI algorithm for solve more and hard engineering problems.

6.2 Future Work

Based on the conclusions above and considering the limitations of the existed work, future research could be carried out in the following.

Most AI-based data-driven methods applied in EV powertrains can achieve high-precision prediction and fault diagnosis without prior knowledge and physical model knowledge. On the other hand, many model-based conventional methods can build more accurate physical models by the intervention of AI.

In EV applications, most of the existing AI-based fault diagnosis methods are supervised learning methods. The supervised learning method is characterized by the need to use a large amount of labeled data for AI model training. In particular, the data labeling work that relies on the method of manually extracting the fault features from signals requires a lot of manpower and time. Although the trained model can achieve higher accuracy, it is not cost-effective in actual EV applications.

At present, there have been a few studies applying transfer learning to solve health monitoring problems in EV powertrains [96]. Using transfer learning can make the AI model transfer knowledge from the existing data to reduce or even eliminate the need for training data of the machine to be tested. Using unlabeled data can directly detect failures.

Self-supervised learning methods can also solve the expensive and time-consuming data labeling problem. Fig. 6-1 demonstrates the diagram of contrastive self-supervised learning for condition classification. For example, in paper [236], more actions of unsupervised learning had been taken than simply feature extraction from historical fault signals and supervised classification. It detects the severity of many types of faults with various motor speed and conduct data analysis with unlabeled data. Interestingly, this method can detect the faults of three different components in the powertrain at the same time (Fig. 6-2) and can diagnose the severity of each fault.

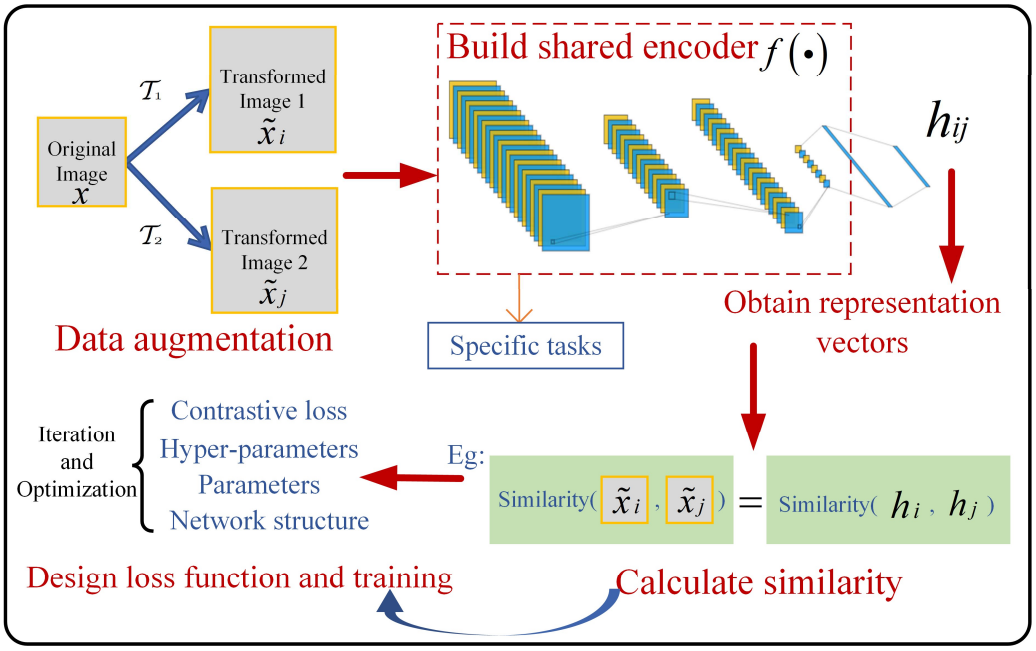


Fig. 6-1. Flowchart of the contrastive self-supervised learning diagram.

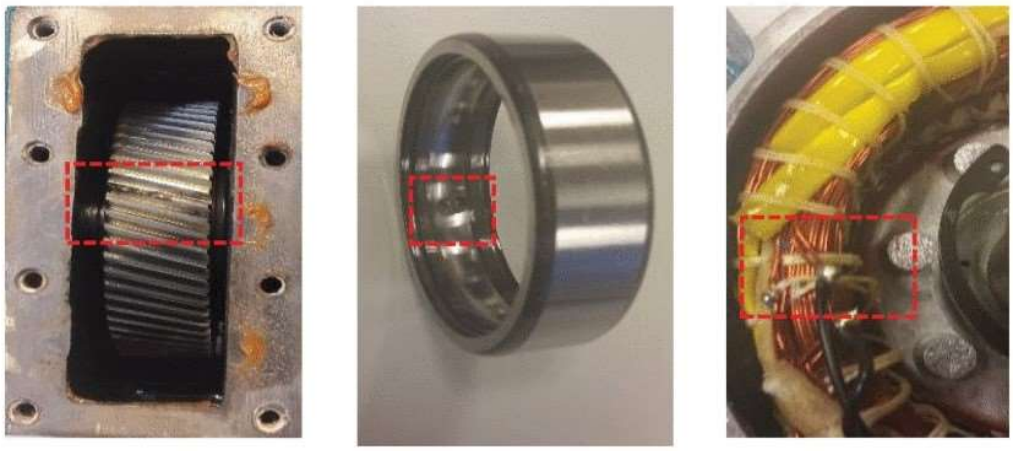


Fig. 6-2 Faulty components of electric powertrain: small gear tooth damage (left), Outer-race damaged bearing (mid) and 10% inter-turn short-circuit fault in the stator (right) [236].

From this example, it can be analysed that another AI application prospect is to propose a single model for multi-component, multi-fault and multi-severity diagnosis. Both the existing AI-based methods and conventional methods can only diagnose the faults of a single component, or even single fault. In order to monitor and diagnose the health status of the entire EV powertrain, multiple methods or models are often needed. Therefore, achieving the single-model full diagnosis can greatly reduce the burden of on-board computing devices.

The EV industry is rapidly evolving, and new technologies, components, and architectures are being introduced. Transferring diagnostic models to new generations of EV powertrains requires continuous updates and adaptations to accommodate new features and fault characteristics. The training of the EV fault diagnosis model is very time-consuming. Aiming at the problem that EV companies need to frequently update and upgrade their condition monitoring and fault diagnosis models, here are two application prospects and corresponding AI technologies: Quickly updating and upgrading the model and quickly rebuilding the model:

- **Quickly updating and upgrading the model**

  Continuous learning technology is applied to the fault diagnosis model of EV powertrains, so that it can add new diagnosis tasks to the previously trained model (Fig. 6-3). In particular, by combining IoT and cloud computing, it can be achieved that quick updates and upgrades of the model in EV powertrain safety applications. This approach significantly reduces the time and cost associated with traditional model updates and

upgrades. However, considerations such as data privacy, network security, and compliance need to be addressed, and the updated and upgraded models should undergo thorough testing and validation to ensure stability and reliability.

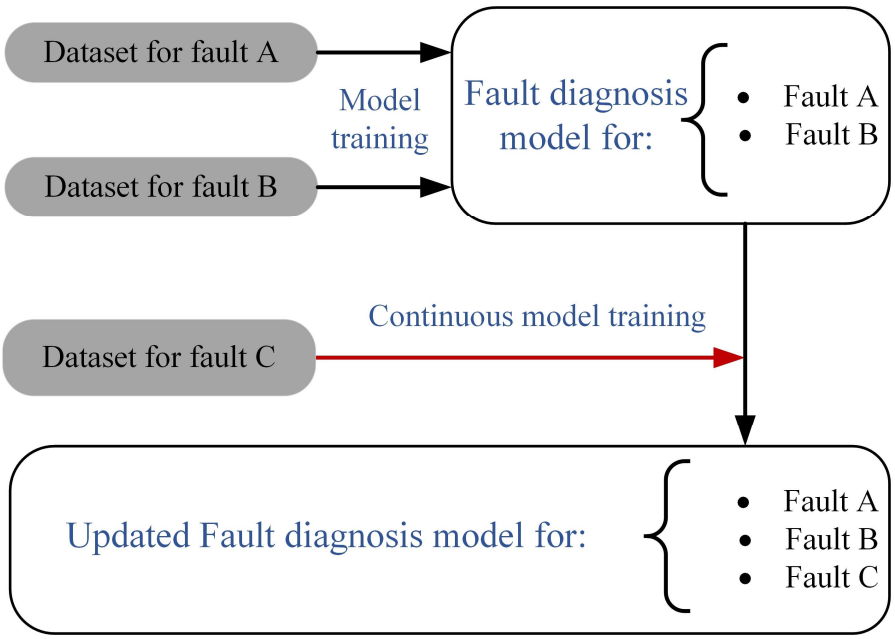


Fig. 6-3. Schematic diagram of quickly updating and upgrading the model.

