



# Model-Based Deep Transfer Learning Method to Fault Detection and Diagnosis in Nuclear Power Plants

Yuantao Yao<sup>1</sup>, Daochuan Ge<sup>1\*</sup>, Jie Yu<sup>1</sup> and Min Xie<sup>2</sup>

<sup>1</sup>Institute of Nuclear Energy Safety Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China,

<sup>2</sup>Department of Advanced Design and System Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR, China

## OPEN ACCESS

### Edited by:

Wenxi Tian,  
Xi'an Jiaotong University, China

### Reviewed by:

Xingang Zhao,  
Oak Ridge National Laboratory (DOE),  
United States

Shripad T. Revankar,  
Purdue University, United States

### \*Correspondence:

Daochuan Ge  
daochuan.ge@inest.cas.cn

### Specialty section:

This article was submitted to  
Nuclear Energy,  
a section of the journal  
Frontiers in Energy Research

**Received:** 27 November 2021

**Accepted:** 07 February 2022

**Published:** 02 March 2022

### Citation:

Yao Y, Ge D, Yu J and Xie M (2022)  
Model-Based Deep Transfer Learning  
Method to Fault Detection and  
Diagnosis in Nuclear Power Plants.  
Front. Energy Res. 10:823395.  
doi: 10.3389/fenrg.2022.823395

Deep learning-based nuclear intelligent fault detection and diagnosis (FDD) methods have been widely developed and have achieved very competitive results with the progress of artificial intelligence technology. However, the pretrained model for diagnosis tasks is hard in achieving good performance when the reactor operation conditions are updated. On the other hand, retraining the model for a new data set will waste computing resources. This article proposes an FDD method for cross-condition and cross-facility tasks based on the optimized transferable convolutional neural network (CNN) model. First, by using the pretrained model's prior knowledge, the model's diagnosis performance to be transferred for source domain data sets is improved. Second, a model-based transfer learning strategy is adopted to freeze the feature extraction layer in a part of the training model. Third, the training data in target domain data sets are used to optimize the model layer by layer to find the optimization model with the transferred layer. Finally, the proposed comprehensive simulation platform provides source and target cross-condition and cross-facility data sets to support case studies. The designed model utilizes the strong nonlinear feature extraction performance of a deep network and applies the prior knowledge of pretrained models to improve the accuracy and timeliness of training. The results show that the proposed method is superior to achieving good generalization performance at less training epoch than the retraining benchmark deep CNN model.

**Keywords:** fault detection and diagnosis, deep learning, transfer learning, freezing and fine-tuning strategy, nuclear power plants

## INTRODUCTION

No matter how advanced the energy systems have progressed with state-of-the-art techniques, operation safety and reliability will be a central research topic all the time. Especially for nuclear systems, safeguards are even more critical and cannot be ignored (Perrault, 2019; Matteo et al., 2021; Yao et al., 2021). Most of the severe nuclear leakage events throughout the history of humankind have been caused by operators' inappropriate responses and solutions. Therefore, it is critical to provide administrators with auxiliary information under different nuclear system operation conditions before an accident worsens (Wahlström, 2018; Yoo et al., 2018).

One predictive maintenance approach that has become increasingly valued is fault detection and diagnosis (FDD), to judge (detection) and identify (diagnosis) the type of fault (Yangping et al., 2000; Ma and Jiang, 2011). According to the review work from Zhao et al. (2021), the development of fault diagnosis in nuclear power plants (NPPs) mainly goes through three essential stages: the model-

based method, knowledge-based rule method, and currently popular data-driven method. Model-based approaches fall into two main categories. One is through the statistical anomaly (fault) and average state residuals, such as parity check, wavelet transform, and time-frequency analysis of quantitative models (Zhong et al., 2018). The other is a qualitative analysis based on the physical or graph structure models. Besides, the rule-based approaches are an essential branch of fault diagnosis research, which by triggering specific “if-then” rules to determine results related to measured/detected fault symptoms. The rule libraries are developed using expert judgment and prior knowledge of systems; the most famous rule of which is the fuzzy rule (Xu et al., 2019).

However, modern industrial systems have a nonlinear, considerable time delay, uncertainty factors, which makes it challenging to build precise mathematical models. Therefore, the application of model-based and rule-based methods is limited.

The data-driven method does not require prior knowledge of the object system (mathematical model or expert experience). It takes the monitoring data as the research object to estimate the state of the target system, avoiding the shortcomings of the physical model-based method. Feature extraction and classifier design are two main parts of this method. The popular method in feature extraction is principal component analysis, which reduces the dimension of the data and extracts critical information (Peng and Wang, 2018). The famous classifiers mainly include Support Vector Machine (Yao et al., 2020a), Extreme Learning Machine (Zheng et al., 2019), and artificial neural network (Xin et al., 2019).

With the advances of the Internet of Things, big data, and the continuous improvement of equipment scale (Lee et al., 2014; Wang et al., 2015; Seabra et al., 2016), deep learning-based fault diagnosis methods such as the automatic deep encoder-decoder (Wang and Zhang, 2018), deep belief networks (DBNs) (Tang et al., 2018), and deep convolutional neural networks (CNNs) (Wen et al., 2017) are gradually coming into view. Compared with traditional data-driven methods, high integrated and end-to-end deep networks with multiple hidden layers can learn and fit any nonlinear relationship under sufficient training data, widely applied in different energy systems.

Correa-Jullian et al. (2020) discussed several deep networks-based methods. They applied them to prognose the performance of solar hot water systems under different meteorological conditions. Xu et al. (2020) combined the CNNs with the variational mode decomposition algorithms to accomplish the fault diagnosis of the rolling bearing of wind turbines. Guo et al. (2018) proposed a fault diagnosis approach using a DBN with a model optimization strategy for building energy saving. In nuclear systems, Peng et al. (2018) utilized the feature selection capability of correlation analysis for dimensionality reduction and DBNs for fault identification. Saeed et al. (2020) proposed a fault diagnosis model based on the deep hybrid networks to achieve FDD with different levels. Mandal et al. (2017) introduced a DBN-based detection and diagnosis method for the thermocouple sensor fault. Yao et al. (2022) presented a residual CNN with an adaptive noise elimination procedure for the FDD in small modular reactors.

The designed deep network-based model can learn features from the original data and have overwhelming advantages in solving various fault classification problems. On the other hand, the end-to-end deep learning model has integration advantages compared with traditional machine learning-based manual feature engineering selection. When training diagnostic models based on data-driven methods, we usually default to the same training and test data distribution. Suppose that there are enough training samples for the fault diagnosis task of a given scene. In that case, the model nonlinear relationship can be fitted through parameter optimization to achieve a high-precision diagnosis.

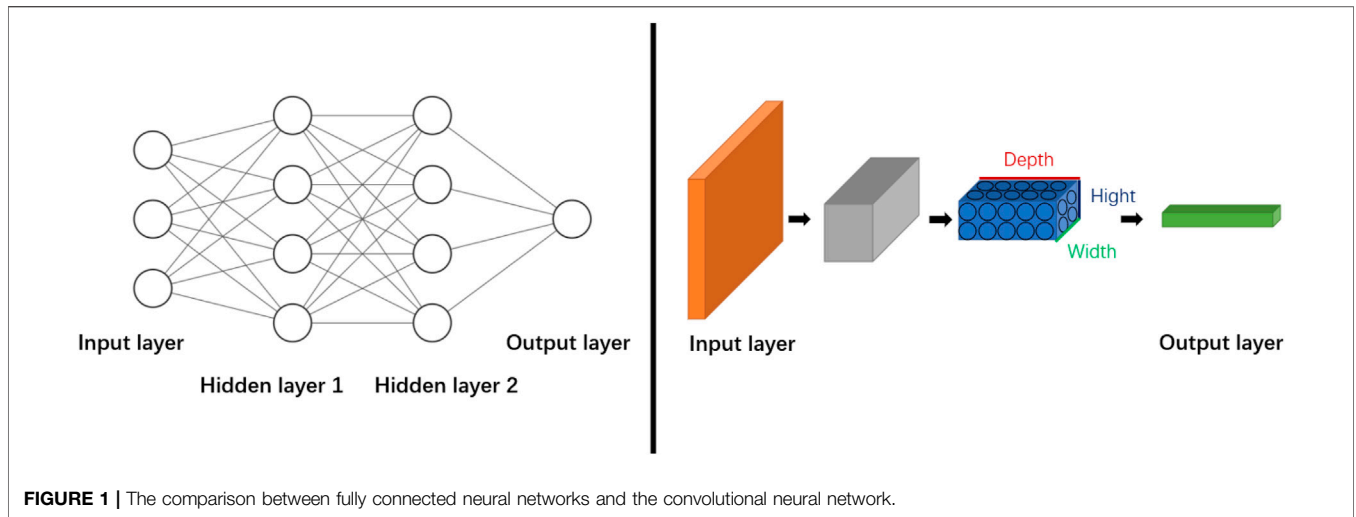
However, in the actual system operation process, the nuclear systems are in a stable operation state most of the time. The difficulty of obtaining fault data leads to a small number of samples, and most faults are in an unmarked condition. Traditional deep learning models such as CNNs and DBNs will overfit when training on small sample data sets and significantly reduce diagnostic performance. On the other hand, the historical training model fails to identify new data once the system runs under different conditions or upgraded or updated environment.

