Multi-Source Spectral Learning Without Target-data Via Dual Regression Heads And Performance-Driven Domain Adaptive Weighting

Zhibin Xu[1], Chuanye He[2], Hailin Chen[1], and Meng Lei[3*]

[1] *China Certification & Inspection Group Hebei Co., Ltd., Shijiazhuang 050071, Hebei, China*
[2] *School of Software, Dalian University of Technology, Dalian 116620, Liaoning, China*
[3] *School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China*
[*] *Corresponding author. E-mail:lmsiee@cumt.edu.cn*

Received: Nov. 11, 2025; Accepted: Nov. 24, 2025

Near-Infrared Spectroscopy (NIRS) has become one of the most widely used non-destructive techniques for rapid coal quality assessment due to its efficiency, low cost, and suitability for on-site industrial deployment. However, NIRS data collected under different measurement conditions often exhibit significant spectral distribution shifts, which substantially weaken the cross-domain generalization ability of conventional prediction models. To address this challenge, this paper proposes a multi-source spectral learning method tailored for NIRS-based coal analysis. The method integrates a unified multi-source domain adaptation framework, a dual-regression-head architecture, and a performance-driven dynamic weighting strategy, enabling effective cross-domain learning without requiring target-domain data. Experiments conducted on real coal NIRS datasets demonstrate that the proposed approach consistently outperforms existing baselines across multiple measurement scenarios, highlighting its strong potential for accurate and robust rapid coal quality detection.

Keywords: Coal quality assessment; multi-source domain adaptation; cross-domain generalization; transfer learning

http://dx.doi.org/10.6180/jase.202607_30.035

1. Introduction

As the primary source of primary energy and electricity production in China, the quality parameters of coal directly affect combustion efficiency, energy utilization, and environmental emissions [1]. Conducting rapid and accurate coal quality testing helps improve coal utilization efficiency, optimize production processes, and reduce pollutant emissions such as carbon dioxide. Although traditional laboratory chemical analysis offers high precision, its complex procedures, high labor and time costs make it unsuitable for industrial sites requiring online, near real-time, and non-destructive testing. In recent years, technologies such as Laser-Induced Breakdown Spectroscopy (LIBS) [1], Near-Infrared Spectroscopy (NIRS) [2], and X-Ray Fluorescence Spectroscopy (XRF) [3] have shown great promise in rapid coal quality testing. With advantages such as high efficiency, speed, and being non-destructive or minimally destructive, these techniques have been widely used for the quantitative analysis of key indicators like moisture, volatile matter, and sulfur content [4]. Combined with deep learning methods, they further facilitate efficient modeling and accurate prediction of related indicators [5].

However, real-world industrial applications commonly face the challenge of domain shift, where a model's cross-domain generalization performance significantly declines when testing conditions deviate from the modeling conditions. A typical scenario involves spectral distribution variations caused by changes in the distance between the sampling device and the coal sample surface (e.g., 5, 10, 15, 20 mm). Alterations in the ranging geometry affect factors such as spot size and the ratio of diffuse to specular reflection, leading to systematic spectral shifts in the same sample at different distances. This consequently causes

a notable drop in cross-domain prediction accuracy [6]. Therefore, achieving effective model transfer and robust generalization under varying measurement conditions has become a critical issue that urgently needs to be addressed for the industrial deployment of coal quality spectral analysis.

To address distribution shifts, researchers have proposed various transfer and calibration methods. In the field of chemometrics, techniques like Direct Standardization (DS) and Piecewise Direct Standardization (PDS) utilize "co-measured samples" from the source and target domains to establish global or local linear mappings, enabling model transfer across instruments or conditions [7]. In machine learning and deep learning, statistical moment matching methods (e.g., Maximum Mean Discrepancy, MMD [8]; CORrelation ALignment, CORAL [9]) and adversarial domain adaptation approaches (e.g., Domain-Adversarial Neural Network, DANN [10]) enhance cross-domain generalization by aligning the feature distributions of the source and target domains.