- **Quickly rebuilding the model**

Meta-learning technology is applied to the fault diagnosis model of EV powertrains. A learn-to-learn model can be trained, which has the knowledge to quickly establish new fault diagnosis models and can directly assign empirical parameters to the new model and achieve excellent performance.

- **Apply advanced powerful AI algorithms**

Although current AI-based methods predict EV motor RUL, battery RUL and SOC, etc. are relatively accurate, there are still errors. The future of AI is aimed at reducing prediction errors. RNN models such as LSTM are currently the main AI technology for dealing with prediction problems, but the disadvantage of LSTM is that it cannot consider global information and cannot be calculated in parallel in the device. At present, no one has used the Transformer model in EV prediction applications. It has a unique attention mechanism and can be calculated in parallel, which reduces the prediction time while improving accuracy. This will be a trend for AI to improve

performance in EV applications in the future.

- **Applications based on large models and large language models (LLMs)**

  Furthermore, recent-emerged large-scale parametric language models such as ChatGPT have a certain impact on EV powertrain safety application. ChatGPT can provide an interactive interface for users to communicate and describe potential issues or symptoms related to the EV powertrain. It can ask clarifying questions, gather relevant information, and assist users in diagnosing faults or determining the next steps for condition monitoring.

References

[1] J. Medina-García, T. Sánchez-Rodríguez, J. Galán, A. Delgado, F. Gómez-Bravo, and R. Jiménez, "A Wireless Sensor System for Real-Time Monitoring and Fault Detection of Motor Arrays," *Sensors,* vol. 17, no. 3, p. 469, Feb. 2017.

[2] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun and J. He, "Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests," *IEEE Sensors Journal,* vol. 17, no. 17, pp. 5581-5588, 1 Sept.1, 2017.

[3] D. Wang, "K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited," *Mechanical Systems and Signal Processing,* vol. 70–71, p. 201–208, Mar. 2016.

[4] F. Ben Abid, S. Zgarni and A. Braham, "Distinct Bearing Faults Detection in Induction Motor by a Hybrid Optimized SWPT and aiNet-DAG SVM," *IEEE Transactions on Energy Conversion,* vol. 33, no. 4, pp. 1692-1699, Dec. 2018.

[5] Q. He, J. Zhao, G. Jiang, and P. Xie, "An Unsupervised Multiview Sparse Filtering Approach for Current-Based Wind Turbine Gearbox Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 8, p. 5569–5578, Aug. 2020.

[6] A. Mohammed and S. Djurovic, "Stator Winding Internal Thermal Monitoring and Analysis Using In Situ FBG Sensing Technology," *IEEE Transactions on Energy Conversion,* vol. 33, no. 3, p. 1508–1518, Sep. 2018.

[7] Siwei Cheng, Yi Du, J. A. Restrepo, Pinjia Zhang, and T. G. Habetler, "A Nonintrusive Thermal Monitoring Method for Induction Motors Fed by Closed-Loop Inverter Drives," *IEEE Transactions on Power Electronics,* vol. 27, no. 9, p. 4122–4131, Sep. 2012.

[8] H. Shao, M. Xia, G. Han, Y. Zhang and J. Wan, "Intelligent Fault Diagnosis of Rotor-Bearing System Under Varying Working Conditions With Modified Transfer Convolutional Neural Network and Thermal Images," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 5, pp. 3488-3496, May 2021.

[9] G. Singh and V. N. A. Naikan, "Infrared thermography based diagnosis of inter-turn fault and cooling system failure in three phase induction motor," *Infrared Physics & Technology,* vol. 87, p. 134–138, Dec. 2017.

[10] G. Singh, T. Ch. Anil Kumar, and V. N. A. Naikan, "Induction motor inter turn fault detection using infrared thermographic analysis," *Infrared Physics & Technology,* vol. 77, p. 277–282, Jul. 2016.

[11] Lu, SiLiang; Wang, Xiaoxian; He, Qingbo; Liu, Fang; Liu, Yongbin;, "Fault diagnosis of motor bearing with speed fluctuation via angular resampling of transient sound signals," *Journal of Sound and Vibration,* vol. 385, pp. 16-32, Dec. 2016.

[12] P. A. Delgado-Arredondo, D. Morinigo-Sotelo, R. A. Osornio-Rios, J. G. Avina-Cervantes, H. Rostro-Gonzalez, and R. de J. Romero-Troncoso, "Methodology for fault detection in induction motors via sound and vibration signals," *Mechanical Systems and Signal Processing,* vol. 83, p. 568–589, Jan. 2017.

[13] A. Ulatowski and A. M. Bazzi, "A Combinational-Logic Method for Electric Vehicle Drivetrain Fault Diagnosis," *IEEE Transactions on Industry Applications,* vol. 52, no. 2, pp. 1796-1807, March-April 2016.

[14] J. Burriel-Valencia et al., "Automatic Fault Diagnostic System for Induction Motors under Transient Regime Optimized with Expert Systems," *Electronics,* vol. 8, no. 1, p. 6, Dec. 2018.

[15] H. Yan, Y. Xu, F. Cai, H. Zhang, W. Zhao, and C. Gerada, "PWM-VSI Fault Diagnosis for a PMSM Drive Based on the Fuzzy Logic Approach," *IEEE Transactions on Power Electronics,* vol. 34, no. 1, p. 759–768, Jan. 2019.

[16] F. Filippetti, G. Franceschini, C. Tassoni and P. Vas, "Recent developments of induction motor drives fault diagnosis using AI techniques," *IEEE Transactions on Industrial Electronics,* vol. 47, no. 5, pp. 994-1004, Oct. 2000.

[17] K. M. Sundaram, A. Hussain, P. Sanjeevikumar, J. B. Holm-Nielsen, V. K. Kaliappan and B. K. Santhoshi, "Deep Learning for Fault Diagnostics in Bearings, Insulators, PV Panels, Power Lines, and Electric Vehicle Applications—The State-of-the-Art Approaches," *IEEE Access,* vol. 9, pp. 41246-41260, 2021.

[18] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing,* vol. 108, p. 33–47, Aug. 2018.

[19] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault Diagnosis for Rotating Machinery Using Vibration Measurement Deep Statistical Feature Learning," *Sensors,* vol. 16, no. 6, p. 895, Jun. 2016.

[20] A. Ngaopitakkul and S. Bunjongjit, "An application of a discrete wavelet transform and a back-propagation neural network algorithm for fault diagnosis on single-circuit transmission line," *International Journal of Systems Science,* vol. 44, no. 9, p. 1745–1761, Sep. 2013.

[21] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation,* vol. 19, no. 90, p. 297, Apr. 1965.

[22] Y. Li, W. Zhao, Q. Li, T. Wang, and G. Wang, "In-Situ Monitoring and Diagnosing for Fused Filament Fabrication Process Based on Vibration Sensors," *Sensors,* vol. 19, no. 11, p. 2589, Jun. 2019.

[23] T. Yang, H. Pen, Z. Wang and C. S. Chang, "Feature Knowledge Based Fault Detection of Induction Motors Through the Analysis of Stator Current Data," *IEEE Transactions on Instrumentation and Measurement,* vol. 65, no. 3, pp. 549-558, March 2016.

[24] R. de Jesus Romero-Troncoso, "Multirate Signal Processing to Improve FFT-Based Analysis for Detecting Faults in Induction Motors," *IEEE Transactions on Industrial Informatics,* vol. 13, no. 3, pp. 1291-1300, June 2017.

[25] Z. Chen, C. Li, and R.-V. Sanchez, "Gearbox Fault Identification and Classification with Convolutional Neural Networks," *Shock and Vibration,* vol. 2015, p. 1–10, 2015.

