Nuclear Science and Techniques (2025) 36:77
https://doi.org/10.1007/s41365-025-01643-1

Simulation-to-reality transferability framework for operating-parameter forecasting in nuclear reactors using domain adaptation

Wei-Qing Lin[1] · Xi-Ren Miao[1] · Jing Chen[1] · Ming-Xin Ye[1] · Yong Xu[2] · Hao Jiang[1] · Yan-Zhen Lu[3]

Received: 3 February 2024 / Revised: 23 March 2024 / Accepted: 27 March 2024 / Published online: 20 March 2025

Abstract
Artificial intelligence has potential for forecasting reactor conditions in the nuclear industry. Owing to economic and security concerns, a common method is to train data generated by simulators. However, achieving a satisfactory performance in practical applications is difficult because simulators imperfectly emulate reality. To bridge this gap, we propose a novel framework called simulation-to-reality domain adaptation (SRDA) for forecasting the operating parameters of nuclear reactors. The SRDA model employs a transformer-based feature extractor to capture dynamic characteristics and temporal dependencies. A parameter predictor with an improved logarithmic loss function is specifically designed to adapt to varying reactor powers. To fuse prior reactor knowledge from simulations with reality, the domain discriminator utilizes an adversarial strategy to ensure the learning of deep domain-invariant features, and the multiple kernel maximum mean discrepancy minimizes their discrepancies. Experiments on neutron fluxes and temperatures from a pressurized water reactor illustrate that the SRDA model surpasses various advanced methods in terms of predictive performance. This study is the first to use domain adaptation for real-world reactor prediction and presents a feasible solution for enhancing the transferability and generalizability of simulated data.

Keywords Nuclear power plant (NPP) · Pressurized water reactor (PWR) · Domain adaptation · Knowledge transfer · Transformer · Forecasting

1 Introduction

The advancement of nuclear energy and its safe use can provide an impetus for human progress and support the production of high-end equipment, energy security, a reduction of climate change, and a shift to greener energy sources [1, 2]. Increasing the autonomous operation of nuclear power plants (NPPs) through digitalization and intellectualization is crucial for enhancing the efficiency and safety of nuclear energy, as well as for lowering operation and maintenance costs [3, 4]. The phrase "unmanned surveillance, few people on duty" describes the automation activities of NPPs in the future [5].

One of the pivotal technologies required to fulfill this desire is the prediction of key parameters in NPPs, especially under transient conditions, to perform timely decision-making and ensure early warning [6, 7]. The precise forecasting of key parameters in reactors is a well-established challenge. In recent years, considerable efforts have been made to predict various operating parameters in reactor transient conditions, including in-core power [8], outlet temperature [9], coolant leakage [10], and pressure [11]. Typical artificial intelligence (AI) models with multidimensional mappability, support vector regression (SVR), artificial neural networks

This work was supported by the Industry-University Cooperation Project in Fujian Province University (No. 2023H6006) and the State Key Laboratory of Reliability and Intelligence of Electrical Equipment (No. EERI-KF20200005).

✉ Xi-Ren Miao
mxr@fzu.edu.cn

[1] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

[2] Fujian Fuqing Nuclear Power Company Limited, Fuqing 350300, China

[3] Fuzhou Power Supply Company, State Grid Fujian Electrical Power Company Limited, Fuzhou 350009, China

(ANNs), and long short-term memory (LSTM) are frequently utilized in operating-parameter forecasting [12–14]. For example, Zeng et al. [15] combined SVR with particle filtering to predict the core power and coolant temperature of a reactor and achieved satisfactory accuracy for reactor reactivity insertion events. Lu et al. [16] developed an ANN-based model for forecasting the thermal–hydraulic parameters in a KLT-40 S nuclear reactor under steady-state operation, and their results were in good agreement with the RELAP5 simulation.

For economic and security purposes, frequently conducting trials in real-world NPPs may trigger uncontrollable events and equipment damage. Therefore, obtaining a large number of transient samples is extremely challenging. A viable alternative for addressing the scarcity of real data is to apply numerical simulations to address class imbalance problems [17]. For example, Xiang et al. [18] proposed a gear-oriented fault detection method to enlarge fault samples by integrating a finite element method simulation and generative adversarial network, achieving satisfactory results. This idea is suitable for a mechanical system with a corresponding well-constructed simulation. Additionally, studies have utilized various programs that can perform system-level simulations (such as RELAP5 [19], TRACE [20], PCTRAN [21], CASMO5 [22], and PANGU [23]) to produce simulated data for AI model training and verification based on nuclear engineering experience and physical knowledge. Li et al. [24] conducted a study using data from the Qinshan 300 MWe NPP full-scope simulator. They combined an automated feed-forward neural network with optimization algorithms, which could effectively forecast the steam mass flow rate and water temperature during transient reactor operation for up to 5 seconds in advance. Tan et al. [25] established a mathematical model to prove the equivalence of simulated and operation data when the mean noise distribution is zero. This indicates that the simulated data can provide a supplemental dataset for the AI model in the initial training and theoretical analysis. Although various simulators have been designed to closely mimic the operations of actual reactors, the simulated and real data still exhibit certain domain discrepancies in terms of noise, numerical distributions, and dynamic characteristics for the following reasons:

(1) The mathematical models in simulators are simplified from real complex nuclear power systems and cannot fully capture the nuanced physical processes.
(2) The operating parameters and states of simulators gradually differ from actual reactors, especially the changes in burnup caused by reactor-lifespan variation.
(3) In certain transient or extreme conditions, the dynamic response of simulators may marginally differ from the operations of actual reactors.

Even if real data accurately capture the intricate features of environmental interactions in NPPs, the difficulty of data collection leads to an inability to encompass all possible scenarios. Data derived from theoretical models and computational simulations of nuclear systems are readily available and inherently safer to obtain. Both the simulated and real data have unique state characteristics. Consequently, AI models trained solely on simulated or real data may be inaccurate and difficult to apply in NPPs with high safety and reliability standards [26]. Thus, investigating the transfer of prior knowledge from sufficiently simulated data to scarce actual data is essential for enhancing the precision of operating-parameter forecasts in NPPs.

However, this scheme raises an open issue: how well do simulated data generalize to real data? Transfer learning, a deep learning technique that aims to leverage pre-existing knowledge to improve the performance in a new task or domain, has become a feasible solution [27]. Lin et al. [28] proposed a transfer learning model using maximum mean discrepancy (MMD) and a convolutional neural network (CNN). The experimental results demonstrated that transferring prior diagnostic knowledge is conducive to expanding the scope of nuclear accident diagnosis in NPPs. Domain adaptation, which is a subset of transfer learning, specifically focuses on addressing domain discrepancies between the source and target domains in scenarios of the same task [29]. For insufficiently learnable samples, numerous domain-adaptation methods have been developed to address situations in mechanical fault diagnosis [30], medical image analysis [31], and robot control [32]. Xiang et al. [33] established a fault diagnosis method using simulations to obtain sufficient faults and domain adaptation to transfer the simulated knowledge to a real-world diagnosis. This approach not only supplements scarce fault samples but also mitigates the gap between simulation and reality. Inspired by their work, domain adaptation theoretically has the potential to transfer knowledge from simulation to reality and learn the common feature subspace for parameter prediction, wherein simulated and real data are considered as the source and target domains, respectively. Thus, the effectiveness of domain-adaptation techniques in bridging the gap between simulation and reality should be investigated, particularly for accurately forecasting critical parameters in nuclear reactors.

Based on the aforementioned discussion, this study aims to devise a transferability architecture for forecasting operating parameters in nuclear reactors using a simulation-to-reality domain adaptation (SRDA) model. Specifically, the SRDA model comprises four components: a feature extractor, parameter predictor, domain discriminator, and multiple kernel maximum mean discrepancy (MK-MMD). The feature extractor, as the backbone network within the SRDA model, is established using transformers that can capture dynamic characteristics and temporal dependencies from

simulated and real data. A parameter predictor containing an improved logarithmic loss function can perform precise forecasting tasks under distinct reactor power levels. The domain discriminator utilizes an adversarial strategy that forces the feature extractor to learn deep domain-invariant features. MK-MMD quantifies the discrepancies between simulated and real data through sophisticated high-dimensional mapping. The key contributions of this study are summarized as follows:

(1) Unlike conventional methods solely using simulated data modeling, the simulation-to-reality transferability model is pioneered by a novel technique in the domain adaptation of computer vision to precisely forecast critical parameters in nuclear reactors.
(2) The transformer uses a multi-head attention mechanism and is embedded as a feature extractor in the SRDA framework to capture both dynamic characteristics and temporal dependencies from simulated and real data. The improved logarithmic loss function in the predictor is refined to adapt varied power levels in reactors.
(3) The SRDA framework is expertly developed to bridge the gap between the simulated and real data and harness the strengths of adversarial strategy (i.e., the extraction of deep domain-invariant features) with the MK-MMD (i.e., the minimization of domain distribution discrepancies) simultaneously.

The remainder of this article is organized as follows. Section 2 analyzes the differences between the simulated and real data. Section 3 introduces the proposed SRDA framework in detail. In Sect. 4, we validate the precision and superiority of the proposed method through comparative experiments. Finally, Sect. 5 concludes the paper.