Most of the above methods are limited to single-source domain scenarios, making it difficult to fully utilize multi-condition data (e.g., multi-range geometries) commonly encountered in industrial settings. In this context, Multi-Source Domain Adaptation (MSDA) has gradually emerged as a research hotspot [11]. By integrating knowledge from multiple source domains and collaboratively transferring it to the target domain, this approach has demonstrated promising performance in fields such as computer vision and medical imaging. Representative methods include M3SDA based on moment matching [12], MDAN based on adversarial confusion [13], and robust transfer methods based on source domain selection [14]. Meanwhile, recent advancements in multimodal and multi-view learning have also provided significant inspiration: for instance, graph-guided information bottleneck multi-view representation [15], progressive partial view alignment clustering [16], consistency-aware incomplete multi-modal alignment methods [17], and complementary modeling in cross-modal generation [18]. These approaches demonstrate that by fully leveraging the complementary information from multiple modalities or views, robust modeling can be achieved without relying on labeled data or annotations from the target domain, offering new insights for multi-source domain extension in coal spectral scenarios. Building on this, this paper attempts to extend single-source transfer to a multi-source transfer framework and explores achieving stable and reliable cross-domain predictions under conditions where target domain data is entirely unavailable.

Based on the above background, the main contributions and innovations of this paper are as follows:

- Multi-source domain adaptation modeling for multi-geometry ranging. To address spectral distribution variations caused by ranging changes (5, 10, 15, 20 mm), a unified multi-source domain modeling framework is constructed to achieve robust generalization under cross-domain conditions, filling the research gap in multi-source cross-domain modeling of coal spectra;

- The collaborative mechanism between the source-domain-specific regression head and the domain-invariant regression head. The source-domain-specific regression head is designed to retain task-specific information from the source domain, while the domain-invariant regression head learns transferable shared features. The two complement each other, effectively balancing task identifiability and cross-domain robustness;

- Domain-adaptive weight allocation based on performance discrepancy. A strategy is proposed to dynamically adjust training weights based on the predictive performance of the source domain, compelling the model to focus more on difficult-to-learn domains and avoid overfitting to easy-to-learn domains.

Systematic experimental results on actual collected coal spectral data demonstrate that the proposed method achieves stable and reliable cross-domain predictions under multiple ranging geometries, significantly outperforming other compared machine learning and deep learning approaches, validating its effectiveness and application potential in rapid coal quality detection scenarios.

2. Related works

2.1. Domain Adaptation Regression

As a crucial branch of transfer learning, DAR aims to address the prediction of continuous variables when there is a distribution discrepancy between the source and target domains. Compared to classification tasks, DAR is more challenging because it requires ensuring the accuracy of regression mapping while aligning distributions [19]. Early research primarily focused on distribution discrepancy metrics and theoretical analysis. For instance, proposed the discrepancy distance based on the Reproducing Kernel Hilbert Space, providing a theoretical foundation for error bounds and algorithm design in DAR. Subsequently, methods like CORAL and MMD improved generalization

by minimizing the statistical distribution differences between the source and target domains [8, 9]. In unsupervised scenarios, DARE-GRAM leverages the inverse Gram matrix to preserve the structural properties of linear regression solutions [20], while the COD method introduces conditional operator discrepancy based on kernel embeddings to achieve conditional distribution alignment [21]. In recent years, DAR applications in multi-source and dynamic environments have gained increasing attention. For example, Wu et al. [22] proposed a similarity-weighted multi-source integration approach, and Zhang et al. [23] combined continual learning to introduce dynamic DAR, addressing prediction challenges in complex industrial operating conditions.

2.2. Multi-Source Domain Adaptation

MSDA aims to enhance the learning performance of the target domain by leveraging knowledge from multiple source domains. Compared to single-source domain adaptation, MSDA better aligns with real-world application needs but also faces challenges such as distribution discrepancies among source domains and negative transfer [24]. Existing research can be broadly categorized into three approaches: The first involves distribution alignment methods, with representative examples including moment matching-based M3SDA [12] and category transfer-based DCTN [25]. The second category comprises adversarial approaches, such as MDAN [13] and maximum classifier discrepancy (MCD) [26], which learn domain-invariant representations through multi-discriminator adversarial training. The third category includes weighting and ensemble methods, which mitigate negative transfer by adaptively balancing source domain contributions via weighting mechanisms or attention [27]. Compared to single-source methods, MSDA can more effectively utilize complementary information from multiple source domains, demonstrating stronger practical value in complex application scenarios such as coal mine spectral analysis.