[26] B. Gou, Y. Xu, Y. Xia, Q. Deng and X. Ge, "An Online Data-Driven Method for Simultaneous Diagnosis of IGBT and Current Sensor Fault of Three-Phase PWM Inverter in Induction Motor Drives," *IEEE Transactions on Power Electronics,* vol. 35, no. 12, pp. 13281-13294, Dec. 2020.

[27] D. R. Musk, "A Comparison of Quantum and Traditional Fourier Transform Computations," *Computing in Science & Engineering,* vol. 22, no. 6, pp. 103-110, 1 Nov.-Dec. 2020.

[28] E. H. E. Bouchikhi, V. Choqueuse, and M. E. H. Benbouzid, "Current Frequency Spectral Subtraction and Its Contribution to Induction Machines' Bearings Condition Monitoring," *IEEE Transactions on Energy Conversion,* vol. 28, no. 1, p. 135–144, Mar. 2013.

[29] C. Shen, D. Wang, F. Kong, and P. W. Tse, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement,* vol. 46, no. 4, p. 1551–1564, May 2013.

[30] M. M. Rahman and M. N. Uddin, "Online Unbalanced Rotor Fault Detection of an IM Drive Based on Both Time and Frequency Domain Analyses," *IEEE Transactions on Industry Applications,* vol. 53, no. 4, pp. 4087-4096, July-Aug. 2017.

[31] I. P. Georgakopoulos, E. D. Mitronikas, and A. N. Safacas, "Detection of Induction Motor Faults in Inverter Drives Using Inverter Input Current Analysis," *IEEE Transactions on Industrial Electronics,* vol. 58, no. 9, p. 4365–4373, Sep. 2011.

[32] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mechanical Systems and Signal Processing,* vol. 25, no. 2, p. 485–520, Feb. 2011.

[33] B. Tang, W. Liu, and T. Song, "Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution," *Renewable Energy,* vol. 35, no. 12, p. 2862–2866, Dec. 2010.

[34] Z. Shen, X. Chen, X. Zhang, and Z. He, "A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM," *Measurement,* vol. 45, no. 1, p. 30–40, Jan. 2012.

[35] Y. Li, R. Zhang, Y. Guo, P. Huan, and M. Zhang, "Nonlinear Soft Fault Diagnosis of Analog Circuits Based on RCCA-SVM," *IEEE Access,* vol. 8, p. 60951–60963, 2020.

[36] T. Gao, J. Yang, and S. Jiang, "A Novel Incipient Fault Diagnosis Method for Analog Circuits Based on GMKL-SVM and Wavelet Fusion Features," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, p. 1–15, 2021.

[37] S.-W. Chen and Y.-H. Chen, "Hardware Design and Implementation of a Wavelet De-Noising Procedure for Medical Signal Preprocessing," *Sensors,* vol. 15, no. 10, p. 26396–26414, Oct. 2015.

[38] J.-H. Zhong, P. Wong, and Z.-X. Yang, "Simultaneous-Fault Diagnosis of Gearboxes Using Probabilistic Committee Machine," *Sensors,* vol. 16, no. 2, p. 185, Feb. 2016.

[39] Y. Wang, G. Xu, L. Liang, and K. Jiang, “Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis,” *Mechanical Systems and Signal Processing,* vol. 54–55, p. 259–276, Mar. 2015.

[40] Y. Wang, P. W. Tse, B. Tang, Y. Qin, L. Deng, and T. Huang, “Kurtogram manifold learning and its application to rolling bearing weak signal detection,” *Measurement,* vol. 127, p. 533–545, Oct. 2018.

[41] P. K. Kankar, S. C. Sharma, and S. P. Harsha, “Rolling element bearing fault diagnosis using wavelet transform,” *Neurocomputing,* vol. 74, no. 10, p. 1638–1645, May 2011.

[42] Y. Yang, J. Cheng, and K. Zhang, “An ensemble local means decomposition method and its application to local rub-impact fault diagnosis of the rotor systems,” *Measurement,* vol. 45, no. 3, p. 561–570, Apr. 2012.

[43] Z. WU and N. E. HUANG, “ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOISE-ASSISTED DATA ANALYSIS METHOD,” *Advances in Adaptive Data Analysis,* vol. 01, no. 01, p. 1–41, Jan. 2009.

[44] M. Beibei, S. Yanxia, W. Dinghui and Z. Zhipu, “Three level inverter fault diagnosis using EMD and support vector machine approach,” in *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Siem Reap, Cambodia, 2017.

[45] H. Fan, S. Shao, X. Zhang, X. Wan, X. Cao and H. Ma, “Intelligent Fault Diagnosis of Rolling Bearing Using FCM Clustering of EMD-PWVD Vibration Images,” *IEEE Access,* vol. 8, pp. 145194-145206, 2020.

[46] J. Yuan, H. Jiang, Q. Zhao, C. Xu, H. Liu and Y. Tian, “Dual-Mode Noise-Reconstructed EMD for Weak Feature Extraction and Fault Diagnosis of Rotating Machinery,” *IEEE Access,* vol. 7, pp. 173541-173548, 2019.

[47] R. Sadeghi, H. Samet and T. Ghanbari, "Detection of Stator Short-Circuit Faults in Induction Motors Using the Concept of Instantaneous Frequency," *IEEE Transactions on Industrial Informatics,* vol. 15, no. 8, pp. 4506-4515, Aug. 2019.

[48] J. Gilles, "Empirical Wavelet Transform," *IEEE Transactions on Signal Processing,* vol. 61, no. 16, p. 3999–4010, Aug. 2013.

[49] F. Gougam, C. Rahmoune, D. Benazzouz, and B. Merainani, "Bearing fault diagnosis based on feature extraction of empirical wavelet transform (EWT) and fuzzy logic system (FLS) under variable operating conditions," *Journal of Vibroengineering,* vol. 21, no. 6, p. 1636–1650, Sep. 2019.

[50] Y. Jiang, H. Zhu, and Z. Li, "A new compound faults detection method for rolling bearings based on empirical wavelet transform and chaotic oscillator," *Chaos, Solitons & Fractals,* vol. 89, p. 8–19, Aug. 2016.

[51] C. Zhang and Y. Liu, "A Two-Step Denoising Strategy for Early-Stage Fault Diagnosis of Rolling Bearings," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 9, p. 6250–6261, Sep. 2020.

[52] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis," *IEEE Transactions on Industrial Electronics,* vol. 63, no. 3, p. 1793–1803, Mar. 2016.

[53] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing Fault Diagnosis Method Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning," *Sensors,* vol. 19, no. 5, p. 1088, Mar. 2019.

[54] S. Guo, T. Yang, W. Gao, C. Zhang, and Y. Zhang, "An Intelligent Fault Diagnosis Method for Bearings with Variable Rotating Speed Based on Pythagorean Spatial Pyramid Pooling CNN," *Sensors,* vol. 18, no. 11, p. 3857, Nov. 2018.

[55] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing,* vol. 272, p. 619–628, Jan. 2018.

[56] X. Wang, S. Si, and Y. Li, "Multiscale diversity entropy: a novel dynamical measure for fault diagnosis of rotating machinery," *IEEE Transactions on Industrial Informatics,* 2020.

[57] X. Wang, S. Si and Y. Li, "Variational embedding multiscale diversity entropy for fault diagnosis of large-scale machinery," *IEEE Transactions on Industrial Electronics,* 2021.

[58] T. Liu, J. Chen, G. Dong, W. Xiao, and X. Zhou, "The fault detection and diagnosis in rolling element bearings using frequency band entropy," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science,* vol. 227, no. 1, p. 87–99, Mar. 2012.

[59] H. Li, T. Liu, X. Wu, and Q. Chen, "Enhanced Frequency Band Entropy Method for Fault Feature Extraction of Rolling Element Bearings," *IEEE Transactions on Industrial Informatics,* vol. 16, no. 9, p. 5780–5791, Sep. 2020.

[60] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," *IEEE Transactions on Signal Processing,* vol. 62, no. 3, pp. 531-544, Feb.1, 2014.

[61] N. u. Rehman and H. Aftab, "Multivariate Variational Mode Decomposition," *IEEE Transactions on Signal Processing,* vol. 67, no. 23, pp. 6039-6052, 1 Dec.1, 2019.