The transfer learning (TL) method is proposed to solve the above problems. It transfers the knowledge learned from the neighboring domain to improve learning performance under insufficient target task training data. In recent years, it has been developed and applied in natural language processing, computer vision, and autonomous driving (Ruder et al., 2019; Zhuang et al., 2020). Furthermore, to apply TL in the FDDs is to relax the constraint that targets domain data, and the source domain data must obey the same distribution. It will reduce the urgency of collecting massive data combined with TL. At present, relevant research in energy is scarce. For nuclear systems, TL-based fault diagnosis, the initial exploration, has not been involved.

To address the above problems, we propose a diagnosis framework based on transferable CNN models to make full use of the prior knowledge of the pretraining model. Compared with the traditional deep network-based diagnosis framework, the proposed method has the following advantages:

- 1) A novel freezing and tuning transfer strategy based on a pretraining model can be applied in the nuclear systems' fault diagnosis under different operating conditions and equipment.
- 2) The proposed method does not need to train and optimize the parameters of each layer of the model one by one. Still, it makes full use of the high-dimensional feature extraction capability of the pretraining model for source domain data.
- 3) The proposed method significantly reduces the model retraining time. Under insufficient data, avoiding data expansion technology minimizes the probability of model overfitting and improves the training performance.
- 4) The proposed method has good portability. After simple optimization for different target transfer environments, future research can achieve ideal results.

The remainder of the article is organized as follows: *Vanilla CNN Structure* briefly introduces the vanilla CNN structure. *Proposed Method* proposes a TL-based fault diagnosis



procedure. The case study is presented in *Case Study*. Results and discussions are shown in *Results and Discussion*. *Conclusion* concludes the article and makes an outlook for the future work.

### VANILLA CONVOLUTIONAL NEURAL NETWORK STRUCTURE

CNNs (LeCun et al., 2015) are traditional deep networks commonly used in classification research. They are mainly based on the feedforward networks to add the corresponding convolution operation process to extract the high-dimensional characteristics of the input data to be analyzed. From the initial application in speech and image recognition to the current machinery, energy, aerospace, and other fields of abnormal detection, fault diagnosis, time-series prediction, and other applications have a wide range of prospects. Compared with traditional neural networks, neurons in the CNN structure are arranged in three dimensions, as is shown in **Figure 1**.

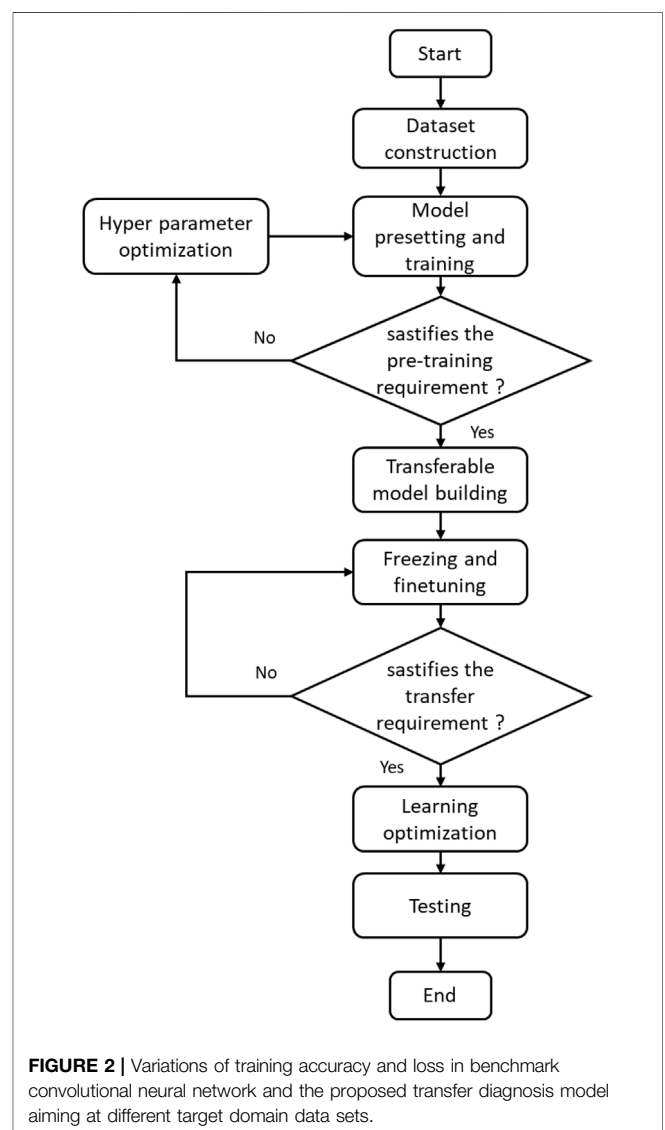
The neurons in the layers are not fully connected traditionally but are only related to a small area of the previous layer. The CNN structure is mainly composed of three essential parts: the convolutional layer, pooling layer, and fully connected layer, which are executed to make feature extraction, filtering, and output with nonlinear combination, respectively.

### PROPOSED METHOD

#### Brief Structure

**Figure 2** shows the scheme of the proposed TL-based diagnosis framework. It mainly contains four critical steps: data set construction, model presetting, transferable model construction, and model optimization and testing.

First, the source and target domains of the research data sets are constructed through the existing comprehensive experimental platform. The source domain data comes from the previous data, and the target domain data comes from different working



**TABLE 1** | Pretrained convolutional neural network model structure.

| Layer no | Module no | Layer type  | Kernel size | Output size     | Stride | Padding |
|----------|-----------|-------------|-------------|-----------------|--------|---------|
| 1        |           | Input       | —           | 200 * 316 * 3   | —      | —       |
| 2        | M 4       | Convolution | 128         | 100 * 158 *,128 | 2      | Yes     |
| 3        |           | BN          | —           | 100 * 158 *,128 | —      | —       |
| 4        |           | ReLU        | —           | 100 * 158 *,128 | —      | —       |
| 5        |           | Max pooling | 128         | 50 * 79 *,128   | —      | —       |
| 6        | M 3       | Convolution | 32          | 25 * 40 * 32    | 2      | Yes     |
| 7        |           | BN          | —           | 25 * 40 * 32    | —      | —       |
| 8        |           | ReLU        | —           | 25 * 40 * 32    | —      | —       |
| 9        |           | Max pooling | 32          | 13 * 20 * 32    | —      | —       |
| 10       | M 2       | Convolution | 64          | 7 * 10 * 64     | 2      | Yes     |
| 11       |           | BN          | —           | 7 * 10 * 64     | —      | —       |
| 12       |           | ReLU        | —           | 7 * 10 * 64     | —      | —       |
| 13       |           | Max pooling | 64          | 4 * 5 * 64      | —      | —       |
| 14       | M 1       | Dense       | —           | 256             | —      | —       |
| 15       |           | Dense       | —           | 128             | —      | —       |
| 16       |           | Softmax     | —           | 11              | —      | —       |

conditions or new reactor types. Second, the CNN-based model designed initially is used to complete sufficient training and hyperparameter optimization for the source domain data. Third, a part of the data set in the target domain data set is selected for the transferred model training, which is completed by freezing and tuning some layer parameters. Finally, the hyperparameter tuning is performed on the transferred model, and the target domain data sets are adopted to test the diagnostic performance of the proposed method. The detailed procedure of model presetting and the transferable model building will be introduced.