2 Preliminary analysis

To visualize the differences between the simulated and real data, neutron fluxes at the same height in each of the four reactor channels are shown in Fig. 1. The four simulated curves of the height coincidence exhibited smooth tendencies over a narrower range, whereas the actual curves exhibited obvious noise over a wider range. Discrepancies arise primarily from model simplifications and potential errors in the parameter estimation. For example, in simulating reactor

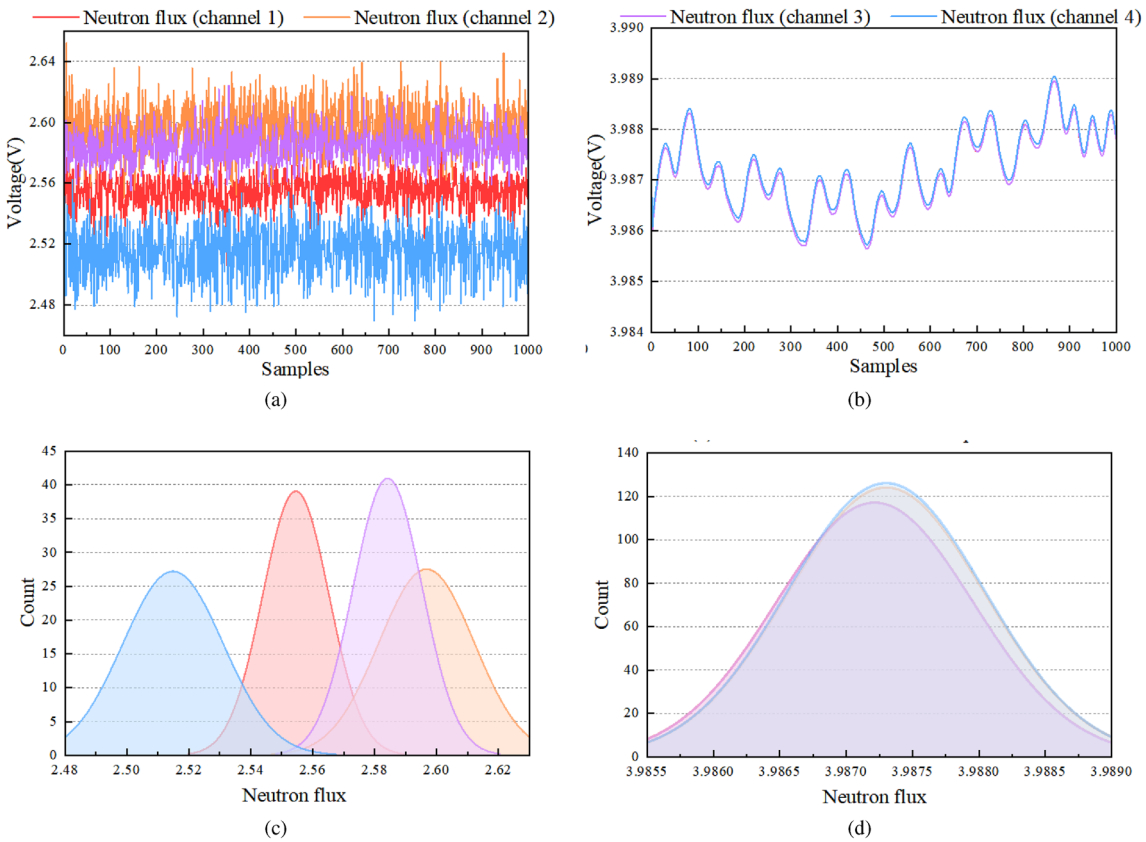

Fig. 1 (Color online) Numerical distributions of the simulated and real data. **a** Neutron fluxes from real-world reactor; **b** neutron fluxes from a full-scope simulator; **c** numerical distributions of real neutron fluxes; and **d** numerical distributions of simulated neutron fluxes

dynamics, certain assumptions must be made for computational feasibility, which can lead to deviations from the actual reactor responses. The dynamic characteristics of an NPP, such as its thermal–hydraulic behavior and neutron kinetics, were approximated in the simulations. However, these approximations can oversimplify real phenomena. In real-world reactors, neutron-flux signals are collected by ex-core detectors, whose signals are primarily induced by neutrons and gamma rays, along with a component of electrical noise [34]. In addition, burnup changes caused by reactor-lifespan variations cause the numerical distributions of the simulation and reality to shift progressively. These observations underscore the limitations of simulations and the complexities of real-world nuclear reactors, which further demonstrate the discrepancies in noise, numerical distributions, and dynamic characteristics between simulations and reality.

3 Methodology

3.1 Problem formulation

The operating-parameter data $\boldsymbol{X} = \{x_1, x_2, \cdots, x_T\} \in \mathbb{R}^T$ are given hypothetically, where T denotes the length of the time series. In the prediction task in nuclear reactors, a sliding window is applied to construct the dataset $\mathcal{D} = \{(\boldsymbol{X}_t, \boldsymbol{Y}_t)\} \in \mathbb{R}^{t \times n}$, where t and n represent the t-th temporal window and total number of windows, respectively. More specifically, $\boldsymbol{X}_t = \{x_{t-\omega-1}, \cdots, x_{t-1}, x_t\}$ represents the input window of length ω, and $\boldsymbol{Y}_t = \{x_{t+1}, x_{t+2}, \cdots, x_{t+\tau}\}$ represents the forecasting window in the future τ steps ($\tau \geq 1$). The process can be viewed as a function F_{P} as follows:

$$\{x_{t+1}, x_{t+2}, \cdots, x_{t+\tau}\} = F_{\mathrm{P}}(\{x_{t-\omega-1}, \cdots, x_{t-1}, x_t\}). \quad (1)$$

The prediction task involves transferring knowledge from ample simulated data (source domain) to scarce actual data (target domain) in NPPs. The simulated data represent the source domain $\mathcal{D}^{\mathrm{S}} = \{(\boldsymbol{X}_t^{\mathrm{S}}, \boldsymbol{Y}_t^{\mathrm{S}})\}$ $(\boldsymbol{X}_t^{\mathrm{S}} \in \mathcal{X}^{\mathrm{S}}, \boldsymbol{Y}_t^{\mathrm{S}} \in \mathcal{Y}^{\mathrm{S}})$, and the real data represent the target domain $\mathcal{D}^{\mathrm{T}} = \{(\boldsymbol{X}_t^{\mathrm{T}}, \boldsymbol{Y}_t^{\mathrm{T}})\}$ $(\boldsymbol{X}_t^{\mathrm{T}} \in \mathcal{X}^{\mathrm{T}}, \boldsymbol{Y}_t^{\mathrm{T}} \in \mathcal{Y}^{\mathrm{T}})$. They share a common feature subspace, that is, $\mathcal{X}^{\mathrm{S}} = \mathcal{X}^{\mathrm{T}}$ and $\mathcal{Y}^{\mathrm{S}} = \mathcal{Y}^{\mathrm{T}}$. Owing to the different actuation modes and system dynamics, the source and target domains have different marginal probability distributions, that is, $P(\boldsymbol{X}^{\mathrm{S}}) \neq P(\boldsymbol{X}^{\mathrm{T}})$. Thus, we design a transformation function F_{DA} to fulfill $P(F_{\mathrm{DA}}(\boldsymbol{X}^{S})) = P(F_{\mathrm{DA}}(\boldsymbol{X}^{\mathrm{T}}))$. Function F_{DA} is combined with function F_{P} to establish the transferability prediction framework $\mathcal{F}_{\mathrm{P}} = F_{\mathrm{DA}}(F_{\mathrm{P}}(\cdot))$, which performs forecasting missions in real-world nuclear reactors.

3.2 Overview of SRDA

The proposed SRDA framework aims to predict real data in nuclear reactors precisely over time by learning and transferring prior knowledge from simulations to reality. As presented in Fig. 2, the architecture of SRDA resembles that of conventional neural networks and possesses two output modules instead of one. The SRDA model comprises four modules. The blue module is the feature extractor G_{f} and serves as the backbone network that directly affects the transfer effect. We use a transformer as the feature extractor, which is described in detail in Sect. 3.4. The orange module, representing the parameter predictor G_{p}, is constructed using an improved logarithmic loss, a fully connected (FC) layer, and a prediction output layer, which forecasts the future variation

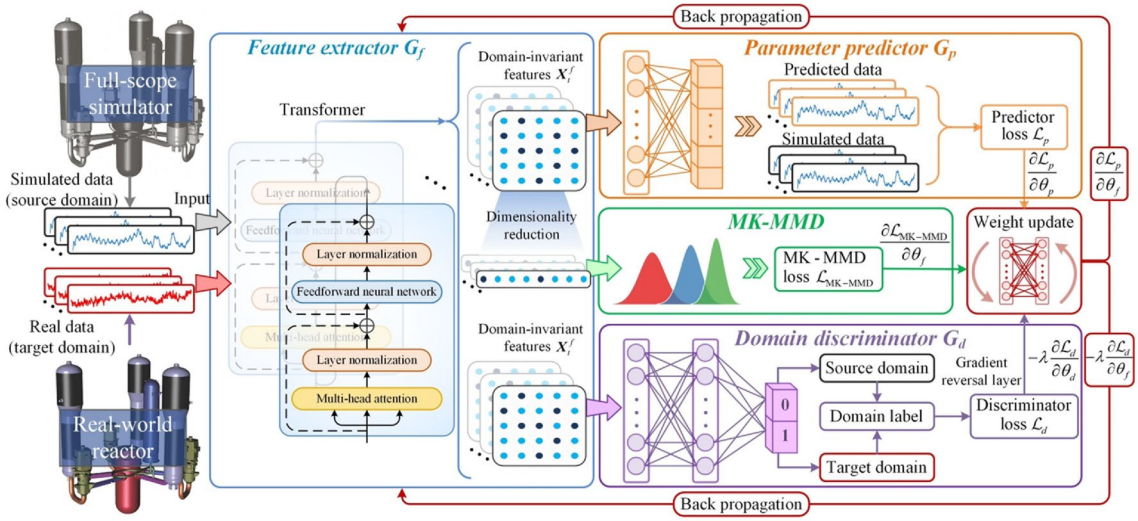

Fig. 2 (Color online) Schematic of the simulation-to-reality domain adaptation (SRDA) framework

of operating parameters in the reactors. The purple module represents the domain discriminator G_d and comprises two FC layers and a classification output layer. The purpose of G_d is to classify the training data from each domain and establish an adversarial learning strategy with feature extractor G_f. In addition, the green module represents the multiple kernel maximum mean discrepancy (MK-MMD) used to estimate the discrepancies between the simulated and real reactor data after the feature extraction.