[62] Y. Li, G. Li, Y. Wei, B. Liu, and X. Liang, "Health condition identification of planetary gearboxes based on variational mode decomposition and generalized composite multi-scale symbolic dynamic entropy," *ISA Transactions,* vol. 81, p. 329–341, Oct. 2018.

[63] F. Li, R. Li, L. Tian, L. Chen, and J. Liu, "Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in

variable working conditions," *Mechanical Systems and Signal Processing,* vol. 116, p. 462–479, Feb. 2019.

[64] J. Yu and J. Lv, "Weak Fault Feature Extraction of Rolling Bearings Using Local Mean Decomposition-Based Multilayer Hybrid Denoising," *IEEE Transactions on Instrumentation and Measurement,* vol. 66, no. 12, p. 3148–3159, Dec. 2017.

[65] R. Yan and R. X. Gao, "Hilbert–Huang Transform-Based Vibration Signal Analysis for Machine Health Monitoring," *IEEE Transactions on Instrumentation and Measurement,* vol. 55, no. 6, p. 2320–2329, Dec. 2006.

[66] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing Health Monitoring Based on Hilbert–Huang Transform, Support Vector Machine, and Regression," *IEEE Transactions on Instrumentation and Measurement,* vol. 64, no. 1, p. 52–62, Jan. 2015.

[67] E. Elbouchikhi, V. Choqueuse, Y. Amirat, M. E. H. Benbouzid, and S. Turri, "An Efficient Hilbert–Huang Transform-Based Bearing Faults Detection in Induction Machines," *IEEE Transactions on Energy Conversion,* vol. 32, no. 2, p. 401–413, Jun. 2017.

[68] X. Zhao and B. Ye, "Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock," *Mechanical Systems and Signal Processing,* vol. 25, no. 5, p. 1617–1631, Jul. 2011.

[69] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen, "Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing,* vol. 52–53, p. 338–359, Feb. 2015.

[70] M. Zhao and X. Jia, "A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery," *Mechanical Systems and Signal Processing,* vol. 94, p. 129–147, Sep. 2017.

[71] Li, Huang, and Ji, "Bearing Fault Diagnosis with a Feature Fusion Method Based on an Ensemble Convolutional Neural Network and Deep Neural Network," *Sensors,* vol. 19, no. 9, p. 2034, Apr. 2019.

[72] A. Widodo and B.-S. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Systems with Applications,* vol. 33, no. 1, p. 241–250, Jul. 2007.

[73] A. Widodo, B.-S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors," *Expert Systems with Applications,* vol. 32, no. 2, p. 299–312, Feb. 2007.

[74] M. Z. Ali, M. N. S. K. Shabbir, S. M. K. Zaman, and X. Liang, "Single- and Multi-Fault Diagnosis Using Machine Learning for Variable Frequency Drive-Fed Induction Motors," *IEEE Transactions on Industry Applications,* vol. 56, no. 3, p. 2324–2337, May 2020.

[75] M. H. Gharavian, F. Almas Ganj, A. R. Ohadi, and H. Heidari Bafroui, "Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes," *Neurocomputing,* vol. 121, p. 150–159, Dec. 2013.

[76] F. Pacheco et al., "A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions," *Neurocomputing,* vol. 194, p. 192–206, Jun. 2016.

[77] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor Bearing Fault Diagnosis Using Trace Ratio Linear Discriminant Analysis," *IEEE Transactions on Industrial Electronics,* vol. 61, no. 5, p. 2441–2451, May 2014.

[78] R. Z. Haddad and E. G. Strangas, "On the Accuracy of Fault Detection and Separation in Permanent Magnet Synchronous Machines Using MCSA/MVSA and LDA," *IEEE Transactions on Energy Conversion,* vol. 31, no. 3, p. 924–934, Sep. 2016.

[79] J. Harmouche, C. Delpha, and D. Diallo, “Improved Fault Diagnosis of Ball Bearings Based on the Global Spectrum of Vibration Signals,” *IEEE Transactions on Energy Conversion,* vol. 30, no. 1, p. 376–383, 2015.

[80] B. R. Nayana and P. Geethanjali, “Improved Identification of Various Conditions of Induction Motor Bearing Faults,” *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 5, p. 1908–1919, May 2020.

[81] D.-H. Lee, J.-H. Ahn, and B.-H. Koh, “Fault Detection of Bearing Systems through EEMD and Optimization Algorithm,” *Sensors,* vol. 17, no. 11, p. 2477, Oct. 2017.

[82] R. N. Toma, A. E. Prosvirin, and J.-M. Kim, “Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers,” *Sensors,* vol. 20, no. 7, p. 1884, Mar. 2020.

[83] B.-S. Yang, X. Di, and T. Han, “Random forests classifier for machine fault diagnosis,” *Journal of Mechanical Science and Technology,* vol. 22, no. 9, p. 1716–1725, Sep. 2008.

[84] C. Gong, Y. Hu, G. Chen, H. Wen, Z. Wang, and K. Ni, “A DC-Bus Capacitor Discharge Strategy for PMSM Drive System With Large Inertia and Small System Safe Current in EVs,” *IEEE Transactions on Industrial Informatics,* vol. 15, no. 8, p. 4709–4718, Aug. 2019.

[85] H. Cherif et al., “Early detection and localization of stator inter-turn faults based on discrete wavelet energy ratio and neural networks in induction motor,” *Energy,* vol. 212, p. 118684, Dec. 2020.

[86] H. Shi, L. Guo, S. Tan, and X. Bai, “Rolling Bearing Initial Fault Detection Using Long Short-Term Memory Recurrent Network,” *IEEE Access,* vol. 7, p. 171559–171569, 2019.

[87] 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, p. 1–14, Jan. 2018.

[88] M. Kang, J. Kim and J. Kim, "An FPGA-Based Multicore System for Real-Time Bearing Fault Diagnosis Using Ultrasampling Rate AE Signals," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 4, pp. 2319-2329, April 2015.

[89] L. Wan, K. Gong, G. Zhang, X. Yuan, C. Li and X. Deng, "An Efficient Rolling Bearing Fault Diagnosis Method Based on Spark and Improved Random Forest Algorithm," *IEEE Access,* vol. 9, pp. 37866-37882, 2021.

[90] M. Kuncan, "An Intelligent Approach for Bearing Fault Diagnosis: Combination of 1D-LBP and GRA," *IEEE Access,* vol. 8, pp. 137517-137529, 2020.

[91] M. Kang, J. Kim, L. M. Wills and J. Kim, "Time-Varying and Multiresolution Envelope Analysis and Discriminative Feature Analysis for Bearing Fault Diagnosis," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 12, pp. 7749-7761, Dec. 2015.

[92] M. Hamadache, D. Lee and K. C. Veluvolu, "Rotor Speed-Based Bearing Fault Diagnosis (RSB-BFD) Under Variable Speed and Constant Load," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 10, pp. 6486-6495, Oct. 2015.

[93] L. Chen, G. Xu, T. Tao and Q. Wu, "Deep Residual Network for Identifying Bearing Fault Location and Fault Severity Concurrently," *IEEE Access,* vol. 8, pp. 168026-168035, 2020.

[94] L. Hou, R. Jiang, Y. Tan and J. Zhang, "Input Feature Mappings-Based Deep Residual Networks for Fault Diagnosis of Rolling Element Bearing With Complicated Dataset," *IEEE Access,* vol. 8, pp. 180967-180976, 2020.

[95] T. Wang, Z. Liu, G. Lu and J. Liu, "Temporal-Spatio Graph Based Spectrum Analysis for Bearing Fault Detection and Diagnosis," *IEEE Transactions on Industrial Electronics,* vol. 68, no. 3, pp. 2598-2607, March 2021.

[96] 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, Sept. 2019.

[97] Q. Jiang, F. Chang and B. Sheng, "Bearing Fault Classification Based on Convolutional Neural Network in Noise Environment," *IEEE Access,* vol. 7, pp. 69795-69807, 2019.

[98] Y. Wang, X. Ding, Q. Zeng, L. Wang and Y. Shao, "Intelligent Rolling Bearing Fault Diagnosis via Vision ConvNet," *IEEE Sensors Journal,* vol. 21, no. 5, pp. 6600-6609, 1 March1, 2021.