## Model Presetting

The basic structure of the transfer model refers to the previous design, which is modified compared with the vanilla CNN structure (Yao et al., 2020b). It has one input layer, three convolutional layers, three max-pooling layers, and two fully connected layers. Meanwhile, the layers in the network are grouped and divided into specific modules to discuss the influence of different model parts on transfer performance during freezing and tuning. The detailed introduction of the model structure is shown in **Table 1**. On this basis, combined with the state information imaging method, the network input is a two-dimensional matrix image. Furthermore, the number of the convolution kernel is set as a more considerable number (128) to enhance the network feature learning ability. The number of the following convolution kernels is settled as 32 and 64, respectively.

The criterion sets the number of convolution kernels from the least to the most. It gradually increases in multiple relationships to obtain more discriminative features at the higher level of the network. In addition, zero padding is used to make the feature output before and after the convolution operation maintain the same size. Moreover, batch normalization is adopted to avoid network overfitting, thereby maximizing saving the original input information. For the pooling layer parameter setting, the number of feature maps in each component block is the same as that in the convolution layer, verified in the previous work. The Softmax function is selected as the classification function. The Adam

optimization method (Kingma and Ba, 2014) is used to make gradient updating, introduced in reference Kingma and Ba (2014) in detail.

## Transferable Model Building

The transferable model is based on a fully pretrained CNN-based diagnosis model for the source domain data set. The output of the Softmax function should be replaced by the number of samples in the target domain when constructing the forward transfer model for samples in the target domain. For the hyperparameters of other network layers, layer-by-layer freezing and optimization are adopted, as is shown in **Figure 3**. The detailed procedure is as follows:

**Step 1:** Replace the output fault categories in the Softmax function according to the target task category.

**Step 2:** Adopt the target domain data training in a small sample environment, freeze the parameters of the pre-sequence network layer.

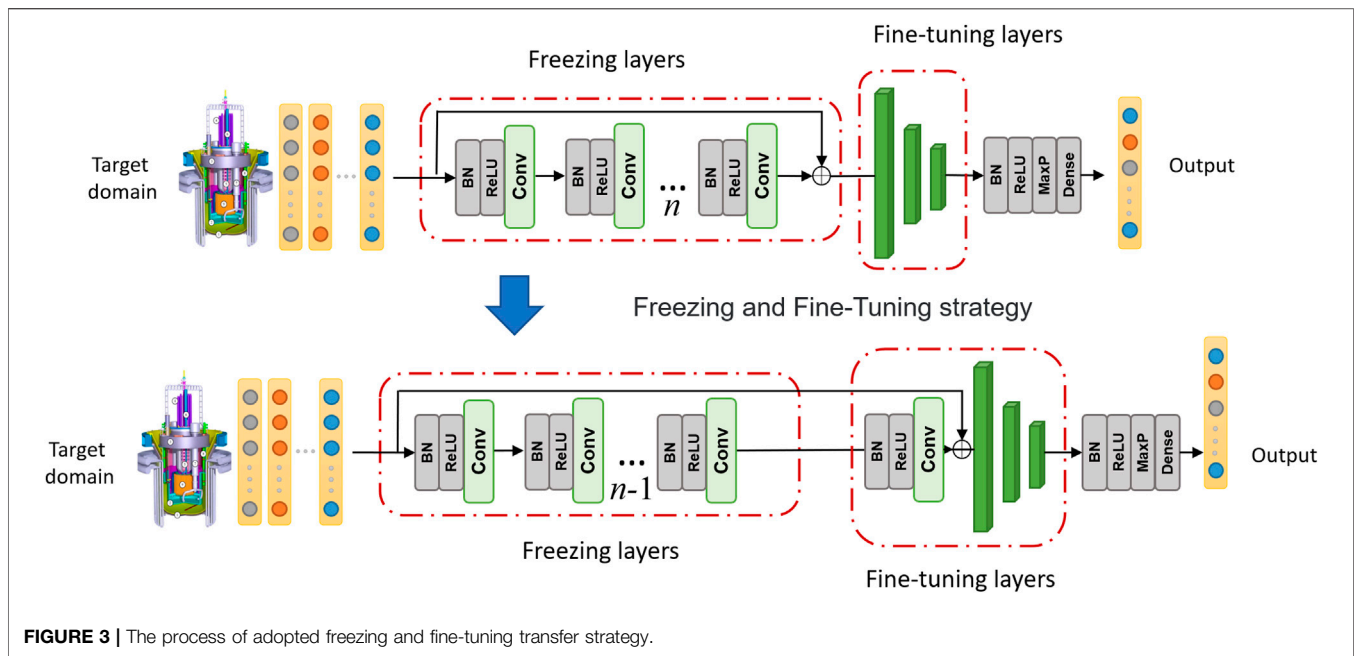
**Step 3:** Adopt a small learning rate design to tune the subsequent connection layer to realize the generalization transfer of the network.

**Step 4:** Reduce the number of frozen, fixed layers and move them to the transfer connection layer. Similarly, a small learning rate is adopted to tune the transfer connection layer, and the whole process is shown in **Figure 3**.

**Step 5:** The test data set samples are substituted into the model to obtain its discriminant classes, and the diagnostic performance of the model is tested.

In the *Proposed Method*, Part A, we have selected the basic structural hyperparameters of the CNN model, which includes the step size, number of kernels, batch size, etc. However, the influence of the learning rate on model training is not to be further considered. When using the gradient descent algorithm, if the learning rate is too low, the convergence of the model will be slow. Furthermore, more epochs are needed to complete the training, thus wasting computational power.

On the other hand, if the setting is too large, the model will not converge, reducing the model's diagnostic performance. The



learning rate decay (LDR) is a commonly used method in deep network training. Although the adaptive gradient descent algorithm (Adam) is adopted in the pretraining model to optimize the updating strategy of the gradient, the LDR is still applicable to the training of the model.

Therefore, to improve the efficiency of freezing and fine-tuning the procedure, we divided the training epoch into several parts and adopted an exponential LDR strategy in each part to enhance the training effect of the model, which is as follows:

$$\alpha = D_p^{\text{epoch}} * \alpha_0 \quad (1)$$

where  $D_p$  represents the LDR in each epoch.  $P$  represents the number of division parts.  $\alpha_0$  represents the initial value of the learning rate.

## CASE STUDY

### Research Platform

Since 2011, the project to build the accelerator-driven systems for nuclear waste transmutation has been developed and researched by the Chinese Academy of Science, including three key stages. Gen-IV China LEAd-based Reactor (CLEAR) was proposed as the reference for the above project (Wu, 2016a).

And then, to test the 1:1 prototype key component and verify the thermal hydraulic performance of the designed CLEAR-I, the integrated nonnuclear test facility CLEAR-S was built commissioning at the end of 2017 (Wu, 2016b). In 2018, the basis experimental hardware system CLEAR assistant simulator was made for the neutron transportation simulation, structure engineering design, and accident security analysis. It utilizes a computerized man-machine interface and digitalized instrumentation and control system. We are currently

conducting research and analysis based on nonnuclear test devices and a hardware-in-the-loop simulation experimental platform, which provides data support for further study.

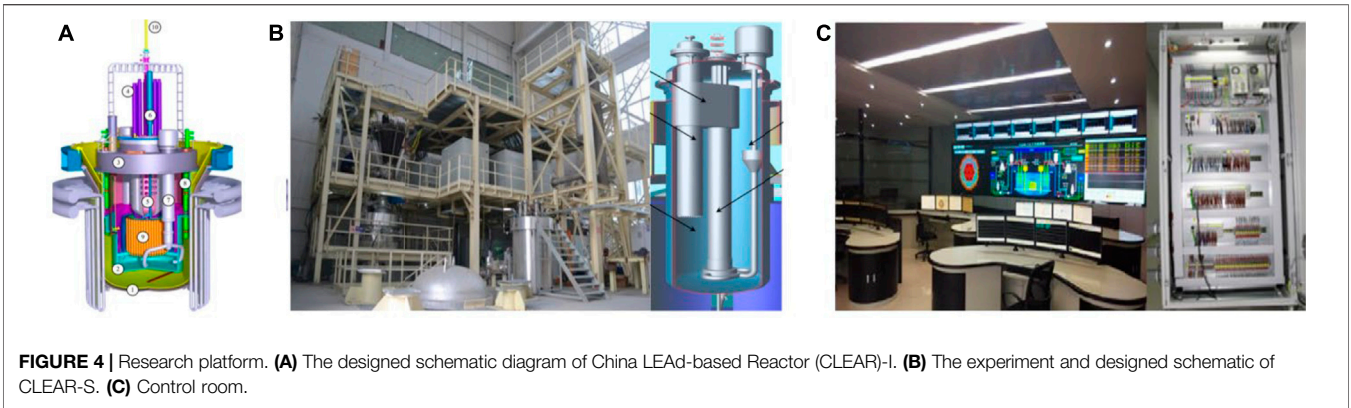
### Data Set Description

The data set in this study comes from the CLEAR-I and CLEAR-S accident simulation data stored in the previous research on the simulation experiment platform with the RELAP5 simulation calculation program-based core, which is shown in Figures 4, 5.