3.3 Principles of domain adversarial strategy

In the training phase, the samples $\boldsymbol{X}_t$ from $\mathcal{D}^S$ and $\mathcal{D}^T$ are fed into the feature extractor G_f to extract temporal features. Then, the obtained features $\boldsymbol{X}_t^f$ are forwarded to both the parameter predictor G_p and domain discriminator G_d to forecast the operating parameters $\tilde{\boldsymbol{Y}}_t$ and generate the domain label $\tilde{\boldsymbol{d}}_t$. This process is expressed as follows:

$$\begin{cases} \boldsymbol{X}_t^f = G_f\left(\boldsymbol{X}_t;\boldsymbol{\theta}_f\right) \\ \tilde{\boldsymbol{Y}}_t = G_p\left(\boldsymbol{X}_t^f;\boldsymbol{\theta}_p\right) \\ \tilde{\boldsymbol{d}}_t = G_d\left(\boldsymbol{X}_t^f;\boldsymbol{\theta}_d\right), \end{cases} \tag{2}$$

where $\boldsymbol{\theta}_f$, $\boldsymbol{\theta}_p$, and $\boldsymbol{\theta}_d$ represent the trainable weight matrices in G_f, G_p, and G_d, respectively.

This framework mitigates domain differences and makes precise parameter predictions by jointly training G_f, G_p, and G_d. More specifically, training has two goals: (1) minimizing the prediction loss for G_p and (2) maximizing the domain loss for G_d simultaneously, such that the domain discriminator cannot distinguish the domain from which the obtained features originate [35]. Feature extractor G_f and domain discriminator G_d are trained adversarially to ensure that G_f maps the simulated and real data into a common subspace and generates domain-invariant features. Consequently, the training convergence learns deep domain-invariant features in the feature extractor, which refers to temporal dependencies or generic patterns that do not significantly change between the simulated and real data. To perform adversarial training, the feature extractor G_f and domain discriminator G_d are interconnected by a gradient reversal layer to achieve optimal results. For efficient backpropagation, the trade-off loss function ($\mathcal{L}_{\text{total}}$) in the framework is built and formalized as follows:

$$\begin{aligned} \mathcal{L}_{\text{total}} &= \frac{1}{n^S}\sum_{t=1}^{n^S}\mathcal{L}_p\left(\left\{G_p\left[G_f\left(\boldsymbol{X}_t;\boldsymbol{\theta}_f\right);\boldsymbol{\theta}_p\right]\right\},\boldsymbol{Y}_t\right) \\ &\quad - \frac{\lambda}{n^S+n^T}\sum_{t=1}^{n^S+n^T}\mathcal{L}_d\left(\left\{G_d\left[G_f\left(\boldsymbol{X}_t;\boldsymbol{\theta}_f\right);\boldsymbol{\theta}_d\right]\right\},\boldsymbol{d}_t\right) \\ &\quad + \mathcal{L}_{\text{MK-MMD}} \\ &= \frac{1}{n^S}\sum_{t=1}^{n^S}\mathcal{L}_p\left(\tilde{\boldsymbol{Y}}_t,\boldsymbol{Y}_t\right) - \frac{\lambda}{n^S+n^T}\sum_{t=1}^{n^S+n^T}\mathcal{L}_d\left(\tilde{\boldsymbol{d}}_t,\boldsymbol{d}_t\right) \\ &\quad + \mathcal{L}_{\text{MK-MMD}}, \end{aligned} \tag{3}$$

where λ denotes the weight coefficient that adjusts the trade-off between the predictor, discriminator, and MK-MMD losses. n^S and n^T are the numbers of samples for training from $\mathcal{D}^S$ and $\mathcal{D}^T$, respectively. $\mathcal{L}_p$, $\mathcal{L}_d$, and $\mathcal{L}_{\text{MK-MMD}}$ denote the loss functions of the predictor, discriminator, and MK-MMD. When the reactor operates at low-power levels, the magnitudes of certain operating parameters vary considerably; thus, the model cannot accurately fit smaller values. The improved logarithmic function is specifically designed as a predictor loss function $\mathcal{L}_p$, and its formula is defined by Eq. (4). The cross-entropy function is used as the discriminator loss function $\mathcal{L}_d$. $\mathcal{L}_{\text{MK-MMD}}$ is described in detail in Sect. 3.5.

$$\mathcal{L}_p = \frac{1}{n^S}\sum_{t=1}^{n^S}\left|\log\left(\tilde{\boldsymbol{Y}}_t+\epsilon\right)-\log\left(\boldsymbol{Y}_t+\epsilon\right)\right|, \tag{4}$$

where ϵ is a small constant that ensures that the predictions inside the logarithmic function are always positive and is set to 0.01.

$\mathcal{L}_p$, $\mathcal{L}_d$, and $\mathcal{L}_{\text{MK-MMD}}$ have different scales corresponding to the losses in prediction, classification, and statistics, respectively. Thus, a trade-off learning strategy is developed for joint training, in which the weight λ is adjusted dynamically and gradually increase from an initial small weight during training. A formal definition of dynamic λ_i is expressed as follows:

$$p_i = \frac{i}{E} \tag{5}$$

$$\lambda_i = \frac{2}{1+\exp\left(-10\times p_i\right)} - 1, \tag{6}$$

where p_i represents the learning progress, which increases linearly from zero to one. i and E denote the i-th epoch being processed and the maximum number of epochs, respectively. This strategy ensures that the domain discrimination is less affected by noisy data in the initial stages of training.

In the testing phase, the trained SRDA model utilizes a feature extractor (i.e., the transformer) to capture temporal characteristics from real data, which are then fed into the predictor to forecast the operating-parameter variations in a

real-world reactor. The domain discriminator and MK-MMD modules are not involved in the testing phase because their purpose is solely to assist the feature extractor in learning domain-invariant features during training.

3.4 Principles of transformer

A transformer [36] has an excellent capacity for handling time series and is utilized as the feature extractor within the SRDA model. A standard transformer has a sequence-to-sequence structure that incorporates an encoder and a decoder. To capture the dynamic characteristics and temporal dependencies from the simulated and real data, the encoder in the transformer is utilized to map the inputs into a high-dimensional domain-invariant feature matrix.

As presented in Fig. 3, the encoder comprises positional coding, multi-head attention, layer normalization, and a feed-forward neural network. The positional information is calculated using sine and cosine functions [37]. Multi-head attention aims to capture the dynamic characteristics of special events that can enhance the sensitivity of the model to critical moments or transient scenarios in reactors. As shown in Fig. 4, multi-head attention, as the basic module in the transformer, first expands the input $\boldsymbol{X}_t$ into a new embedding $\boldsymbol{X}'_t$ by an FC layer, which is described as follows:

$$\boldsymbol{X}'_t = \boldsymbol{X}_t W^{\mathrm{I}}, \boldsymbol{X}'_t \in \mathbb{R}^{k\times d\times 3}, \tag{7}$$

where W^{I} is the weight of the input FC layer. k and d are the head numbers of the attention mechanism and feature dimension, respectively.

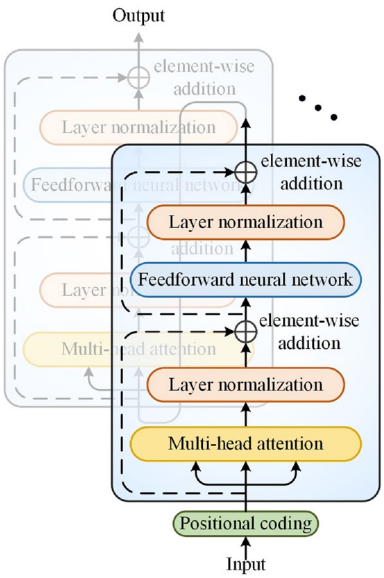

Fig. 3 Structure of transformer

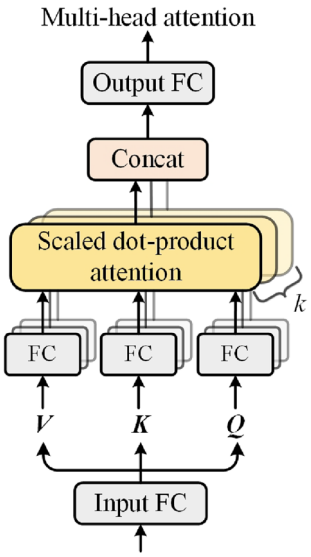

Fig. 4 Structure of multi-head attention

Embedding $\boldsymbol{X}'_t$ is further propagated to multiple heads where the weights are not shared among them. Each head has three FC layers and a scaled dot-product attention. The FC layers are employed to map $\boldsymbol{X}'_t$ into a query ($\boldsymbol{Q} \in \mathbb{R}^{k\times d}$), key ($\boldsymbol{K} \in \mathbb{R}^{k\times d}$), and value ($\boldsymbol{V} \in \mathbb{R}^{k\times d}$), which are expressed as follows:

$$\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} = \boldsymbol{X}'_t W^{\mathrm{Q}}, \boldsymbol{X}'_t W^{\mathrm{K}}, \boldsymbol{X}'_t W^{\mathrm{V}}, \tag{8}$$

where $\boldsymbol{W}^{\mathrm{Q}}$, $\boldsymbol{W}^{\mathrm{K}}$, and $\boldsymbol{W}^{\mathrm{V}}$ are the weights of the FC layers.