[99] L. Yuan, D. Lian, X. Kang, Y. Chen and K. Zhai, "Rolling Bearing Fault Diagnosis Based on Convolutional Neural Network and Support Vector Machine," *IEEE Access,* vol. 8, pp. 137395-137406, 2020.

[100] S. Zgarni, H. Keskes, and A. Braham, "Nested SVDD in DAG SVM for induction motor condition monitoring," *Engineering Applications of Artificial Intelligence,* vol. 71, p. 210–215, May 2018.

[101] H. Keskes and A. Braham, "Recursive Undecimated Wavelet Packet Transform and DAG SVM for Induction Motor Diagnosis," *IEEE Transactions on Industrial Informatics,* vol. 11, no. 5, pp. 1059-1066, Oct. 2015.

[102] B. Bessam, A. Menacer, M. Boumehraz, and H. Cherif, "Detection of broken rotor bar faults in induction motor at low load using neural network," *ISA Transactions,* vol. 64, p. 241–246, Sep. 2016.

[103] 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, April 2018.

[104] 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,* vol. 8, pp. 172043-172056, 2020.

[105] X. Zhu et al., "Rotor fault diagnosis using a convolutional neural network with symmetrized dot pattern images," *Measurement,* vol. 138, p. 526–535, May 2019.

[106] X. Yan, C. Zhang, and Y. Liu, "Multi-branch convolutional neural network with generalized shaft orbit for fault diagnosis of active magnetic bearing-rotor system," *Measurement,* vol. 171, p. 108778, Feb. 2021.

[107] J. Seshadrinath, B. Singh, and B. K. Panigrahi, "Vibration Analysis Based Interturn Fault Diagnosis in Induction Machines," *IEEE Transactions on Industrial Informatics,* vol. 10, no. 1, p. 340–350, Feb. 2014.

[108] S. Toma, L. Capocchi and G. Capolino, "Wound-Rotor Induction Generator Inter-Turn Short-Circuits Diagnosis Using a New Digital Neural Network," *IEEE Transactions on Industrial Electronics,* vol. 60, no. 9, pp. 4043-4052, Sept. 2013.

[109] Y. Nyanteh, C. Edrington, S. Srivastava and D. Cartes, "Application of Artificial Intelligence to Real-Time Fault Detection in Permanent-Magnet Synchronous Machines," *IEEE Transactions on Industry Applications,* vol. 49, no. 3, pp. 1205-1214, May-June 2013.

[110] 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.

[111] M. Singh and A. G. Shaik, "Incipient Fault Detection in Stator Windings of an Induction Motor Using Stockwell Transform and SVM," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 12, pp. 9496-9504, Dec. 2020.

[112] N. R. Devi, D. V. S. S. Siva Sarma and P. V. Ramana Rao, "Diagnosis and classification of stator winding insulation faults on a three-phase induction motor using wavelet and MNN," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 23, no. 5, pp. 2543-2555, October 2016.

[113] P. C. M. Lamim Filho, R. Pederiva, and J. N. Brito, "Detection of stator winding faults in induction machines using flux and vibration analysis," *Mechanical Systems and Signal Processing,* vol. 42, no. 1–2, p. 377–387, Jan. 2014.

[114] M. Riera-Guasp, J. A. Antonino-Daviu and G. Capolino, "Advances in Electrical Machine, Power Electronic, and Drive Condition Monitoring and Fault Detection: State of the Art,," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 3, pp. 1746-1759, March 2015.

[115] A. Sarikhani and O. A. Mohammed, "Inter-Turn Fault Detection in PM Synchronous Machines by Physics-Based Back Electromotive Force Estimation," *IEEE Transactions on Industrial Electronics,* vol. 60, no. 8, pp. 3472-3484, Aug. 2013.

[116] L. Chen, J. Wang and Z. Sun, "Electromagnetic-thermal coupled modelling and analysis of inter-turn short-circuit faults of a permanent magnet alternator," *J. Eng.,* vol. 2019, no. 17, pp. 4426-4431, Jun. 2019.

[117] C. Kral, A. Haumer and S. B. Lee, "A Practical Thermal Model for the Estimation of Permanent Magnet and Stator Winding Temperatures," *IEEE Transactions on Power Electronics,* vol. 29, no. 1, pp. 455-464, Jan. 2014.

[118] D. Reigosa, D. Fernandez, T. Tanimoto, T. Kato and F. Briz, "Comparative Analysis of BEMF and Pulsating High-Frequency Current Injection Methods for PM

Temperature Estimation in PMSMs," *IEEE Transactions on Power Electronics,* vol. 32, no. 5, pp. 3691-3699, May 2017.

[119] G. Feng, C. Lai, K. L. V. Iyer and N. C. Kar, "Improved High-Frequency Voltage Injection Based Permanent Magnet Temperature Estimation for PMSM Condition Monitoring for EV Applications," *IEEE Transactions on Vehicular Technology,* vol. 67, no. 1, pp. 216-225, Jan. 2018.

[120] W. Kirchgässner, O. Wallscheid and J. Böcker, "Estimating Electric Motor Temperatures With Deep Residual Machine Learning," *IEEE Transactions on Power Electronics,* vol. 36, no. 7, pp. 7480-7488, July 2021.

[121] R. K. Singleton, E. G. Strangas and S. Aviyente, "Extended Kalman Filtering for Remaining-Useful-Life Estimation of Bearings," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 3, pp. 1781-1790, March 2015.

[122] Y. Qian and R. Yan, "Remaining Useful Life Prediction of Rolling Bearings Using an Enhanced Particle Filter," *IEEE Transactions on Instrumentation and Measurement,* vol. 64, no. 10, pp. 2696-2707, Oct. 2015.

[123] C. Zhang, P. Lim, A. K. Qin and K. C. Tan, "Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 28, no. 10, pp. 2306-2318, Oct. 2017.

[124] L. Ren, Y. Sun, H. Wang and L. Zhang, "Prediction of Bearing Remaining Useful Life With Deep Convolution Neural Network," *IEEE Access,* vol. 6, pp. 13041-13049, 2018.

[125] R. Liu, B. Yang and A. G. Hauptmann, "Simultaneous Bearing Fault Recognition and Remaining Useful Life Prediction Using Joint-Loss Convolutional Neural Network," *IEEE Transactions on Industrial Informatics,* vol. 16, no. 1, pp. 87-96, Jan. 2020.

[126] J. Zhu, N. Chen and W. Peng, "Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 4, pp. 3208-3216, April 2019.

[127] 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,* vol. 69, no. 1, pp. 401-412, March 2020.

[128] "ISO 15243:2017," ISO, 2017. [Online]. Available: https://www.iso.org/standard/59619.html. [Accessed 24 Sep 2021].

[129] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification," *PHME_CONF,* vol. 3, no. 1, Jul. 2016.

[130] M. Zeng and Z. Chen, "SOSO boosting of the K-SVD denoising algorithm for enhancing fault-induced impulse responses of rolling element bearings," *IEEE Transctions on Industrial Electronics,* vol. 67, no. 2, pp. 1282-1292, Feb. 2020.

[131] Y. Qin, J. Zou, B. Tang and Y. Wang, "Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary," *IEEE Transctions on Industrial Informatics,* vol. 16, no. 1, pp. 215-227, Jan. 2020.

[132] Z. Zhao, S. Wu, B. Qiao and S. Wang, "Enhanced sparse period-group lasso for bearing fault diagnosis," *IEEE Transctions on Industrial Electronics,* vol. 66, no. 3, pp. 2143-2153, Mar. 2019.

[133] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko and T. Darrell, "Deep Domain Confusion: Maximizing for Domain Invariance," arXiv.org, 2014. [Online]. Available: Available: https://arxiv.org/abs/1412.3474v1.

[134] Y. Ganin et al., “Domain-Adversarial Training of Neural Networks,” arXiv.org, 2016. [Online]. Available: https://arxiv.org/abs/1505.07818.

[135] S. J. Pan, I. W. Tsang, J. T. Kwok and Q. Yang, “Domain Adaptation via Transfer Component Analysis,,” *IEEE Transactions on Neural Networks,* vol. 22, no. 2, pp. 199-210, Feb. 2011.