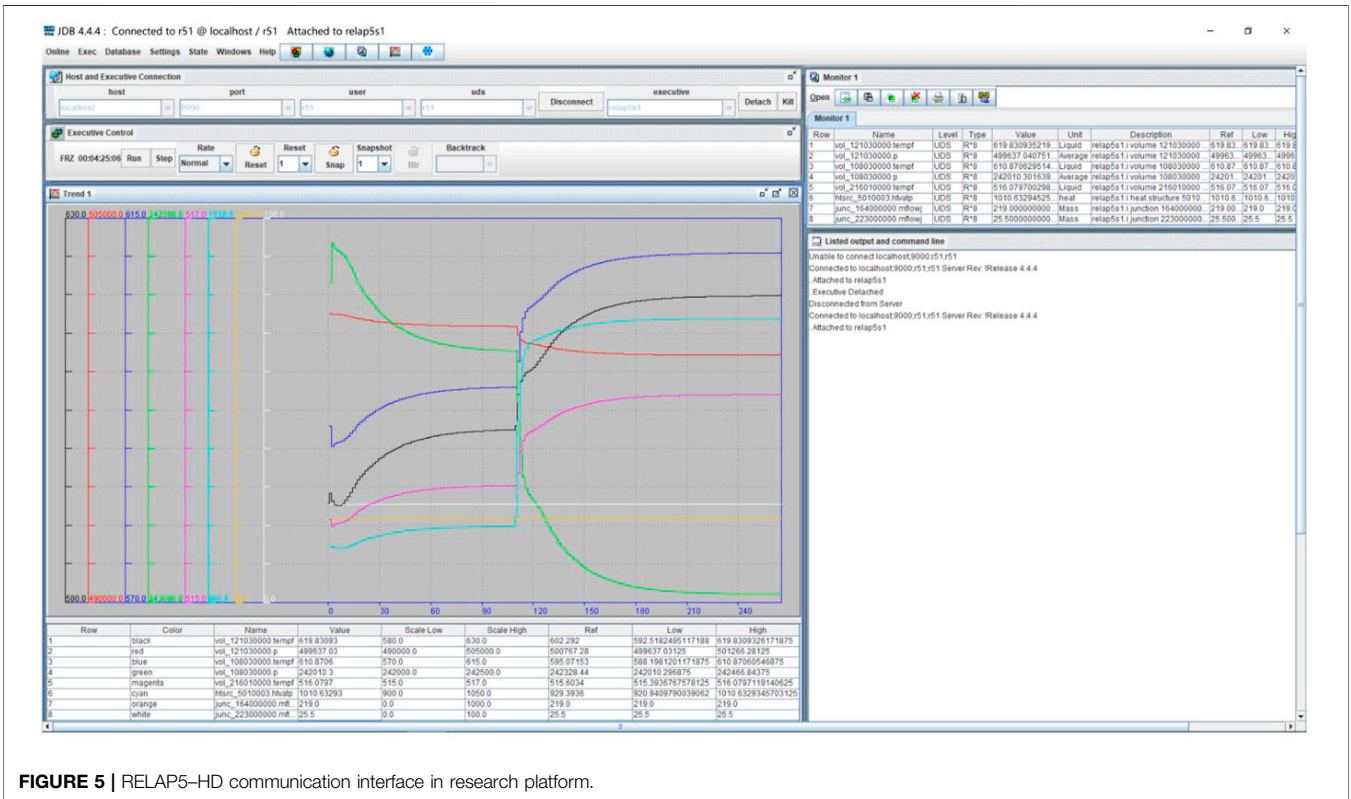
The RELAP program is a thermo-hydraulic program developed by Idaho National Laboratory to simulate a transient accident in a light water reactor (Li et al., 2014). It is a one-dimensional transient, two-phase fluid, six hydrodynamic equation and one-dimensional heat conduction. Moreover, point reactor dynamics models are widely used in NPP's accident safety analysis, accident evaluation, experimental analysis, and other fields. RELAP5-HD is a new version developed based on RELAP5. Its most significant feature is that it highly integrates the functions of RELAP5-HD and its three-dimensional (3D) thermo-hydraulic and neutron dynamics modeling capabilities, which can achieve more accurate 3D reactor construction. It can meet the real-time simulation requirements of the simulator. It can be adopted as a simulation program for the thermal-hydraulic system of the reactor simulator.

### Source Domain Data Set

The source domain data set  $D_s$  is acquired from the CLEAR-I operation data. In the study, 10 typical operation scenarios are included which contain one scenario under standard steady scenario [100% rate full power (RFP)], two power step scenarios (from 100 to 120% and 150% RFP), a severe accident scenario (loss of coolant with a small break), and six scenarios of rotating machinery of component faults in the fan or



**FIGURE 4 |** Research platform. (A) The designed schematic diagram of China LEAd-based Reactor (CLEAR)-I. (B) The experiment and designed schematic of CLEAR-S. (C) Control room.



**FIGURE 5 |** RELAP5-HD communication interface in research platform.

a pump. **Table 2** shows the detailed information about each operation scenario in  $D_s$ .

Each scenario starts instantaneously except for scenario 1 (steady-state). The corresponding transient operation or fault is introduced from  $t = 0$  to track the development trend. The sampling time set in the program is 0.25 s, and the sample length is 200, i.e., a total of 50 s. Each scenario contains 100 data samples, 50% of which are obtained through data enhancement. The method adopted is the sliding window method introduced in a previous work (Yao et al., 2020b). Each sample consists of 316 monitoring points from different components. The 316

monitoring points are distributed in 17 key node parameters, which are detailed as follows:

**Core (5):** Reactivity, power, control rod position, core temperature, and flow rate.

**Steam Generator (6):** SG primary side temperature, flow rate, and pressure; SG secondary side temperature, flow rate, and pressure.

**Pump (2):** Main pump flow rate and secondary pump flow rate.

**Fan (4):** Primary temperature, flow rate, and secondary temperature and flow rate.

**TABLE 2** | China LEAd-based Reactor-I source domain data set.

| Data set type  | Scenario no | Scenario description                         | Data size      |
|----------------|-------------|--|----------------|
| D <sub>s</sub> | 1           | Steady-state with 100% rated full power      | 200 * 316 *100 |
|                | 2           | 20% increasing power steps from steady-state | 200 * 316 *100 |
|                | 3           | 50% increasing power steps from steady-state | 200 * 316 *100 |
|                | 4           | LOCA with the small break                    | 200 * 316 *100 |
|                | 5           | One of the main pump rotor seizures          | 200 * 316 *100 |
|                | 6           | Two main pump rotor seizure                  | 200 * 316 *100 |
|                | 7           | One of the feed pump rotor seizures          | 200 * 316 *100 |
|                | 8           | Two feed pump rotor seizure                  | 200 * 316 *100 |
|                | 9           | Fan rotor seizures                           | 200 * 316 *100 |
|                | 10          | Fan speed decreases by 10%                   | 200 * 316 *100 |
|                | 11          | Fan speed decreases by 30%                   | 200 * 316 *100 |

LOCA, loss of coolant accident.