The scaled dot-product attention can calculate the correlation between $\boldsymbol{Q}$ and $\boldsymbol{K}$ to produce an attention map, which is employed as the weight of $\boldsymbol{V}$; the calculation of which is described formulas follows:

$$F_S(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \sigma\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V}, \tag{9}$$

where F_S is the mapping function of the scaled dot-product attention. $\sigma(\cdot)$ denotes a Softmax activation function.

The output of each head is concatenated and calculated using an output FC layer. This process can be simplified as follows:

$$S_i = F_S\left(\boldsymbol{X}'_t \boldsymbol{W}_i^{\mathrm{Q}}, \boldsymbol{X}'_t \boldsymbol{W}_i^{\mathrm{K}}, \boldsymbol{X}'_t \boldsymbol{W}_i^{\mathrm{V}}\right), i \in (0, k] \tag{10}$$

$$F_{\mathrm{A}}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \gamma\left(S_1, \ldots, S_k\right)\boldsymbol{W}^{O}, \tag{11}$$

where F_{A} denotes the mapping function of multi-head attention. $\boldsymbol{W}_i^{\mathrm{Q}}$, $\boldsymbol{W}_i^{\mathrm{K}}$, and $\boldsymbol{W}_i^{\mathrm{V}}$ represent the weights of the FC layers in the i-th head, whereas $\boldsymbol{W}^{\mathrm{O}}$ represents the weight of the final FC layer. γ denotes the concatenated operation.

The feed-forward neural network, consisting of two FC layers and a rectified linear unit activation function, primarily boosts the nonlinear fitting capability of feature extraction, which is utilized separately for each position.

3.5 Principles of MK-MMD

A standalone domain adversarial strategy may have suboptimal effects or instability in simulation-to-reality knowledge transfer. Combining the adversarial strategy with MK-MMD can compensate for these shortcomings while promoting the stability and robustness of the model. In the SRDA framework, the dimensionality of the feature matrix $\boldsymbol{X}_{\mathrm{f}}$ obtained by the feature extractor is reduced to eigenvectors $\boldsymbol{X}_t^{f^{\mathrm{S}}}$ and $\boldsymbol{X}_t^{f^{\mathrm{T}}}$, whose MK-MMD is then calculated using Eq. (12). $\mathcal{L}_{\text{MK-MMD}}$ can quantify the distribution discrepancies between simulated and real data using sophisticated high-dimensional mapping. Compared to traditional MMD, MK-MMD employs a set of kernel functions to fully analyze data across different scales and dimensions, which enhances the ability to identify distributional discrepancies [38].

$$\mathcal{L}_{\text{MK-MMD}} = \left\| \frac{1}{n^{\mathrm{S}}} \sum_{t=1}^{n^{\mathrm{S}}} \varphi\left(\boldsymbol{X}_t^{f^{\mathrm{S}}}\right) - \frac{1}{n^{\mathrm{T}}} \sum_{t=1}^{n^{\mathrm{T}}} \varphi\left(\boldsymbol{X}_t^{f^{\mathrm{T}}}\right) \right\|_{\mathcal{H}}^2. \tag{12}$$

$\varphi(\cdot)$ denotes the function that maps the feature to the reproducing kernel Hilbert space. A kernel function K, which is a convex combination of m positive semi-definite kernels K_u, is defined to avoid a complicated mapping.

$$\mathcal{K} = \left\{ K = \sum_{u=1}^{\tau} \beta_u K_u \ : \ \sum_{u=1}^{\tau} \beta_u = 1, \beta_u \geq 0, \forall u \right\}, \tag{13}$$

where K_u and β_u represent the u-th kernel function defined by the Gaussian kernel and its coefficient, respectively. u denotes the number of kernels, which is set to five.

4 Experiments

In this section, the proposed SRDA model is evaluated using two types of data (simulated and real), which are regarded as the source and target domains. Typical operating parameters are selected for forecasting, including twenty-four neutron fluxes $(N_1, \cdots, N_{24})$ and six temperatures $(T_1, \cdots, T_6)$ at different locations in the reactor. Neutron fluxes and temperatures play critical roles in reactors, as they are essential for monitoring the power distributions and levels of the reactor. As presented in Fig. 5, the neutron fluxes are gathered by ex-core neutron detectors (i.e., uncompensated ion chambers) to generate channel currents at six distinct heights in the four channels. The currents are amplified and converted into voltage signals. The temperatures recorded by the resistance thermometers correspond to the inlet and outlet temperatures in the three primary loops.

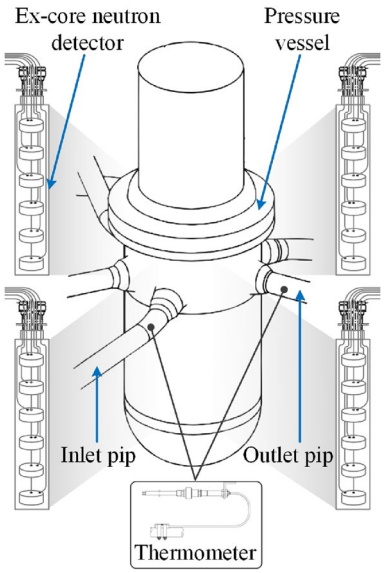

Fig. 5 Monitoring devices for neutron fluxes and temperatures

The simulated data with a 1-second sampling interval are produced by the full-scope simulator of a pressurized water reactor (PWR), which is meticulously designed to match the actual control station of an NPP, ensuring that every component and system are precisely simulated. All critical nuclear reactor systems such as the reactor core, cooling systems, control systems, and emergency response systems are integrated into the simulator to provide a comprehensive simulation environment. To extract abundant information and transfer knowledge, a full-scope simulator produces various transitory data under varying power. The actual data originate from a real-world digital instrumentation and the control system in a PWR.

Nuclear reactors mostly operate in a steady state, owing to their operating characteristics. Transient operation rarely occurs, except when it is caused by external factors such as grid peaking, shutdowns, and faults. To demonstrate the reliability of the results, the target-domain data contain two sets: 37,000 samples of shutdown data and 60,000 samples of power variation with 1-second and 10-second sampling intervals, respectively. In the two transient scenarios, the control rods are manipulated to induce perturbations in the three-dimensional power distribution, which characterizes the different degrees of change in the reactor. In the joint training phase, the training set is composed of all source data and the first 5% of the target-domain data (only steady-state operation). The remaining target-domain data are used as the test set. For the two test sets with different sampling intervals, the past 180 steps (3 and 30 min) of the historical

data are applied to recursively predict the data of the next 60 steps (1 and 10 min, respectively).

4.1 Experimental setup

Examining the effects of various feature extractors on the SRDA model can provide valuable insights. Six representative deep learning networks are applied to explore the generalizability of the proposed framework: autoencoder (AE), CNN, recurrent neural network (RNN), LSTM, gated recurrent unit (GRU), and temporal convolutional network (TCN). The key parameters of each model are listed in Table 1.

In addition to comparing different feature extractors, the SRDA model is compared with six advanced domain-adaptation methods in parallel to prove its superiority. Owing to the limited domain-adaptation methods available for forecasting tasks, we modified the existing methods proposed for time-series or visual classification. Comparison methods include deep domain confusion (DDC) [39], correlation alignment via domain adaptation (CA-DA) [40], minimum discrepancy estimation for domain adaptation (MDE-DA) [41], a DIRT-T approach to domain-adversarial adaptation (DIRT-T) [42], an adaptive domain-adversarial neural network (ADANN) [43], and adversarial spectral-kernel matching for domain adaptation (ASKM-DA) [44]. The hyperparameters of the aforementioned approaches are rationally set in accordance with corresponding studies to ensure fairness. In the training phase, the learning rate and batch size in all models are critical parameters adjusted by grid optimization. For the proposed SRDA, the parameter settings are listed in Tab. 2. All the AI models are developed using PyTorch 2.0.1 in Python version 3.8.

Three precision metrics, namely, the root mean square error (δ_{RMSE}), mean absolute error (δ_{MAE}), and symmetric mean absolute percentage error (δ_{SMAPE}), are adopted to evaluate the forecasting performance. The smaller the δ_{RMSE}, δ_{MAE}, and δ_{SMAPE} metrics, the higher the prediction accuracy. These can be calculated as follows:

$$\delta_{\mathrm{RMSE}} = \sqrt{\frac{\sum_{t=1}^{n}\left(y_t - \tilde{y}_t\right)^2}{n}}, \tag{14}$$

$$\delta_{\mathrm{MAE}} = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \tilde{y}_t\right|, \tag{15}$$

$$\delta_{\mathrm{SMAPE}} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t - \tilde{y}_t\right|}{\left[y_t + \tilde{y}_t\right]/2} \times 100, \tag{16}$$

where n is the total number of test samples. y_t and $\tilde{y}_t$ are the actual and predicted values at time t, resepctively.