[136] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang and T. Zhang, “Deep Model Based Domain Adaptation for Fault Diagnosis,,” *IEEE Transactions on Industrial Electronics,* vol. 64, no. 3, pp. 2296-2305, March 2017.

[137] K.-H. Chao, L.-Y. Chang, and F.-Q. Xu, “Three-Level T-Type Inverter Fault Diagnosis and Tolerant Control Using Single-Phase Line Voltage,” *IEEE Access,* vol. 8, p. 44075–44086, 2020.

[138] M. S. Haque, S. Choi and J. Baek, “Auxiliary Particle Filtering-Based Estimation of Remaining Useful Life of IGBT,” *IEEE Transactions on Industrial Electronics,* vol. 65, no. 3, pp. 2693-2703, March 2018.

[139] Z. Rao, M. Huang and X. Zha, “IGBT Remaining Useful Life Prediction Based on Particle Filter With Fusing Precursor,” *IEEE Access,* vol. 8, pp. 154281-154289, 2020.

[140] S. H. Ali, M. Heydarzadeh, S. Dusmez, X. Li, A. S. Kamath and B. Akin, “Lifetime Estimation of Discrete IGBT Devices Based on Gaussian Process,” *IEEE Transactions on Industry Applications,* vol. 54, no. 1, pp. 395-403, Jan.-Feb. 2018.

[141] A. Alghassi, S. Perinpanayagam and M. Samie, “Stochastic RUL Calculation Enhanced With TDNN-Based IGBT Failure Modeling,” *IEEE Transactions on Reliability,* vol. 65, no. 2, pp. 558-573, June 2016.

[142] T. Wang, J. Qi, H. Xu, Y. Wang, L. Liu, and D. Gao, “Fault diagnosis method based on FFT-RPCA-SVM for Cascaded-Multilevel Inverter,” *ISA Transactions,* vol. 60, p. 156–163, Jan. 2016.

[143] S. S. Moosavi, A. Kazemi, and H. Akbari, “A comparison of various open-circuit fault detection methods in the IGBT-based DC/AC inverter used in electric vehicle,” *Engineering Failure Analysis,* vol. 96, p. 223–235, Feb. 2019.

[144] Y. Cheng, W. Dong, F. Gao, and G. Xin, “Open-circuit fault diagnosis of traction inverter based on compressed sensing theory,” *Chinese Journal of Electrical Engineering,* vol. 6, no. 1, p. 52–60, Mar. 2020.

[145] D. Chen, Y. Ye, and R. Hua, “Fault Diagnosis of NPC Inverter Based on Multi-layer SVM,” *Communications in Computer and Information Science,* p. 611–621, 2012.

[146] H. Hu, F. Feng, and T. Wang, “Open-circuit fault diagnosis of NPC inverter IGBT based on independent component analysis and neural network,” *Energy Reports,* vol. 6, p. 134–143, Dec. 2020.

[147] X. Wan, H. Hu, Y. Yu, L. Kang, and F. Zeng, “Open-circuit Fault Diagnosis for Grid-connected NPC Inverter based on Independent Component Analysis and Neural Network,” *TELKOMNIKA (Telecommunication Computing Electronics and Control),* vol. 15, no. 1, p. 36, Mar. 2017.

[148] D. Chen, Y. Ye, and R. Hua, “Fault Diagnosis System for NPC Inverter based on Multi-Layer Principal Component Neural Network,” *Journal of Computers,* vol. 8, no. 6, Jun. 2013.

[149] J. Liu, J. Wang, W. Yu, Z. Wang, and G. Zhong, “Open-circuit fault diagnosis of traction inverter based on improved convolutional neural network,” *Journal of Physics: Conference Series,* vol. 1633, p. 012099, Sep. 2020.

[150] Z. Ji and W. Liu, “Open-circuit fault detection for three-phase inverter based on backpropagation neural network,” *Neural Computing and Applications,* vol. 31, no. 9, p. 4665–4674, Jul. 2018.

[151] B. D. E. Cherif, A. Bendiabdellah, and M. Tabbakh, “An Automatic Diagnosis of an Inverter IGBT Open-Circuit Fault Based on HHT-ANN,” *Electric Power Components and Systems,* vol. 48, no. 6–7, p. 589–602, Apr. 2020.

[152] R. B. Dhumale and S. D. Lokhande, “Neural Network Fault Diagnosis of Voltage Source Inverter under variable load conditions at different frequencies,” *Measurement,* vol. 91, p. 565–575, Sep. 2016.

[153] W. Gong, H. Chen, Z. Zhang, M. Zhang, and H. Gao, “A Data-Driven-Based Fault Diagnosis Approach for Electrical Power DC-DC Inverter by Using Modified Convolutional Neural Network With Global Average Pooling and 2-D Feature Image,” *IEEE Access,* vol. 8, p. 73677–73697, 2020.

[154] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, “Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification,” *PHME_CONF,* vol. 3, no. 1, Jul. 2016.

[155] R. Duan, Y. Liao and S. Wang, “Adaptive Morphological Analysis Method and Its Application for Bearing Fault Diagnosis,” *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-10, 2021.

[156] K. Zheng, T. Li, Z. Su and B. Zhang, “Sparse Elitist Group Lasso Denoising in Frequency Domain for Bearing Fault Diagnosis,” *IEEE Transactions on Industrial Informatics,* vol. 17, no. 7, pp. 4681-4691, July 2021.

[157] Y. Qin, L. Jin, A. Zhang and B. He, “Rolling Bearing Fault Diagnosis With Adaptive Harmonic Kurtosis and Improved Bat Algorithm,” *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-12, 2021.

[158] F. Dalvand, S. Dalvand, F. Sharafi and M. Pecht, "Current Noise Cancellation for Bearing Fault Diagnosis Using Time Shifting," *IEEE Transactions on Industrial Electronics,* vol. 64, no. 10, pp. 8138-8147, Oct. 2017.

[159] T. Wang, Z. Liu, G. Lu and J. Liu, "Temporal-Spatio Graph Based Spectrum Analysis for Bearing Fault Detection and Diagnosis," *IEEE Transactions on Industrial Electronics,* vol. 68, no. 3, pp. 2598-2607, March 2021.

[160] D. T. Hoang and H. J. Kang, "A Motor Current Signal-Based Bearing Fault Diagnosis Using Deep Learning and Information Fusion," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 6, pp. 3325-3333, June 2020.

[161] H. Shao, M. Xia, G. Han, Y. Zhang and J. Wan, "Intelligent Fault Diagnosis of Rotor-Bearing System Under Varying Working Conditions With Modified Transfer Convolutional Neural Network and Thermal Images," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 5, pp. 3488-3496, May 2021.

[162] T. Lu, F. Yu, B. Han and J. Wang, "A Generic Intelligent Bearing Fault Diagnosis System Using Convolutional Neural Networks With Transfer Learning," *IEEE Access,* vol. 8, pp. 164807-164814, 2020.

[163] 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, Sept. 2019.

[164] M. Zhao, S. Zhong, X. Fu, B. Tang and M. Pecht, "Deep Residual Shrinkage Networks for Fault Diagnosis," *IEEE Transactions on Industrial Informatics,* vol. 16, no. 7, pp. 4681-4690, July 2020.

[165] S. Shao, S. McAleer, R. Yan and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning," *IEEE Transactions on Industrial Informatics,* vol. 15, no. 4, pp. 2446-2455, April 2019.

[166] Y. Hou et al., "Bearing Fault Diagnosis Under Small Data Set Condition: A Bayesian Network Method With Transfer Learning for Parameter Estimation," *IEEE Access,* vol. 10, pp. 35768-35783, 2022.

[167] C. Liu and K. Gryllias, "Simulation-Driven Domain Adaptation for Rolling Element Bearing Fault Diagnosis," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 9, pp. 5760-5770, Sept. 2022.

[168] Z. Shen and W. Guo, "An Intelligent Bearing Fault Diagnosis based on Modified Probabilistic Knowledge Distillation," in *2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing)*, 2021.

[169] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate,," arXiv.org, 19 May 2016. [Online]. Available: https://arxiv.org/abs/1409.0473. [Accessed 13 Jul 2021].