**TABLE 3** | China LEAd-based Reactor (CLEAR)-I source domain data set.

| Data set type   | Reactor type | Scenario label | Scenario description                           | Data size         |
|-----------------|--------------|----------------|--|-------------------|
| D <sub>t1</sub> | CLEAR-I      | 1              | Steady state with 120% rated full power        | 200 * 316 * 80    |
|                 |              | 2              | 20% increasing power steps from steady state   | 200 * 316 * 80    |
|                 |              | 3              | LOCA with the small break                      | 200 * 316 * 80    |
|                 |              | 4              | One primary pump rotor seizure                 | 200 * 316 * 80    |
|                 |              | 5              | Double primary pump rotor seizures             | 200 * 316 * 80    |
|                 |              | 6              | One secondary pump rotor seizure               | 200 * 316 * 80    |
|                 |              | 7              | Double secondary pump rotor seizures           | 200 * 316 * 80    |
|                 |              | 8              | Fan rotor seizures                             | 200 * 316 * 80    |
|                 |              | 9              | Fan speed decreases by 30%                     | 200 * 316 * 80    |
| D <sub>t2</sub> | CLEAR-I      | 1              | Steady state with 80% or 120% rated full power | 200 * 316 * 40&40 |
|                 |              | 2              | 20% increasing power steps from steady state   | 200 * 316 * 40&40 |
|                 |              | 3              | LOCA with the small break                      | 200 * 316 * 40&40 |
|                 |              | 4              | One primary pump rotor seizure                 | 200 * 316 * 40&40 |
|                 |              | 5              | Double primary pump rotor seizures             | 200 * 316 * 40&40 |
|                 |              | 6              | One secondary pump rotor seizure               | 200 * 316 * 40&40 |
|                 |              | 7              | Double secondary pump rotor seizures           | 200 * 316 * 40&40 |
|                 |              | 8              | Fan rotor seizures                             | 200 * 316 * 40&40 |
|                 |              | 9              | Fan speed decreases by 30%                     | 200 * 316 * 40&40 |

LOCA, loss of coolant accident.

## Target Domain Data Sets

The target domain data sets in this study are constructed according to specific transfer tasks. Two different target data sets for the cross-condition and cross-facility mission are built in the case study to verify the validity of the proposed model-based transfer method.

The target domain data sets D<sub>t1</sub> and D<sub>t2</sub> are settled to evaluate the ability of the network's transfer and generalization capability under different steady states. D<sub>t1</sub> shares the same reactor type (CLEAR-I) as D<sub>s</sub> but operates at a much higher steady-state power (120% RFP). The sample size in each failure case is 30% of the source domain data set, but the data sample's length is the same. D<sub>t2</sub>, based on D<sub>t1</sub>, improves the difficulty of cross-operating conditions transfer, including 80 and 120% RFP steady-state operating conditions of the reactor. Because the monitored parameters will be changed with different operating conditions under accident, the diagnosis task for D<sub>t2</sub> will be more challenging than that of D<sub>t1</sub>. To ensure the comparability of experimental results, the total number of samples between two target data sets

is unified. **Table 3** shows details of the cross-condition target domain data set.

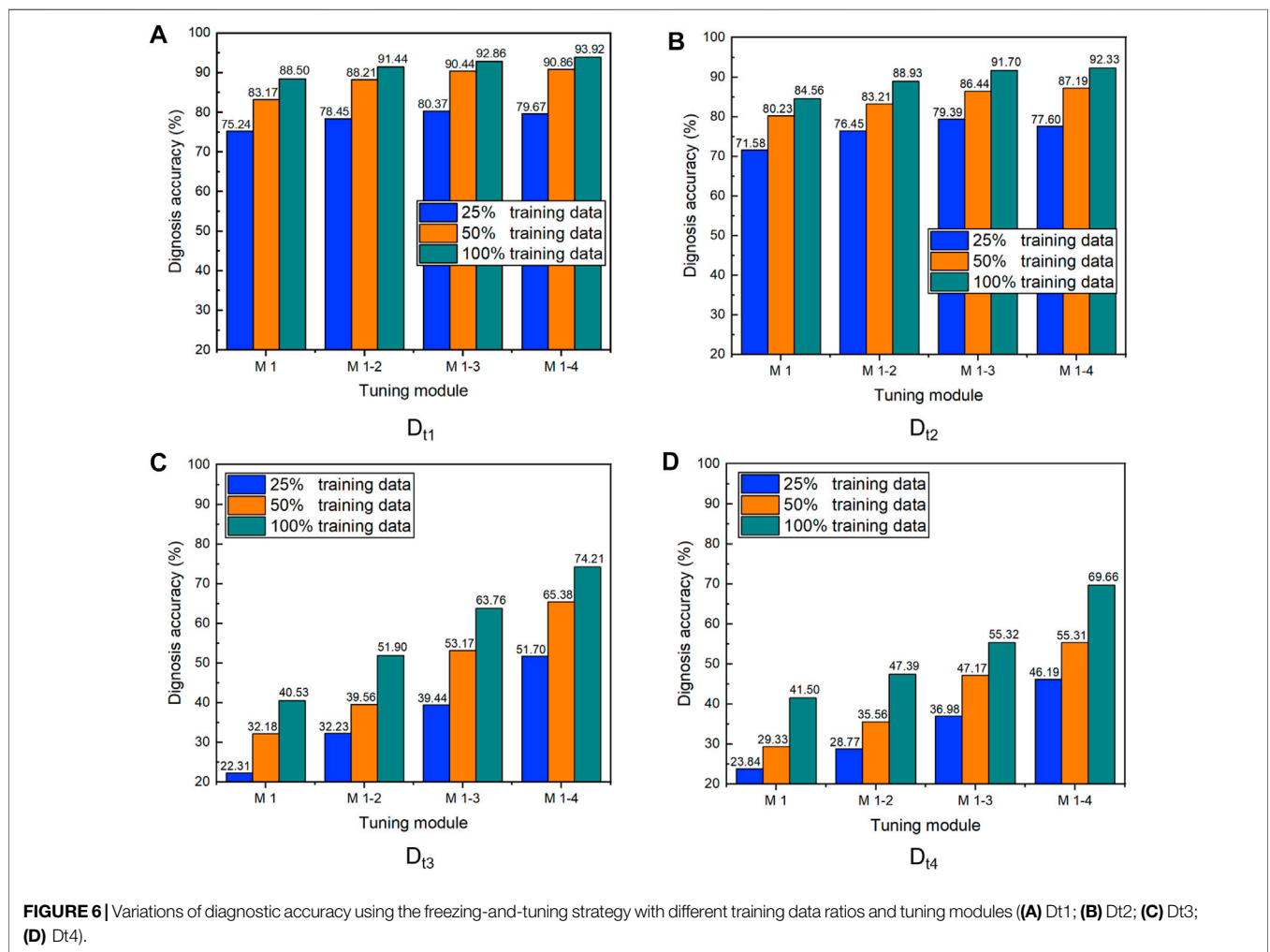
D<sub>t3</sub> and D<sub>t4</sub> are constructed for cross-facility transfer tasks built from a completely different reactor (CLEAR-S), as is shown in **Table 4**. The steady-state and power step operation data were derived from the experimental data on the existing facility. Compared with cross-condition data sets, cross-facility data sets differ significantly from source domain data in distribution characteristics, improving transfer complexity. The purpose of building this type of target domain data set is to explore the ability of the proposed transfer strategy between different facilities. It is worth noting that there are differences in structure between CLEAR-S and CLEAR-I nodes (CLEAR-I nodes are more precisely divided). To ensure that the cross-facility data set is dimensionally the same as the source data, we use zero padding for the default nodes to unify the data dimension, avoiding data heterogeneity.