Table 1 Parameter settings of various feature extractors

Model	Parameter	Value
AE	Neurons of encoder	64
	Neurons of decoder	64
CNN	Number of filters	32
	Filter size	3
RNN	Bi-direction structure	True
	Neurons of hidden layer	[32,32]
LSTM	Bi-direction structure	True
	Neurons of hidden layer	[32,32]
GRU	Bi-direction structure	True
	Neurons of hidden layer	[32,32]
TCN	Number of filters	32
	Number of residual layer	2
	Filter size	13
	Dilated factor	[1,2]

Table 2 Parameter settings of the SRDA model

Module	Parameter	Value
SRDA (transferability framework)	Input length	180
	Output length	60
	Neurons of predictor	[32, 64, 60]
	Neurons of discriminator	[32, 16, 16, 2]
	Number of kernels in MK-MMD	5
	Optimizer	Adam
	Epoch	200
	Learning rate	0.001
	Batch size	32
Transformer (feature extractor)	Number of multi-head	2
	Feature dimension	32
	Number of encoder layers	4
	Neurons of feed-forward neural network	[32, 128, 32]
	Dropout	0.1

4.2 Forecasting results

Experiments on forecasting tasks are conducted using source-only, target-only, and source-target models as baselines to validate the effectiveness of the proposed SRDA model for knowledge transfer from simulation to reality in nuclear reactors. The source-only model is trained exclusively on the source-domain training set and directly tested on the target-domain test set. Similarly, the target-only model is trained on the target-domain training set and directly tested on the target-domain test set. The source-target model is trained on the source-domain data using conventional transfer learning, and it fine-tunes the weights of its final layer on the target-domain data. Furthermore, a transformer and improved logarithmic loss are used to

construct the three baselines mentioned above. Domain adaptation is not employed in this process. Forecasting of the neutron flux N_1 and hot-leg temperature T_1 are performed as examples. As shown in Fig. 6, the target-only model trained using the first 5% of the steady-operation data exhibits the largest predictive deviation. After learning a sufficient number of simulated samples, the trained source-only model can adapt to the basic variational trend of real neutron fluxes and temperatures, resulting in suboptimal effects attributable to the difference between the simulation and reality. Although the source-target model outperforms the above two models, it exhibits a certain deviation in the local area. Compared with the three baselines, the predictive trend obtained by the SRDA model with domain adaptation is generally closer to the real curves and free from the interference of operational noise in complex nuclear systems.

Table 3 provides the specific average and standard deviation of the errors ($\overline{\delta}_{\mathrm{RMSE}}$, $\overline{\delta}_{\mathrm{MAE}}$, and $\overline{\delta}_{\mathrm{SMAPE}}$) for all the models in the two test sets. For neutron-flux forecasts, the SRDA model demonstrates superior performance during shutdown, with a remarkably low $\overline{\delta}_{\mathrm{RMSE}}$ of 0.010 V and $\overline{\delta}_{\mathrm{MAE}}$ of 0.012 V. Moreover, $\overline{\delta}_{\mathrm{SMAPE}}$ is an order of magnitude lower than that of its counterparts, at 1.248%. This precision is also reflected in the power-variation scenarios, where the SRDA model achieves an $\overline{\delta}_{\mathrm{RMSE}}$ of 0.008 V and $\overline{\delta}_{\mathrm{MAE}}$ of 0.009 V, along with a notably low $\overline{\delta}_{\mathrm{SMAPE}}$ of 0.636%. For the inlet and outlet temperature forecasts, the SRDA model achieves excellent results, with the lowest $\overline{\delta}_{\mathrm{RMSE}}$, $\overline{\delta}_{\mathrm{MAE}}$, and $\overline{\delta}_{\mathrm{SMAPE}}$ under the shutdown and power-variation conditions. In addition, the low standard deviation further highlights the stability of the SRDA model. Case experiments demonstrate the

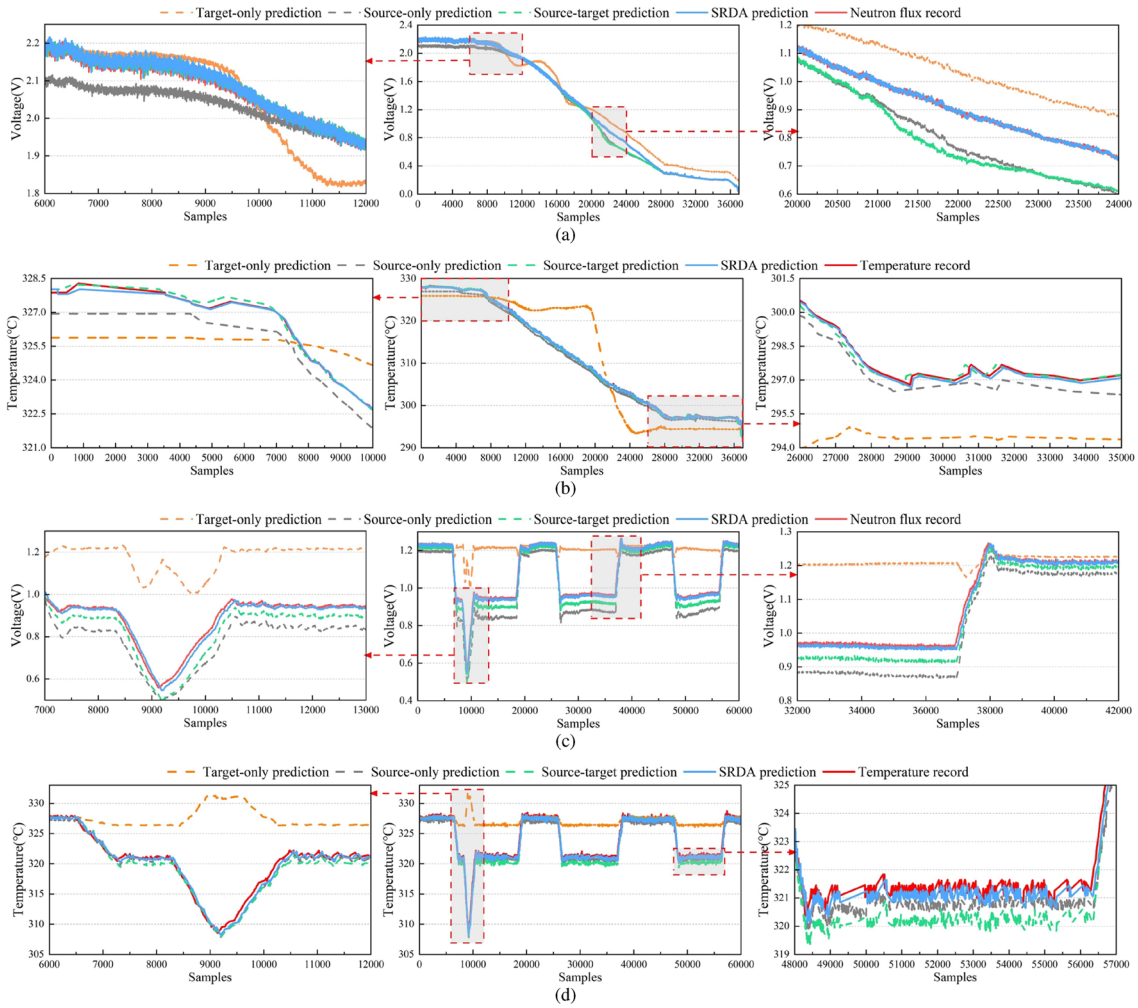

Fig. 6 (Color online) Forecast curves of the SRDA model compared with various baselines. **a** Forecast curves for neutron flux N_1 in shutdown; **b** forecast curves for outlet temperature T_1 in shutdown; **c** forecast curves for neutron flux N_1 in power variation; and **d** forecast curves for outlet temperature T_1 in power variation

Table 3 Forecast errors of the SRDA model compared with various baselines

Forecasting target	Model	Shutdown			Power variation		
		$\overline{\delta}_{\mathrm{RMSE}}$ (V/°C)	$\overline{\delta}_{\mathrm{MAE}}$ (V/°C)	$\overline{\delta}_{\mathrm{SMAPE}}$ (%)	$\overline{\delta}_{\mathrm{RMSE}}$ (V/°C)	$\overline{\delta}_{\mathrm{MAE}}$ (V/°C)	$\overline{\delta}_{\mathrm{SMAPE}}$ (%)
Neutron flux	Target-only	0.902±0.688	0.903±0.689	61.549 ± 23.977	0.107±0.051	0.105±0.049	8.710±4.543
	Source-only	0.034±0.011	0.041±0.011	6.615±1.639	0.035±0.014	0.036±0.014	3.083±1.733
	Source-target	0.028±0.005	0.030±0.005	5.587±1.277	0.026±0.035	0.027±0.035	2.192±3.081
	SRDA*	**0.010±0.002**	**0.012±0.003**	**1.248±0.287**	**0.008±0.003**	**0.009±0.003**	**0.636±0.178**
Temperature	Target-only	4.918±4.379	4.914±4.377	1.653±1.482	1.246±1.092	1.252±1.096	0.393±0.331
	Source-only	0.825±0.697	0.825±0.696	0.280±0.239	0.357±0.134	0.369±0.123	0.118±0.049
	Source-target	0.777±0.749	0.778±0.748	0.264±0.256	0.351±0.140	0.363±0.130	0.116±0.051
	SRDA*	**0.113±0.079**	**0.118±0.085**	**0.037±0.026**	**0.223±0.111**	**0.231±0.110**	**0.073±0.038**

effectiveness of domain adaptation in transferring knowledge from simulation to reality in a practical reactor.