[170] C. Blakeney, X. Li, Y. Yan and Z. Zong, "Parallel Blockwise Knowledge Distillation for Deep Neural Network Compression," *IEEE Transactions on Parallel and Distributed Systems,* vol. 32, no. 7, pp. 1765-1776, 1 July 2021.

[171] H. Mao, "A Survey on Self-supervised Pre-training for Sequential Transfer Learning in Neural Networks," arXiv.org, 2020. [Online]. Available: https://arxiv.org/abs/2007.00800. [Accessed 7 Sep 2021].

[172] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale,," arXiv.org, 03 June 2021. [Online]. Available: https://arxiv.org/abs/2010.11929.

[173] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," arXiv.org, 06 Dec 2017. [Online]. Available: https://arxiv.org/abs/1706.03762v5. [Accessed 16 Jul 2021].

[174] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 21 Jul 2016. [Online]. Available: https://arxiv.org/abs/1607.06450.

[175] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," arXiv.org, 11 Nov 2018. [Online]. Available: https://arxiv.org/abs/1606.08415v3.

[176] Z. -X. Hu, Y. Wang, M. -F. Ge and J. Liu, "Data-Driven Fault Diagnosis Method Based on Compressed Sensing and Improved Multiscale Network," *IEEE Transactions on Industrial Electronics,* vol. 67, no. 4, pp. 3216-3225, April 2020.

[177] 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, pp. 1066-1090, 2006.

[178] "Case Western Reserve University Bearing Data Center Website," [Online]. Available: https://engineering.case.edu/bearingdatacenter/.

[179] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," arXiv:1412.3474, 2014. [Online].

[180] Y. Ganin et al., "Domain-adversarial training of neural networks," *Journal of Machine Learning Research,* vol. 17, no. 59, pp. 1-35, 2016.

[181] C. Gong, Y. Hu, H. Wen, G. Chen, W. Li and J. Gao, "Reliable Winding-Based DC-Bus Capacitor Discharge Technique Over Full-Speed Range for IPMSM Drive in Electric Vehicles Without Position Sensor," *IEEE Transactions on Industrial Electronics,* vol. 67, no. 10, pp. 8131-8142, Oct. 2020.

[182] O. Wallscheid and J. Böcker, “Global identification of a low-order lumped-parameter thermal network for permanent magnet synchronous motors,” *IEEE Transactions on Energy Conversion,* vol. 31, no. 1, pp. 354-365, March 2016.

[183] O. Wallscheid and J. Bocker, "Design and identification of a lumped-parameter thermal network for permanent magnet synchronous motors based on heat transfer theory and particle swarm optimisation," in *2015 17th European Conference on Power Electronics and Applications (EPE’15 ECCE-Europe)*, 2015.

[184] A. Specht, O. Wallscheid and J. Böcker, "Determination of rotor temperature for an interior permanent magnet synchronous machine using a precise flux observer," in *Proc. Int. Power Electron. Conf.*, 2014.

[185] O. Wallscheid, A. Specht and J. Böcker, “Observing the Permanent-Magnet Temperature of Synchronous Motors Based on Electrical Fundamental Wave Model Quantities,” *IEEE Transactions on Industrial Electronics,* vol. 64, no. 5, pp. 3921-3929, May 2017.

[186] D. D. Reigosa, D. Fernandez, H. Yoshida, T. Kato and F. Briz, “Permanent-Magnet Temperature Estimation in PMSMs Using Pulsating High-Frequency Current Injection,” *IEEE Transactions on Industry Applications,* vol. 51, no. 4, pp. 3159-3168, July-Aug. 2015.

[187] D. Reigosa, D. Fernández, M. Martínez, J. M. Guerrero, A. B. Diez and F. Briz, "Magnet Temperature Estimation in Permanent Magnet Synchronous Machines Using the High Frequency Inductance," *IEEE Transactions on Industry Applications,* vol. 55, no. 3, pp. 2750-2757, May-June 2019.

[188] Z. Li, G. Feng, C. Lai, W. Li and N. C. Kar, “Current Injection-Based Simultaneous Stator Winding and PM Temperature Estimation for Dual Three-Phase PMSMs,” *IEEE Transactions on Industry Applications,* vol. 57, no. 5, pp. 4933-4945, Sept.-Oct. 2021.

[189] G. Feng, C. Lai, J. Tjong and N. C. Kar, "Noninvasive Kalman Filter Based Permanent Magnet Temperature Estimation for Permanent Magnet Synchronous Machines," *IEEE Transactions on Power Electronics,* vol. 33, no. 12, pp. 10673-10682, Dec. 2018.

[190] G. Feng, C. Lai, W. Li, M. Kelly, and N. C. Kar, "Simultaneous stator winding and permanent magnet temperature estimation for permanent magnet synchronous machines," in *2018 XIII International Conference on Electrical Machines (ICEM)*, 2018.

[191] D. E. G. Erazo, O. Wallscheid and J. Böcker, "Improved Fusion of Permanent Magnet Temperature Estimation Techniques for Synchronous Motors Using a Kalman Filter," *IEEE Transactions on Industrial Electronics,* vol. 67, no. 3, pp. 1708-1717, March 2020.

[192] E. G. Gedlu, O. Wallscheid and J. Böcker, "PERMANENT MAGNET SYNCHRONOUS MACHINE TEMPERATURE ESTIMATION USING LOW-ORDER LUMPED-PARAMETER THERMAL NETWORK WITH EXTENDED IRON LOSS MODEL," in *The 10th International Conference on Power Electronics, Machines and Drives (PEMD 2020)*, 2020.

[193] M. Ganchev, C. Kral, H. Oberguggenberger and T. Wolbank, "Sensorless rotor temperature estimation of permanent magnet synchronous motor," in *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, 2018-2023, 2011.

[194] W. Kirchgässner, O. Wallscheid and J. Böcker, "Estimating Electric Motor Temperatures With Deep Residual Machine Learning," *IEEE Transactions on Power Electronics,* vol. 36, no. 7, pp. 7480-7488, July 2021.

[195] W. Kirchgässner, O. Wallscheid and J. Böcker, "Data-Driven Permanent Magnet Temperature Estimation in Synchronous Motors With Supervised Machine Learning: A Benchmark," *IEEE Transactions on Energy Conversion,* vol. 36, no. 3, pp. 2059-2067, Sept. 2021.

[196] F. Masi, I. Stefanou, P. Vannucci, and V. Maffi-Berthier, "Thermodynamics-based artificial neural networks for Constitutive Modeling," *Journal of the Mechanics and Physics of Solids,* vol. 147, p. 104277, 2021.

[197] W. Kirchgässner, O. Wallscheid, and J. Böcker, "Thermal neural networks: Lumped-parameter thermal modeling with state-space machine learning,," *Engineering Applications of Artificial Intelligence,* vol. 117, p. 105537, 2023.

[198] E. G. Gedlu, O. Wallscheid and J. Böcker, "Temperature estimation of electric machines using a hybrid model of feed-forward neural and low-order lumped-parameter thermal networks," in *2021 IEEE International Electric Machines & Drives Conference (IEMDC)*, Hartford, CT, USA, 2021.

[199] J. Lee and J. -I. Ha, "Temperature Estimation of PMSM Using a Difference-Estimating Feedforward Neural Network," *IEEE Access,* vol. 8, pp. 130855-130865, 2020.

[200] P. Sharma, A. Vijayvargiya, B. Singh and R. Kumar, "Data Driven Temperature Estimation of PMSM with Regression Models," in *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, Raipur, India, 2022.

[201] C. Liu and K. Gryllias, "Simulation-Driven Domain Adaptation for Rolling Element Bearing Fault Diagnosis," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 9, pp. 5760-5770, Sept. 2022.

[202] 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, Sept. 2019.

[203] W. Mao, J. Chen, J. Liu and X. Liang, "Self-supervised Deep Domain-Adversarial Regression Adaptation for Online Remaining Useful Life Prediction of Rolling

Bearing under Unknown Working Condition," *IEEE Transactions on Industrial Informatics.*

[204] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of Neural Networks," arXiv.org, 2016 May 26. [Online]. Available: https://arxiv.org/abs/1505.07818.