To be similar to the procedure in the cross-condition transfer task, when compared with D<sub>t3</sub>, D<sub>t4</sub> is replaced with the data in

**Table 4** | Cross-facility target domain data set.

| Data set type   | Reactor type | Scenario label | Scenario description                            | Data size         |
|-----------------|--------------|----------------|---|-------------------|
| D <sub>t3</sub> | CLEAR-S      | 1              | Steady state with 100% rated full power         | 200 * 316 * 40    |
|                 |              | 2              | 20% increasing power steps from steady state    | 200 * 316 * 40    |
|                 |              | 3              | 50% increasing power steps from steady state    | 200 * 316 * 40    |
|                 |              | 4              | LOCA with the small break                       | 200 * 316 * 40    |
|                 |              | 5              | Main pump rotor seizures                        | 200 * 316 * 40    |
|                 |              | 6              | Secondary pump rotor seizures                   | 200 * 316 * 40    |
|                 |              | 7              | Fan rotor seizures                              | 200 * 316 * 40    |
|                 |              | 8              | Fan speed decreases by 10%                      | 200 * 316 * 40    |
|                 |              | 9              | Fan speed decreases by 30%                      | 200 * 316 * 40    |
| D <sub>t4</sub> | CLEAR-S      | 1              | Steady state with 100% or 120% rated full power | 200 * 316 * 20&20 |
|                 |              | 2              | 20% increasing power steps from steady state    | 200 * 316 * 20&20 |
|                 |              | 3              | 50% increasing power steps from steady state    | 200 * 316 * 20&20 |
|                 |              | 4              | LOCA with the small break                       | 200 * 316 * 20&20 |
|                 |              | 5              | Main pump rotor seizures                        | 200 * 316 * 20&20 |
|                 |              | 6              | Secondary pump rotor seizures                   | 200 * 316 * 20&20 |
|                 |              | 7              | Fan rotor seizures                              | 200 * 316 * 20&20 |
|                 |              | 8              | Fan speed decreases by 10%                      | 200 * 316 * 20&20 |
|                 |              | 9              | Fan speed decreases by 30%                      | 200 * 316 * 20&20 |

LOCA, loss of coolant accident.





120% operating conditions to increase the transfer complexity. It is more challenging to transfer the model under different steady-state operating conditions simultaneously under cross-facility conditions.

## RESULTS AND DISCUSSION

To highlight the effectiveness of the proposed transfer model, we discussed and analyzed the screening results of model layers. The comparison results between transfer and non-transfer models. We also considered the final training and optimization results for cross-condition and cross-facility transfer tasks. The training procedure of the proposed transfer model was implemented in Keras v2.2.4 on Microsoft Windows 10 operating system based on an Intel Core i7-10750 2.6 GHz CPU with 64 GB RAM and accelerated by the Cuda v11.1 environment with NVIDIA RTX 3070 GPU.

### Freezing and Fine-Tuning Result

To discuss the influence of the ratio of training samples on the transfer model's freezing-and-tuning process, we further divide  $D_{t1}$  to  $D_{t4}$  into three parts, which include 25, 50, and 100% training samples. Meanwhile, the training and test samples' ratio is the same as the pretraining model in the source domain, set as 4:1. All training and testing sites were cross-validated by a percentage of five folds to ensure accuracy. Meanwhile, all the test results are averaged 10 times.

**Figures 6A,B** show the influence of different training data ratios and tuning modules on the diagnosis accuracy in cross-working condition data sets ( $D_{t1}$  to  $D_{t2}$ ). The horizontal axis shows the included tuning modules. For example, M 1 represents that only the last fully connected layer is tuned. At the same time, M 1–4 illustrate that all modules from 1 to 4 are adjusted layer by layer. The vertical axis shows the diagnostic accuracy of the target domain test data set.

It can be concluded that the diagnostic accuracy of all transfer models aiming at  $D_{t1}$  and  $D_{t2}$  can reach more than 70% under the most extreme condition of the 25% training samples when only the latter fully connected layers in the model are fine-tuned. These results indicate that the designed model transfer scheme is feasible for cross-condition target domain data sets. On the other hand, the diagnostic accuracy is also improved in the increasing proportion of fine-tuning modules after transfer. Significant improvement has been made in learning the lowest level convolutional pooling module with the M 1–2 strategy. However, the effect of this promotion is gradually decreasing, indicating that the low-level features of the training process of the transfer model have universal value. By contrast, the transfer of high-level features is more complex and abstract. It is challenging to obtain ideal training effects.

When all modules are fine-tuned, the diagnosis accuracy only improves 0.66% compared to that of the M 1–3 strategy under the condition of 100% training samples. To make matters worse, the diagnostic performance of the model decreased by 1.79% under the condition of 25% training samples. According to the structure of the model in the *Proposed Method*, Part B, it can be seen that

the large size of the convolution and pooling layer makes it impossible to train a large number of neurons for parameter optimization and update when the number of samples is small. Therefore, over-fitting problems occurred in model training, introducing the decrease in diagnostic performance.

However, we do not need to be pessimistic because this situation will be improved as the number of training samples increases. Therefore, it is often necessary to fix the weight of the underlying parameters rather than fine-tune all model parameters in the whole transfer procedure.

**Figures 6C,D** show the results of variations of diagnostic accuracy in cross-facility data sets ( $D_{t3}$  to  $D_{t4}$ ). It can be found that when all the training data are used for training and the number of tuning layers is more remarkable than three, the diagnostic accuracy of the target domain sample can reach an ideal result, which is more than 75%. By comparing the data of different transfer modules, it can be found that the freezing and tuning transfer strategies significantly improve the diagnosis accuracy under the cross-facility transfer task.

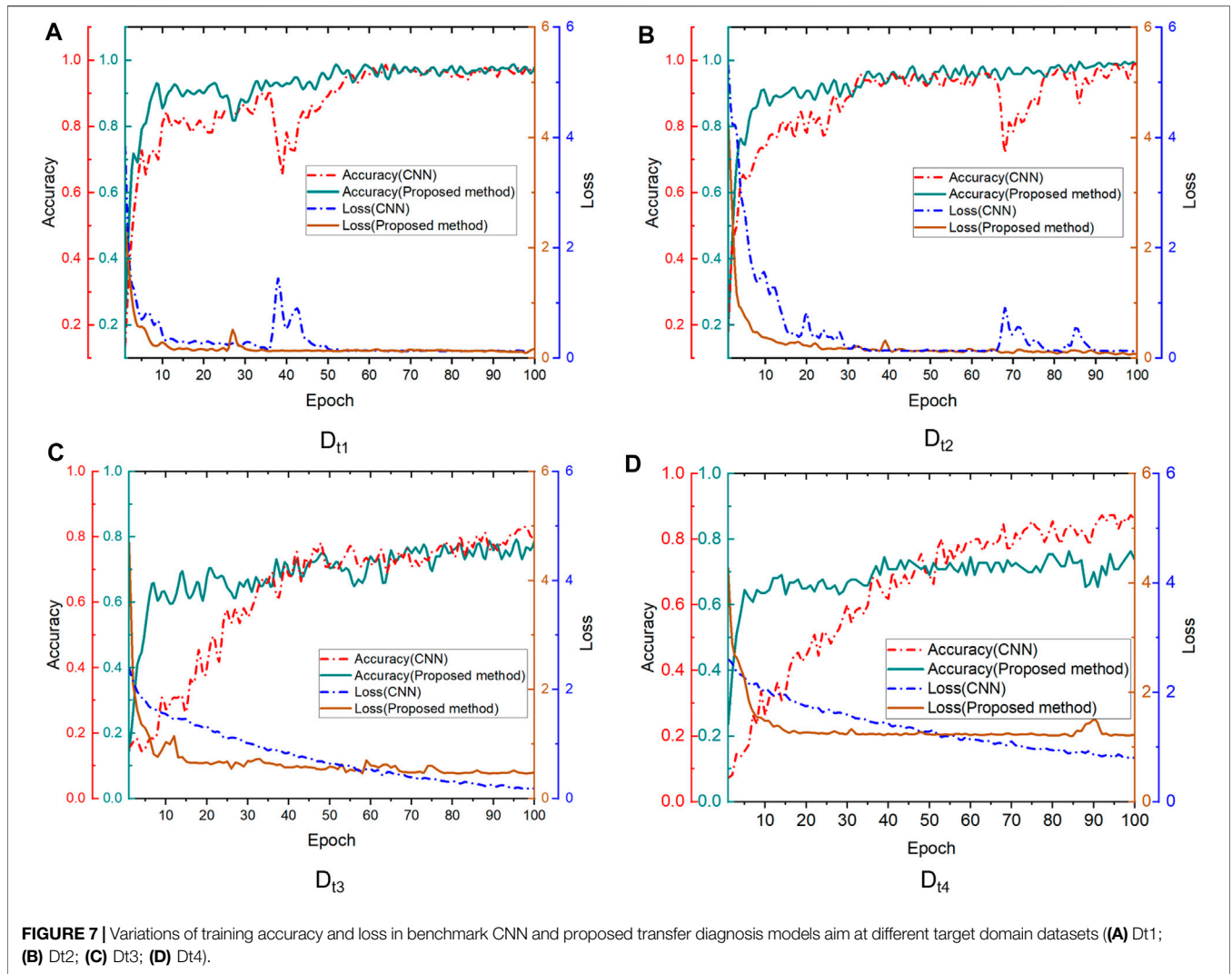
It shows that the high-dimensional features of the source domain in the transferred layer are beneficial to the generalization of the model in the target domain. In addition, the improvement of diagnostic accuracy of data samples is more significant, indicating that for cross-facility condition data with apparent differences in distribution, sufficient trainable samples are more important for advancing model performance.