4.3 Forecasting comparison and analysis

Temporal feature extraction within the SRDA model is interlinked with the operating-parameter prediction performance. To analyze the impact of various feature extractors, experiments are conducted to replace the transformer in the SRDA framework with the six representative neural networks specified in Sect. 4.1: AE, CNN, RNN, LSTM, GRU, and TCN. As depicted in Fig. 7, the SRDA framework consistently achieves favorable outcomes across various feature extractors, demonstrating its broad versatility. However, different backbone networks moderately affect the prediction accuracy for neutron fluxes and temperatures during the shutdown and power-variation phases. Owing to the lack of inherent architecture in AE and CNN for recording temporal dependencies, the SRDA (AE) and SRDA (CNN) models exhibit insufficient feature-extraction capabilities. The SRDA (RNN) makes unstable predictions, as reflected in its δ_{SMAPE}, owing to the absence of effective memory mechanisms. The SRDA (LSTM) and SRDA (GRU) models incorporate memory cells and gating mechanisms to mitigate issues, such as vanishing gradients, thereby bolstering the temporal feature-extraction process. SRDA (TCN), which incorporates dilated causal convolutions with a larger receptive field to capture long-term characteristics, exhibits precision comparable to that of SRDA (Trans) and is a robust contender. The multi-head attention block of the proposed SRDA (Trans) allows it to capture both subtle long- and short-term dependencies, which makes it superior to other feature extractors in adapting to complex variations. The experimental results are demonstrated by the steady $\overline{\delta}_{\mathrm{SMAPE}}$ in the forecasting under reactor-shutdown and power-variation conditions.

Table 4 presents a detailed comparison of the predictive precision across various feature extractors for the two test sets. In summary, the models intricately designed for time-series analysis, such as TCN and the transformer within the SRDA framework, possess advanced temporal feature-extraction capabilities. This facilitates more effective domain adaptation, resulting in enhanced predictive performance. Although SRDA (TCN) yields formidable and competitive results, SRDA (Trans) demonstrates unparalleled performance for both neutron fluxes and temperatures. For

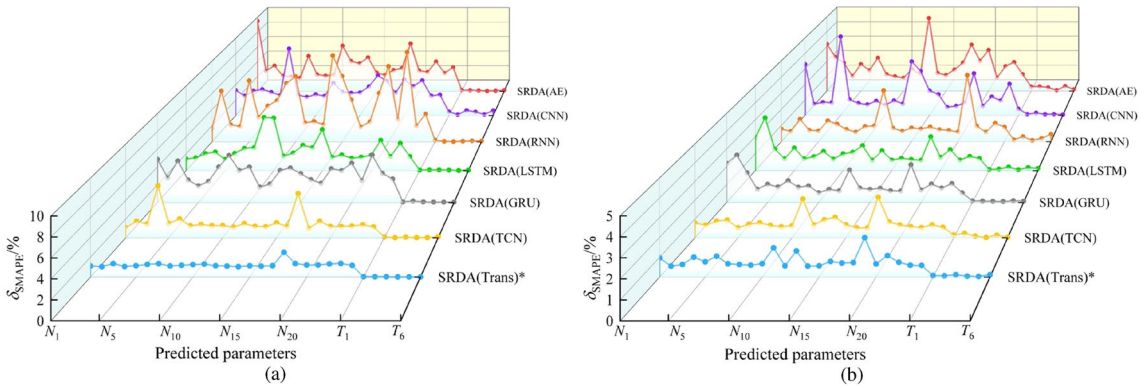

Fig. 7 (Color online) Comparison of δ_{SMAPE} by various feature extractors in the SRDA model. **a** Prediction errors in shutdown and **b** prediction errors in power variation

Table 4 Comparison of forecast errors by various feature extractors in the SRDA model

Forecasting target	Feature extractor	Shutdown			Power variation		
		$\overline{\delta}_{\mathrm{RMSE}}$ (V/°C)	$\overline{\delta}_{\mathrm{MAE}}$ (V/°C)	$\overline{\delta}_{\mathrm{SMAPE}}$ (%)	$\overline{\delta}_{\mathrm{RMSE}}$ (V/°C)	$\overline{\delta}_{\mathrm{MAE}}$ (V/°C)	$\overline{\delta}_{\mathrm{SMAPE}}$ (%)
Neutron flux	SRDA (AE)	0.023±0.013	0.038±0.044	3.496±1.791	0.022±0.009	0.028±0.011	1.817±0.942
	SRDA (CNN)	0.026±0.007	0.030±0.010	3.525±1.344	0.019±0.015	0.020±0.015	1.464±1.176
	SRDA (RNN)	0.044±0.035	0.048±0.036	4.902±3.013	0.014±0.012	0.015±0.013	1.077±0.829
	SRDA (LSTM)	0.021±0.013	0.024±0.015	2.563±1.410	0.014±0.008	0.015±0.008	1.076±0.589
	SRDA (GRU)	0.024±0.009	0.026±0.009	3.375±1.065	0.013±0.006	0.015±0.006	1.076±0.577
	SRDA (TCN)	**0.010**±0.003	0.013±**0.003**	1.365±**0.280**	0.009±**0.002**	0.010±**0.003**	0.686±0.198
	SRDA (Trans)*	**0.010**±0.002	**0.012±0.003**	**1.248**±0.287	**0.008**±0.003	**0.009±0.003**	**0.636±0.178**
Temperature	SRDA (AE)	0.376±0.148	0.497±0.239	0.123±0.047	0.707±0.365	1.050±0.583	0.226±0.115
	SRDA (CNN)	1.276±0.730	1.628±0.977	0.419±0.238	0.429±0.468	0.513±0.514	0.135±0.143
	SRDA (RNN)	0.300±0.195	0.339±0.264	0.098±0.062	1.074±0.982	1.120±0.981	0.345±0.305
	SRDA (LSTM)	0.241±0.094	0.252±0.093	0.079±0.030	0.379±0.197	0.397±0.206	0.122±0.063
	SRDA (GRU)	0.256±0.153	0.271±0.154	0.084±0.047	0.241±**0.048**	0.258±**0.048**	0.078±**0.015**
	SRDA (TCN)	0.177±0.131	0.181±0.133	0.058±0.043	0.254±0.179	0.270±0.180	0.083±0.056
	SRDA (Trans)*	**0.113±0.079**	**0.118±0.085**	**0.037±0.026**	**0.223**±0.111	**0.231**±0.110	**0.073**±0.038

example, in inlet and outlet temperature prediction, compared with the average errors ($\overline{\delta}_{\mathrm{RMSE}}$, $\overline{\delta}_{\mathrm{MAE}}$, and $\overline{\delta}_{\mathrm{SMAPE}}$) of SRDA (TCN), SRDA (Trans) is improved by 36.158%, 34.807%, and 36.207% in shutdown, as well as by 12.205%, 14.444%, and 12.048% during power variation, respectively, further showcasing the superiority of the transformer for capturing temporal dependencies in complex, dynamic nuclear reactor systems.

To demonstrate the knowledge transfer superiority of the proposed method from simulation to reality, the SRDA model is compared with the six advanced domain-adaptation methods specified in Sect. 4.1, namely, the DDC, CA-DA, MDE-DA, DIRT-T, ADANN, and ASKM-DA models. To ensure a fair evaluation, the transformer and improved logarithm losses are applied in the aforementioned methods. As presented in Fig. 8, the DDC and CA-DA models are statistic-based domain adaptations, and are designed to mitigate the distribution discrepancies between the simulated and real features. Their knowledge transfer ability for neutron fluxes and temperatures during shutdown is inadequate, leading to larger errors. MDE-DA is a composite method that improves the predictive stability on the two test sets by fusing second-order statistics in CA-DA with MMD in DDC. With the incorporation of conditional entropy and a teacher model, the adversarial-based DIRT-T approach effectively forces the feature extractor to align domain-invariant features. This methodology demonstrates moderate performance in terms of inlet and outlet temperature predictions. In contrast, ADANN and ASKM-DA, specially designed for time-series analysis tasks, exhibit more refined accuracy because of their marginally smaller errors compared to DIRT-T. Based on Fig. 8, adversarial-based methods tend to perform consistently better, thereby enhancing the generalization from simulation to reality. The finesse of the SRDA model involves adopting a domain adversarial strategy to extract deep domain-invariant features between the simulated and real data, in addition to utilizing MK-MMD to mitigate their distribution discrepancies. This dual approach guarantees that the model maintains the essential characteristics of the simulation while adjusting to the distribution and fluctuations that exist in a real-world reactor, which considerably improves its forecast precision.

4.4 Model interpretation

The proposed trade-off loss function includes the predictor $\mathcal{L}_{\mathrm{p}}$, discriminator $\mathcal{L}_{\mathrm{d}}$, and MK-MMD losses $\mathcal{L}_{\mathrm{MK\text{-}MMD}}$. Because $\mathcal{L}_{\mathrm{d}}$ and $\mathcal{L}_{\mathrm{MK\text{-}MMD}}$ have been proven to enhance performance in Sect. 4.3, we focus on investigating the contribution of $\mathcal{L}_{\mathrm{p}}$ to the prediction precision. $\mathcal{L}_{\mathrm{p}}$ acts as an improved logarithmic loss and is responsible for adapting the predictor losses under different power levels in nuclear reactors. We compared the logarithmic loss with conventional mean square error (MSE) loss. An example (only six layers of neutron flux in one channel are shown) of the results for the neutron-flux forecasting task is shown in Fig. 9. The SRDA model using the MSE loss achieves qualified predictions for the SRDA model, with only a marginal divergence from the actual values observed at lower reactor power levels. This deviation can be attributed to situations in which the MSE loss changes significantly with data scaling, making the model less sensitive at low powers, and the value is close to zero. The proposed SRDA model exhibits superior predictive accuracy for different power operations by balancing different numerical scales. Enhanced precision is critical