[205] S. K. Chowdhury and P. K. Baski, "A simple lumped parameter thermal model for electrical machine of TEFC design," in *2010 Joint International Conference on Power Electronics, Drives and Energy Systems & 2010 Power India*, India, 2010.

[206] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," arXiv.org, 28 Mar 2021. [Online]. Available: https://arxiv.org/abs/2012.07436. [Accessed 03 Dec 2022].

[207] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv.org, 30 Jan 2017. [Online]. Available: https://arxiv.org/abs/1412.6980.

[208] "scipy.stats.wasserstein_distance," SciPy.org, 8 Jul 2019. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wasserstein_distance.html#scipy.stats.wasserstein_distance.

[209] Y. Hui, S. Wenzhu, Z. Xiuzhi, Z. Guotao and H. Wenting, "Heuristic sample reduction based support vector regression method," in *2016 IEEE International Conference on Mechatronics and Automation*, Harbin, China, 2016.

[210] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation,* vol. 9, no. 8, p. 1735–1780, 1997.

[211] S. Mohsenzade, "A Gate Driving Strategy for the Series-Connected IGBTs to Improve the Resilience Against IGBTs Short-Circuit Failures," *IEEE Transactions on Industrial Electronics,* vol. 69, no. 10, pp. 9961-9971, Oct. 2022.

[212] X. Chen, L. Wang, Y. Liu and Q. Wu, "Bang-Bang Funnel Control of Three-Phase Full-Bridge Inverter Under Dual-Buck Scheme," *IEEE Transactions on Industrial Electronics,* vol. 70, no. 6, pp. 5399-5409, June 2023.

[213] R. Li, C. Ho and C. Xu, "Active virtual ground - single phase transformerless grid-connected voltage source inverter topology," *IEEE Trans. Power Electron.,* vol. 33, no. 2, pp. 1335-1346, Feb. 2018.

[214] J. Kim, S. -W. Ryu, M. Kim and J. -W. Jung, "Triple-Mode Isolated Resonant Buck–Boost Converter Over Wide Input Voltage Range for Residential Applications," *IEEE Transactions on Industrial Electronics,* vol. 68, no. 11, pp. 11087-11099, Nov. 2021.

[215] Y. Xu, C. N. M. Ho, A. Ghosh and D. Muthumuni, "An Electrical Transient Model of IGBT-Diode Switching Cell for Power Semiconductor Loss Estimation in Electromagnetic Transient Simulation," *IEEE Transactions on Power Electronics,* vol. 35, no. 3, pp. 2979-2989, March 2020.

[216] J. Qi et al., "Temperature Dependence of Dynamic Performance Characterization of 1.2-kV SiC Power mosfets Compared With Si IGBTs for Wide Temperature Applications," *IEEE Transactions on Power Electronics,* vol. 34, no. 9, pp. 9105-9117, Sept. 2019.

[217] W. Wang, Z. Shen and V. Dinavahi, "Physics-Based Device-Level Power Electronic Circuit Hardware Emulation on FPGA," *IEEE Transactions on Industrial Informatics,* vol. 10, no. 4, pp. 2166-2179, Nov. 2014.

[218] C. Liu, H. Bai, S. Zhuo, X. Zhang, R. Ma and F. Gao, "Real-Time Simulation of Power Electronic Systems Based on Predictive Behavior," *IEEE Transactions on Industrial Electronics,* vol. 67, no. 9, pp. 8044-8053, Sept. 2020.

[219] H. Bai, C. Liu, A. K. Rathore, D. Paire and F. Gao, "An FPGA-Based IGBT Behavioral Model With High Transient Resolution for Real-Time Simulation of Power Electronic

Circuits," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 8, pp. 6581-6591, Aug. 2019.

[220] P. Górecki and D. Wojciechowski, “Accurate Electrothermal Modeling of High Frequency DC–DC Converters With Discrete IGBTs in PLECS Software,” *IEEE Transactions on Industrial Electronics,* vol. 70, no. 6, pp. 5739-5746, June 2023.

[221] Z. Wang and W. Qiao, “A Physics-Based Improved Cauer-Type Thermal Equivalent Circuit for IGBT Modules,” *IEEE Transactions on Power Electronics,* vol. 31, no. 10, pp. 6781-6786, Oct. 2016.

[222] R. Wu et al., “A Temperature-Dependent Thermal Model of IGBT Modules Suitable for Circuit-Level Simulations,” *IEEE Transactions on Industry Applications,* vol. 52, no. 4, pp. 3306-3314, July-Aug. 2016.

[223] R. Kraus and H. J. Mattausch, "Status and trends of power semiconductor device models for circuit simulation," *IEEE Transctions on Power Electronics,* vol. 13, no. 3, pp. 452-465, May 1998.

[224] A. R. Hefner and D. M. Diebolt, "An experimentally verified IGBT model implemented in the saber circuit simulator," *IEEE Transctions on Power Electronics,* vol. 9, no. 5, pp. 532-542, Sep. 1994.

[225] Y. Duan, F. Xiao, Y. Luo and F. Iannuzzo, "A lumped-charge approach based physical SPICE-model for high power soft-punch through IGBT," *IEEE Journal of Emerging and Selected Topics in Power Electronics,* vol. 7, no. 1, pp. 62-70, March 2019.

[226] K. Górecki and P. Górecki, "Nonlinear Compact Thermal Model of the IGBT Dedicated to SPICE," *IEEE Transactions on Power Electronics,* vol. 35, no. 12, pp. 13420-13428, Dec. 2020.

[227] J. L. Tichenor, S. D. Sudhoff and J. L. Drewniak, “Behavioral IGBT modeling for predicting high frequency effects in motor drives,” *IEEE Transctions on Power Electronics,* vol. 15, no. 2, pp. 354-360, Mar. 2000.

[228] A. D. Rajapakse, A. M. Gole and R. P. Jayasinghe, "An improved representation of FACTS controller semiconductor losses in EMTP-type programs using accurate loss-power injection into network solution," *IEEE Trans. Power Del.,* vol. 24, no. 1, pp. 381-389, Jan. 2009.

[229] A. D. Rajapakse, A. M. Gole and P. L. Wilson, "Electromagnetic transients simulation models for accurate representation of switching losses and thermal performance in power electronic systems," *IEEE Trans. Power Del.,* vol. 20, no. 1, pp. 319-327, Jan. 2005.

[230] H. Luo, W. Li and X. He, "Online High-Power P-i-N Diode Chip Temperature Extraction and Prediction Method With Maximum Recovery Current di/dt," *IEEE Transactions on Power Electronics,* vol. 30, no. 5, pp. 2395-2404, May 2015.

[231] J. Xu, H. Ke, Z. Chen, X. Fan, T. Peng and C. Yang, "Oversmoothing Relief Graph Convolutional Network-Based Fault Diagnosis Method With Application to the Rectifier of High-Speed Trains," *IEEE Transactions on Industrial Informatics,* vol. 19, no. 1, pp. 771-779, Jan. 2023.

[232] A. Vaswani et al., "Attention is all you need," arXiv.org, 06 Dec 2017. [Online]. Available: https://arxiv.org/abs/1706.03762.

[233] K. Jia, Z. Yang, L. Zheng, Z. Zhu and T. Bi, "Spearman Correlation-Based Pilot Protection for Transmission Line Connected to PMSGs and DFIGs," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 7, pp. 4532-4544, July 2021.

[234] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A Deep Learning Framework for traffic forecasting," arXiv.org, 12 Jul. 2018. [Online].

[235] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," arXiv.org, 22 February 2017. [Online]. Available: https://arxiv.org/abs/1609.02907.

[236] J. S. L. Senanayaka, H. Van Khang and K. G. Robbersmyr, “Toward Self-Supervised Feature Learning for Online Diagnosis of Multiple Faults in Electric Powertrains,” *IEEE Transactions on Industrial Informatics,* vol. 17, no. 6, pp. 3772-3781, June 2021.

[237] J. Jung et al., “Monitoring Journal-Bearing Faults: Making Use of Motor Current Signature Analysis for Induction Motors,” *IEEE Industry Applications Magazine,* vol. 23, no. 4, p. 12–21, Jul. 2017.

[238] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” arXiv.org, 06 Dec 2017. [Online]. Available: https://arxiv.org/abs/1706.03762.

END