Meanwhile, **Table 4** shows the influence of the amount of training data and the number of transferred layers on the training time of the model. It can be found that the increase in the number of tuning layers will significantly prolong the model training time. Compared with M 1–3, tuning all layers in the model (M 1–4) improved 97.36, 86.06, and 77.25% in three different training data levels. The above results mean that tuning a high level with high dimensional characteristics increases model complexity. However, using more training data does not significantly increase the training time of the model compared with tuning more layers. Combined with the above discussion results related to **Figure 6**, we finally selected M 1–3 as the transferred tuning model structure compared with the benchmark CNN in the following discussions.

### Compared With the Benchmark Convolutional Neural Network Model

In this section, we verify the effectiveness of the proposed method by comparing the accuracy and loss changes of different models. The benchmark CNN model and the pretraining model are identical in structure to ensure the fairness of comparison results, as is shown in **Table 1**. The former directly uses the insufficient data of the target domain for direct training. By contrast, the latter uses the M1-3 structure for different transfers to perform full pretraining on the CLEAR-I source domain data and then migrates the source domain knowledge to the target domain network.

**Figure 7** shows the variations of accuracy and loss of the benchmark CNN model and the proposed transfer model in the training process of 100 epochs for different transfer target data sets. According to the cross-condition target data set results,



**FIGURE 7 |** Variations of training accuracy and loss in benchmark CNN and proposed transfer diagnosis models aim at different target domain datasets ((A)  $D_{t1}$ ; (B)  $D_{t2}$ ; (C)  $D_{t3}$ ; (D)  $D_{t4}$ ).

which are shown in **Figures 7A,B**, both models get 98.5% accuracy after 100 epochs training. The training loss in the proposed transfer model gradually approaches a fixed value. It remains stable after 50 epochs, but the similar target losses of the benchmark CNN model for  $D_{t1}$  and  $D_{t2}$  gradually stabilized after 60 and 90 epochs, respectively.

The results indicate that the CNN model is more prone to over-fitting for sample data in target domains, thus falling into local optimal solutions. The proposed method adopts the transfer strategy based on the pretraining model, effectively reducing the network’s dependence on training parameters and sample numbers. On the other hand, it makes the network parameters establish a better initial value in the searched parameter space, conducive to faster training and convergence of the model.

**Figures 7C,D** show the training results of the transfer model for cross-facility data sets. It can be found that compared with the benchmark CNN model, the training loss of the transfer model decreases faster in the initial stage, indicating that source domain knowledge plays a specific role in the transfer process. However, with the epoch increases in the later period, training loss was not

**TABLE 5 |** Cross-facility target domain data set.

| Tuning module | Training time cost/epoch (s) |                   |                    |
|---------------|------------------------------|-------------------|--------------------|
|               | 25% training data            | 50% training data | 100% training data |
| M 1           | 0.22                         | 0.44              | 0.53               |
| M 1–2         | 0.69                         | 0.83              | 1.21               |
| M 1–3         | 1.14                         | 1.65              | 2.11               |
| M 1–4         | 2.25                         | 3.07              | 3.74               |

further reduced, mainly because of the significant difference in sample distribution among different devices. The prior knowledge obtained from the source domain data could not be further generalized to target domain data set to improve diagnostic accuracy. As the CNN model is retrained, its accuracy could be further improved with the epoch increase, which is better than the transfer model after 100 epochs. However, the ideal training effect cannot be achieved due to insufficient samples in the cross-facility target domain.

**TABLE 6** | Cross-facility target domain data set.

| Data set        | Total training time (s) |                 |                    | Diagnosis accuracy (%) |                 |                    |
|-----------------|-------------------------|-----------------|--------------------|------------------------|-----------------|--------------------|
|                 | CNN                     | Proposed method | Proposed method-50 | CNN                    | Proposed method | Proposed method-50 |
| D <sub>t1</sub> | 314.6                   | 147.2           | 80.3               | 93.9                   | 94.1            | 88.6               |
| D <sub>t2</sub> | 319.7                   | 145.6           | 74.6               | 89.5                   | 93.2            | 90.8               |
| D <sub>t3</sub> | 201.7                   | 109.7           | 59.0               | 71.2                   | 63.7            | 59.2               |
| D <sub>t4</sub> | 203.6                   | 107.5           | 58.2               | 73.5                   | 57.3            | 52.2               |

CNN, convolutional neural network.

Since the proposed strategy can achieve good generalization performance at the initial training stage, we discuss the proposed method with the benchmark CNN's training time and diagnostic accuracy after only 50 epochs. At the same time, we also give the results after 100 epochs, as is shown in **Table 5**.

It can be found that compared with the benchmark CNN model, after training with 100-epoch, the proposed transfer strategy saves 53.21% on average in the overall training time for the target domain data sets. Similar training accuracy is achieved on the cross-condition data sets (D<sub>t1</sub> and D<sub>t2</sub>). What is more noteworthy is that after 50 training sessions, the test and diagnosis accuracy of the proposed method for D<sub>t1</sub> and D<sub>t2</sub> is close to 90%, which achieves a high model training effect in less time, proving the effectiveness of the proposed transfer strategy. However, the transfer strategy has not achieved satisfactory results in cross-facility data sets (D<sub>t3</sub> and D<sub>t4</sub>). Although the accuracy was significantly improved initially, it could not be further improved due to the difference in data distribution, which proved that the existing high-dimensional feature extractor did not realize its function in the target domain samples.

## CONCLUSION

In this article, we proposed an FDD method based on the optimized transferable CNN model. The priority knowledge and proposed fine-tuning strategy improved the diagnosis performance of the pretrained model aiming at a new target domain data set. It saved 53.21% of the training time compared with the benchmark CNN model. In addition, acceptable training accuracy could be achieved no more than 50-epoch training, proving that the proposed method has good generalization performance and timeliness.

On the other hand, although the proposed transfer strategy could not achieve ideal diagnostic accuracy for the cross-facility diagnosis task, the model performance increased at the initial stage of training. It indicates that the training model with characteristic information in the source domain data set provided a specific help. However, the data distribution

difference is too big between the two data sets. The available training data are limited, leading to worse diagnosis results. Similarly, the CNN model could not obtain ideal training results when the data were missing.

In general, the proposed method ideally solved the problem of cross-condition transfer. Besides, collecting fault data of different domains at the initial stage is time-consuming and essential, which is also to prepare for future TL-related tasks. We will optimize the diagnostic performance of the deep TL model to resolve data distribution differences. In addition, the transferred non-deep learning method can be equally valuable, which will be further discussed and attempted.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

CRedit authorship contribution statement: YY: methodology, analysis, software, writing—original draft. DG: conceptualization, writing—review and editing, funding acquisition. JY: methodology, supervision, project administration. MX: methodology, supervision, funding acquisition.

## FUNDING

This work is supported by the Anhui Foreign Science and Technology Cooperation Project- Intelligent Fault Diagnosis in Nuclear Power Plants (No. 201904b11020046) and China's National Key R&D Program (No.2018YFB1900301) and the National Natural Science Foundation of China (No.71901203, 71971181). This work is also funded by Research Grant Council of Hong Kong (11203519 and 11200621), Hong Kong ITC (InnoHK Project CIMDA) and HKIDS (Project 9360163).