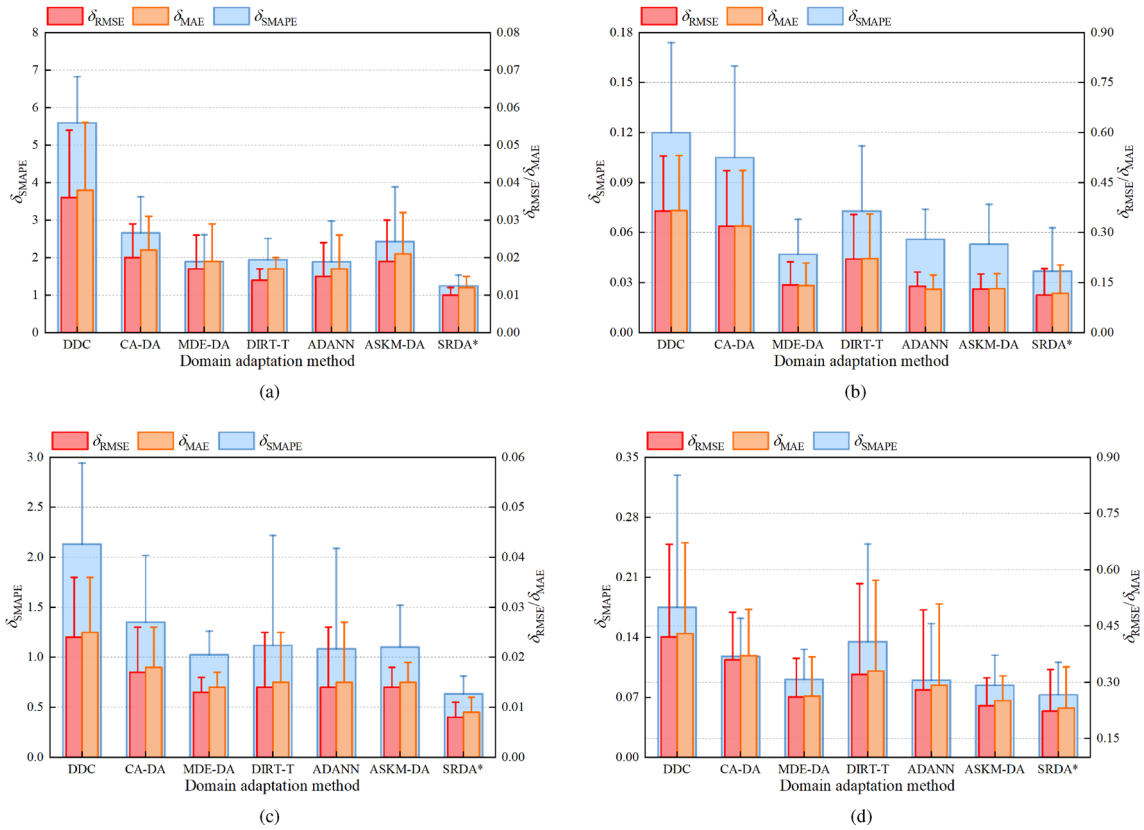

Fig. 8 (Color online) Comparison of forecast errors by various domain-adaptation methods. **a** Forecast errors for neutron fluxes in shutdown; **b** forecast errors for inlet and outlet temperatures in shutdown; **c** forecast errors for neutron fluxes in power variation; and **d** forecast errors for inlet and outlet temperatures in power variation

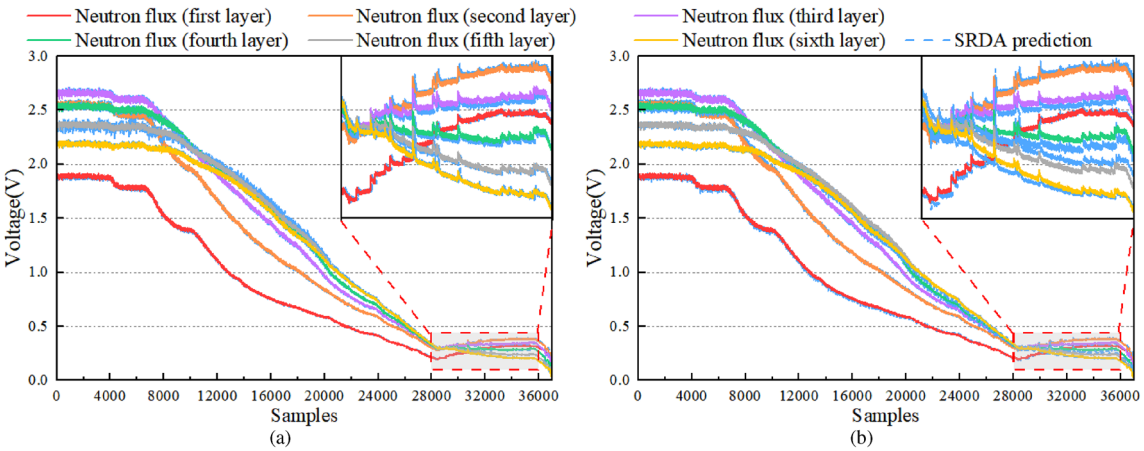

Fig. 9 (Color online) Forecast curves of different predictor loss functions in the SRDA model. **a** Forecast curves of the proposed SRDA model. **b** Forecast curves of the SRDA model using MSE loss

for fine-tuned applications in nuclear reactor operations, where minute deviations can have significant implications.

To present the transferability effect of the domain adaptation intuitively, feature distributions are visualized using t-distributed stochastic embedding (t-SNE). t-SNE is a nonlinear dimensionality reduction algorithm that can transform a high-dimensional feature matrix into a two-dimensional eigenvector for visualization. As depicted in Fig. 10, the two domains are color-coded, with red denoting the simulated data and blue denoting the real data. In

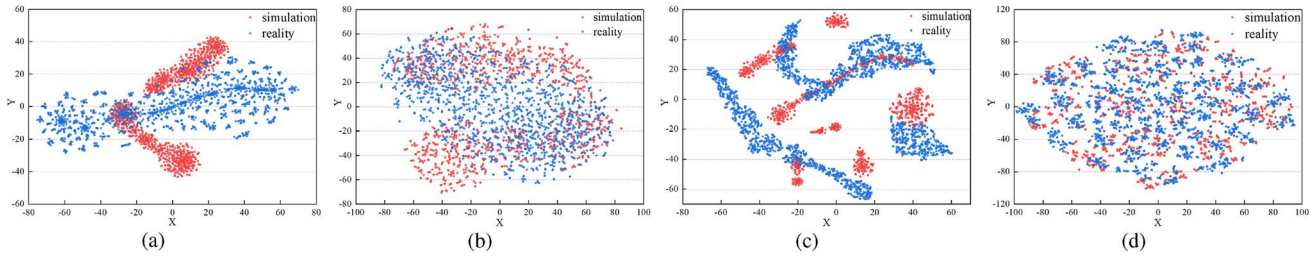

Fig. 10 (Color online) Feature visualization of simulation-to-reality knowledge transfer. **a** Without domain adaptation in shutdown; **b** domain adaptation in shutdown; **c** without domain adaptation in power variation; and **d** domain adaptation in power variation

detail, Fig. 10a and c shows the feature distributions without domain adaptation for both datasets. The features of the two domains only overlap locally, illustrating the similarities and discrepancies that commonly exist between simulation and reality. Thus, directly applying a model trained on simulated data to real data results in unsatisfactory forecasting results owing to a domain shift. In contrast, Fig. 10b and d shows the feature distribution after the feature extraction in the SRDA model. Notably, after the domain adaptation, the distributions of the extracted features from the simulation and reality are uniformly mixed, illustrating that the SRDA model can mitigate the domain discrepancies effectively to enhance the operating-parameter prediction in reactors.

5 Conclusion

Simulators imperfectly emulate reality in NPPs due to their different actuation modes and system dynamics. This study aimed to mitigate the discrepancies in noise, numerical distributions, and dynamic characteristics between simulated and real data. A novel transferability framework, called the SRDA model, was proposed for forecasting critical parameters in nuclear reactors. The SRDA framework comprised a feature extractor, parameter predictor, domain discriminator, and MK-MMD. Relative to several advanced domain-adaptation methods, the results indicated that the SRDA model demonstrates superior knowledge transfer by leveraging ample simulated and finite real data. The transformer-based feature extractor adeptly captured the dynamic characteristics and temporal dependencies in transient conditions, such as reactor shutdown and power variation, as evidenced by comparisons with various feature extractors. The improved logarithmic loss within the predictor was conducive to enhancing forecasting precision at various power levels. Furthermore, the integration of the domain adversarial strategy and MK-MMD effectively adapted to the distributions and fluctuations in real-world reactors while retaining the essential characteristics in the simulation. Considering the significant impact of different feature extractors, the versatility of the SRDA model enables the substitution of backbone networks tailored to specific scenarios in NPPs, which is another intriguing problem.