## REFERENCES

Correa-Jullian, C., Cardemil, J. M., López Droguett, E., and Behzad, M. (2020). Assessment of Deep Learning Techniques for Prognosis of Solar

thermal Systems. *Renew. Energ.* 145, 2178–2191. doi:10.1016/j.renene.2019.07.100

Guo, Y., Tan, Z., Chen, H., Li, G., Wang, J., Huang, R., et al. (2018). Deep Learning-Based Fault Diagnosis of Variable Refrigerant Flow Air-Conditioning System for Building Energy Saving. *Appl. Energ.* 225, 732–745. doi:10.1016/j.apenergy.2018.05.075

- Kingma, D. P., and Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. arXiv preprint arXiv:1412.6980.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *nature* 521 (7553), 436–444. doi:10.1038/nature14539
- Lee, J., Kao, H.-A., and Yang, S. (2014). Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. *Proced. Cirp* 16, 3–8. doi:10.1016/j.procir.2014.02.001
- Li, W., Wu, X., Zhang, D., Su, G., Tian, W., and Qiu, S. (2014). Preliminary Study of Coupling CFD Code FLUENT and System Code RELAP5. *Ann. Nucl. Energ.* 73, 96–107. doi:10.1016/j.anucene.2014.06.042
- Ma, J., and Jiang, J. (2011). Applications of Fault Detection and Diagnosis Methods in Nuclear Power Plants: A Review. *Prog. Nucl. Energ.* 53 (3), 255–266. doi:10.1016/j.pnucene.2010.12.001
- Mandal, S., Santhi, B., Sridhar, S., Vinolia, K., and Swaminathan, P. (2017). Nuclear Power Plant Thermocouple Sensor-Fault Detection and Classification Using Deep Learning and Generalized Likelihood Ratio Test. *IEEE Trans. Nucl. Sci.* 64 (6), 1526–1534. doi:10.1109/tns.2017.2697919
- Matteo, F., Carlo, G., Federico, P., and Enrico, Z. (2021). Time-dependent Reliability Analysis of the Reactor Building of a Nuclear Power Plant for Accounting of its Aging and Degradation. *Reliability Eng. Syst. Saf.* 205, 107173. doi:10.1016/j.res.2020.107173
- Peng, B.-S., Xia, H., Liu, Y.-K., Yang, B., Guo, D., and Zhu, S.-M. (2018). Research on Intelligent Fault Diagnosis Method for Nuclear Power Plant Based on Correlation Analysis and Deep Belief Network. *Prog. Nucl. Energ.* 108, 419–427. doi:10.1016/j.pnucene.2018.06.003
- Peng, M., and Wang, Q. (2018). False Alarm Reducing in PCA Method for Sensor Fault Detection in a Nuclear Power Plant. *Ann. Nucl. Energ.* 118, 131–139.
- Perrault, D. (2019). Nuclear Safety Aspects on the Road towards Fusion Energy. *Fusion Eng. Des.* 146, 130–134. doi:10.1016/j.fusengdes.2018.11.053
- Ruder, S., Peters, M. E., Swayamdipta, S., and Wolf, T. (2019). “Transfer Learning in Natural Language Processing,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials, 15–18. June.
- Saeed, H. A., Peng, M.-j., Wang, H., and Zhang, B.-w. (2020). Novel Fault Diagnosis Scheme Utilizing Deep Learning Networks. *Prog. Nucl. Energ.* 118, 103066. doi:10.1016/j.pnucene.2019.103066
- Seabra, J. C., Costa, M. A., and Lucena, M. M. (2016). “September). IoT Based Intelligent System for Fault Detection and Diagnosis in Domestic Appliances,” in 2016 IEEE 6th International Conference on Consumer Electronics-Berlin (ICCE-Berlin) (IEEE), 205–208.
- Tang, Q., Chai, Y., Qu, J., and Ren, H. (2018). Fisher Discriminative Sparse Representation Based on DBN for Fault Diagnosis of Complex System. *Appl. Sci.* 8 (5), 795. doi:10.3390/app8050795
- Wahlström, B. (2018). Systemic Thinking in Support of Safety Management in Nuclear Power Plants. *Saf. Sci.* 109, 201–218. doi:10.1016/j.ssci.2018.06.001
- Wang, C., Vo, H. T., and Ni, P. (2015). “December)An IoT Application for Fault Diagnosis and Prediction,” in 2015 IEEE International Conference on Data Science and Data Intensive Systems (IEEE), 726–731.
- Wang, J., and Zhang, C. (2018). Software Reliability Prediction Using a Deep Learning Model Based on the RNN Encoder-Decoder. *Reliability Eng. Syst. Saf.* 170, 73–82. doi:10.1016/j.res.2017.10.019
- Wen, L., Li, X., Gao, L., and Zhang, Y. (2017). A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Elect.* 65 (7), 5990–5998.
- Wu, Y. (2016). CLEAR-S: an Integrated Non-nuclear Test Facility for China lead-based Research Reactor. *Int. J. Energ. Res.* 40 (14), 1951–1956. doi:10.1002/er.3569
- Wu, Y. (2016). Design and R&D Progress of China Lead-Based Reactor for ADS Research Facility. *Engineering* 2 (1), 124–131. doi:10.1016/j.eng.2016.01.023
- Xin, M., Jiao, W., and Da-zi, L. (2019). “Fault Diagnosis of Nuclear Power Plant Based on Simplified Signed Directed Graph with Principal Component Analysis and Support Vector Machine,” in *Proc. 2019 Chin* (Hangzhou, China: Autom. Congr.), 3082–3087. doi:10.1109/CAC48633.2019.8997001
- Xu, B., Li, H., Pang, W., Chen, D., Tian, Y., Lei, X., et al. (2019). Bayesian Network Approach to Fault Diagnosis of a Hydroelectric Generation System. *Energy Sci Eng* 7, 1669–1677. doi:10.1002/ese3.383
- Xu, Z., Li, C., and Yang, Y. (2020). Fault Diagnosis of Rolling Bearing of Wind Turbines Based on the Variational Mode Decomposition and Deep Convolutional Neural Networks. *Appl. Soft Comput.* 95, 106515. doi:10.1016/j.asoc.2020.106515
- Yangping, Z., Bingquan, Z., and DongXin, W. (2000). Application of Genetic Algorithms to Fault Diagnosis in Nuclear Power Plants. *Reliability Eng. Syst. Saf.* 67 (2), 153–160. doi:10.1016/s0951-8320(99)00061-7
- Yao, Y., Wang, J., Ge, D., and Xie, M. (2021). “October)Intelligent PHM Based Auxiliary Decision Framework of Advanced Modular Nuclear,” in *2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing)* (IEEE), 1–5.
- Yao, Y., Wang, J., Long, P., Xie, M., and Wang, J. (2020). Small-batch-size Convolutional Neural Network Based Fault Diagnosis System for Nuclear Energy Production Safety with Big-data Environment. *Int. J. Energ. Res* 44 (7), 5841–5855. doi:10.1002/er.5348
- Yao, Y., Wang, J., and Xie, M. (2022). Adaptive Residual CNN-Based Fault Detection and Diagnosis System of Small Modular Reactors. *Appl. Soft Comput.* 114, 108064. doi:10.1016/j.asoc.2021.108064
- Yao, Y., Wang, J., Xie, M., Hu, L., and Wang, J. (2020). A New Approach for Fault Diagnosis with Full-Scope Simulator Based on State Information Imaging in Nuclear Power Plant. *Ann. Nucl. Energ.* 141, 107274. doi:10.1016/j.anucene.2019.107274
- Yoo, K. H., Back, J. H., Na, M. G., Hur, S., and Kim, H. (2018). Smart Support System for Diagnosing Severe Accidents in Nuclear Power Plants. *Nucl. Eng. Tech.* 50 (4), 562–569. doi:10.1016/j.net.2018.03.007
- Zhao, X., Kim, J., Warns, K., Wang, X., Ramuhalli, P., Cetiner, S., and Golay, M. (2021). Prognostics and Health Management in Nuclear Power Plants: An Updated Method-Centric Review with Special Focus on Data-Driven Methods. *Front. Energ. Res.* 9, 294. doi:10.3389/fenrg.2021.696785
- Zheng, J., Dong, Z., Pan, H., Ni, Q., Liu, T., and Zhang, J. (2019). Composite Multi-Scale Weighted Permutation Entropy and Extreme Learning Machine Based Intelligent Fault Diagnosis for Rolling Bearing. *Measurement* 143, 69–80. doi:10.1016/j.measurement.2019.05.002
- Zhong, M., Xue, T., and Ding, S. X. (2018). A Survey on Model-Based Fault Diagnosis for Linear Discrete Time-Varying Systems. *Neurocomputing* 306, 51–60. doi:10.1016/j.neucom.2018.04.037
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A Comprehensive Survey on Transfer Learning. *Proc. IEEE* 109 (1), 43–76.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yao, Ge, Yu and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.