Author contributions All authors contributed to the study conception and design. Material preparation and analysis were performed by Ming-Xin Ye and Yan-Zhen Lu. The first draft of the manuscript was written by Wei-Qing Lin. The manuscript was reviewed by Xi-Ren Miao, Jing Chen, and Hao Jiang. Data are collected by Yong Xu. All authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. W. Xu, J. Li, H. Xie et al., Conceptual design and safety characteristics of a new multi-mission high flux research reactor. Nucl. Sci. Tech. **34**, 34 (2023). https://doi.org/10.1007/s41365-023-01191-6
2. S.T. Hassan, D. Khan, B. Zhu et al., Is public service transportation increase environmental contamination in china? the role of nuclear energy consumption and technological change. Energy **238**, 929–962 (2022). https://doi.org/10.1016/j.energy.2021.121890
3. A.J. Dave, T. Lee, R. Ponciroli et al., Design of a supervisory control system for autonomous operation of advanced reactors. Ann. Nucl. Energy **182**, 109593 (2023). https://doi.org/10.1016/j.anucene.2022.109593
4. J. Chen, Y. Lu, H. Jiang et al., Twin model-based fault detection and tolerance approach for in-core self-powered neutron detectors. Nucl. Sci. Tech. **34**, 117 (2023). https://doi.org/10.1007/s41365-023-01276-2
5. D. Lv, R. Chen, X. Li, *In Lecture Notes in Electrical Engineering, Preliminary study on improving the automation level of large commercial pressurized water reactor*, vol. 779 (China, Beijing, 2021), pp.429–435
6. J. Ling, G. Liu, J. Li et al., Fault prediction method for nuclear power machinery based on bayesian ppca recurrent neural network model. Nucl. Sci. Tech. **31**, 75 (2020). https://doi.org/10.1007/s41365-020-00792-9
7. H. Song, X. Liu, M. Song, Comparative study of data-driven and model-driven approaches in prediction of nuclear power plants operating parameters. Appl. Energ. **341**, 121077 (2023). https://doi.org/10.1016/j.apenergy.2023.121077

8. W. Li, P. Ding, W. Xia et al., Artificial neural network reconstructs core power distribution. Nucl. Eng. Technol. **54**, 617–626 (2022). https://doi.org/10.1016/j.net.2021.08.015
9. D.A. Ejigu, X. Liu, Gradient descent-particle swarm optimization based deep neural network predictive control of pressurized water reactor power. Prog. Nucl. Energ. **145**, 104108 (2022). https://doi.org/10.1016/j.pnucene.2021.104108
10. H.P. Nguyen, P. Baraldi, E. Zio, Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants. Appl. Energ. **283**, 116346 (2021). https://doi.org/10.1016/j.apenergy.2020.116346
11. I.M. Radaideh, C. Pigg, T. Kozlowski et al., Neural-based time series forecasting of loss of coolant accidents in nuclear power plants. Expert Syst. Appl. **160**, 113699 (2020). https://doi.org/10.1016/j.eswa.2020.113699
12. K. Moshkbar-Bakhshayesh, Comparative study of application of different supervised learning methods in forecasting future states of npps operating parameters. Ann. Nucl. Energy **132**, 87–99 (2019). https://doi.org/10.1016/j.anucene.2019.04.031
13. H. Xu, T. Tang, B. Zhang et al., Application of artificial neural network for the critical flow prediction of discharge nozzle. Nucl. Eng. Technol. **54**, 834–841 (2022). https://doi.org/10.1016/j.net.2021.08.038
14. W. Lin, X. Miao, J. Chen et al., Forecasting thermal parameters for ultra-high voltage transformers using long- and short-term time-series network with conditional mutual information. IET Electr. Power Appl. **16**, 548–564 (2022). https://doi.org/10.1049/elp2.12175
15. Y. Zeng, J. Liu, K. Sun et al., Machine learning based system performance prediction model for reactor control. Ann. Nucl. Energy **113**, 270–278 (2018). https://doi.org/10.1016/j.anucene.2017.11.014
16. Q. Lu, Y. Yuan, F. Li et al., Prediction method for thermal-hydraulic parameters of nuclear reactor system based on deep learning algorithm. Appl. Therm. Eng. **196**, 117272 (2021). https://doi.org/10.1016/j.applthermaleng.2021.117272
17. J. Xiang, Y. Zhong, A novel personalized diagnosis methodology using numerical simulation and an intelligent method to detect faults in a shaft. Appl. Sci-Basel **6**, 414 (2016). https://doi.org/10.3390/app6120414
18. Y. Gao, X. Liu, J. Xiang, Fault detection in gears using fault samples enlarged by a combination of numerical simulation and a generative adversarial network. IEEE-ASME Trans. Mechatron. **27**, 3798–3805 (2022). https://doi.org/10.1109/TMECH.2021.3132459
19. H. Song, M. Song, X. Liu, Online autonomous calibration of digital twins using machine learning with application to nuclear power plants. Appl. Energ. **326**, 1–14 (2022). https://doi.org/10.1016/j.apenergy.2022.119995
20. O. Safarzadeh, A. Pourrostam, The large-break LOCA uncertainty analysis in a VVER-1000 reactor using TRACE and DAKOTA. Nucl. Eng. Des. **412**, 112459 (2023). https://doi.org/10.1016/j.nucengdes.2023.112459
21. B. Qi, X. Xiao, J. Liang et al., An open time-series simulated dataset covering various accidents for nuclear power plants. Sci. Data **9**, 714–766 (2022). https://doi.org/10.1038/s41597-022-01879-1
22. M. Hou, W. Lv, M. Kong et al., Efficient predictor of pressurized water reactor safety parameters by topological information embedded convolutional neural network. Ann. Nucl. Energy **192**, 110004 (2023). https://doi.org/10.1016/j.anucene.2023.110004
23. D. She, B. Xia, J. Guo et al., Prediction calculations for the first criticality of the htr-pm using the pangu code. Nucl. Sci. Tech. **32**, 90 (2021). https://doi.org/10.1007/s41365-021-00936-5
24. X. Li, K. Cheng, T. Huang et al., Research on short term prediction method of thermal hydraulic transient operation parameters based on automated deep learning. Ann. Nucl. Energy **165**, 108777 (2022). https://doi.org/10.1016/j.anucene.2021.108777
25. X. Li, K. Cheng, T. Huang et al., Equivalence analysis of simulation data and operation data of nuclear power plant based on machine learning. Ann. Nucl. Energy **163**, 108507 (2021). https://doi.org/10.1016/j.anucene.2021.108507
26. S. Siddharth, Artificial intelligence in nuclear industry: Chimera or solution? J. Clean. Prod. **278**, 124022 (2021). https://doi.org/10.1016/j.jclepro.2020.124022
27. K.I.K. Wang, X. Zhou, W. Liang et al., Federated transfer learning based cross-domain prediction for smart manufacturing. IEEE Trans. Ind. Inf. **18**, 4088–4096 (2022). https://doi.org/10.1109/TII.2021.3088057
28. J. Li, M. Lin, Y. Li et al., Transfer learning network for nuclear power plant fault diagnosis with unlabeled data under varying operating conditions. Energy **254**, 124358 (2022). https://doi.org/10.1016/j.energy.2022.124358
29. Y. Shi, X. Ying, J. Yang, Deep unsupervised domain adaptation with time series sensor data: a survey. Sensors-basel. **22**, 5507 (2022). https://doi.org/10.3390/s22155507
30. Z. Wang, X. He, B. Yang et al., Subdomain adaptation transfer learning network for fault diagnosis of roller bearings. IEEE Trans. Ind. Electron. **69**, 8430–8439 (2022). https://doi.org/10.1109/TIE.2021.3108726
31. H. Guan, M. Liu, Domain adaptation for medical image analysis: a survey. IEEE Trans. Biomed. Eng. **69**, 1173–1185 (2022). https://doi.org/10.1109/TBME.2021.3117407
32. H. Ju, R. Juan, R. Gomez et al., Transferring policy of deep reinforcement learning from simulation to reality for robotics. Nat. Mach. Intell. **4**, 1077–1087 (2022). https://doi.org/10.1038/s42256-022-00573-6
33. Y. Lou, A. Kumar, J. Xiang, Machinery fault diagnosis based on domain adaptation to bridge the gap between simulation and measured signals. IEEE Trans. Instrum. Meas. **71**, 3514709 (2022). https://doi.org/10.1109/TIM.2022.3180416
34. Y.B. Kim, F.P. Vista, S.B. Cho et al., Digitalization of the ex-core neutron flux monitoring system for apr1400 nuclear power plant. Appl. Sci.-Basel **10**, 8331 (2020). https://doi.org/10.3390/app10238331
35. M. Ragab, E. Eldele, W.L. Tan et al., ADATIME: a benchmarking suite for domain adaptation on time series data. ACM Trans. Knowl. Discov. From Data **17**, 1–18 (2023). https://doi.org/10.1145/3587937
36. C. Wang, Y. Wang, Z. Ding et al., A transformer-based method of multienergy load forecasting in integrated energy system. IEEE Trans. Smart Grid **13**, 2703–2714 (2022). https://doi.org/10.1109/TSG.2022.3166600
37. A. Vaswani, N. Shazeer, N. Parmar et al., in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Attention is all you need. 2017, pp. 6000–6010
38. M. Long, Y. Cao, J. Wang et al., in *The 32nd International Conference on Machine Learning, ICML 2015*, Learning transferable features with deep adaptation networks. Vol. 1, Lile, France, 2015, pp. 97–105
39. T. Eric, H. Judy, Z. Ning et al., Deep domain confusion: maximizing for domain invariance. (2014). arXiv:1412.3474
40. B. Sun, J. Feng, K. Saenko, *Correlation alignment for unsupervised domain adaptation*, (Springer International Publishing, Cham, 2017), pp. 153–171. https://doi.org/10.1007/978-3-319-58347-1_8
41. M.M. Rahman, C. Fookes, M. Baktashmotlagh et al., *On minimum discrepancy estimation for deep domain adaptation*, (Springer International Publishing, 2020), pp. 81–94. https://doi.org/10.1007/978-3-030-30671-7_6

42. R. Shu, H.H. Bui, H. Narui et al., in *arXiv*, A dirt-t approach to unsupervised domain adaptation. 2018, pp. 1–19,https://doi.org/10.48550/arXiv.1802.08735
43. Y. Ma, Z. Zhang, H.L. Yang et al., An adaptive adversarial domain adaptation approach for corn yield prediction. Comput. Electron. Agr. **187**, 106314 (2021). https://doi.org/10.1016/j.compag.2021.106314
44. Q. Liu, H. Xue, in *IJCAI International Joint Conference on Artificial Intelligence*, Adversarial spectral kernel matching for unsupervised time series domain adaptation. Virtual, Online, Canada, 2021, pp. 2744–2750, https://www.ijcai.org/proceedings/2021/0378.pdf