2.3. Domain-Adversarial Learning

DAL is a key approach to addressing cross-domain distribution discrepancies, with its core idea rooted in adversarial training mechanisms. By introducing a game-theoretic process between the feature extractor and domain discriminator, it compels the model to learn domain-invariant features, thereby enhancing generalization capabilities in the target domain. The initially proposed DANN framework systematically established the domain-adversarial concept and achieved end-to-end optimization through a gradient reversal layer [28]. Building on this, CDAN [29] further incorporated conditional distribution constraints to enhance feature discriminability, MADA [30] achieved fine-grained modeling of category-domain relationships via multi-domain discriminators, while AFN [31] leveraged feature norm regularization to improve cross-domain robustness. These methods effectively mitigate the limitations of traditional global adversarial approaches in modeling complex cross-domain relationships. In recent years, domain-adversarial learning has gradually expanded to multi-source domain adaptation [13] and unsupervised scenarios [26], better meeting the demands of diverse real-world applications.

3. Proposed method

3.1. Problem Definition

In the multi-source domain adaptation problem, we assume the existence of K source domains $\mathcal{D}_s = \{D_s^{(1)}, D_s^{(2)}, ..., D_s^{(K)}\}$ and one target domain $\mathcal{D}_t$. Each domain $D^{(i)}$ consists of an input space $\mathcal{X}^{(i)}$ and an output space $\mathcal{Y}^{(i)}$. The source domain data is $\{(x_s^{(i)}, y_s^{(i)})\}_{i=1}^{K}$ where $x_s^{(i)} \in \mathcal{X}_s^{(i)}$, and $y_s^{(i)} \in \mathcal{Y}_s^{(i)}$, the target domain data is $\{x_t\}$ where $x_t \in \mathcal{X}_t$, and the target domain labels are unknown.

In near-infrared spectroscopy analysis, different measurement conditions can lead to distribution shifts in spectral data. Let $X^{(i)}$ represent the spectral data of the i-th domain and $Y^{(i)}$ the corresponding target values. The domain adaptation problem can then be formulated as:

$$\min_{\theta} \sum_{i=1}^{K} \mathcal{L}_{task}(f_\theta(X_s^{(i)}), Y_s^{(i)}) + \lambda \mathcal{L}_{adapt}(X_s^{(1)}, ..., X_s^{(K)}) \quad (1)$$

Among them, $\mathcal{L}_{task}$ is the task loss, $\mathcal{L}_{adapt}$ is the domain adaptation loss, and λ is the balance parameter.

3.2. Feature Representation Preprocessing

To enhance the expressive power of spectral features, this paper designs a three-channel feature representation that transforms the original spectral data into a multichannel representation encompassing different information dimensions.As shown in Fig. 1 is the corresponding three-channel feature processing.

Channel 1 is the original spectrum, which retains all the information of the spectral data.

$$X_{raw} = X \quad (2)$$

Channel 2 employs Savitzky-Golay filtering, where the processed data not only accentuate spectral peak variations but also effectively suppress noise. In this step, a window size of 11 and a polynomial order of 2 are used.

$$X_{deriv} = \mathrm{SG}(X) \quad (3)$$

Channel 3 is the standard normal variation (SNV) transformation, which can eliminate spectral amplitude differences caused by scattering, uneven particle size, or variations in measurement conditions.

$$X_{snv} = \frac{X - \mu_X}{\sigma_X} \tag{4}$$

Among them, μ_X and σ_X are the mean and standard deviation of the sample, respectively.

The final feature is constructed by stacking the three channels, each with the same dimensionality, into a unified three-channel representation.

$$X_{3ch} = \text{Stack}([X_{raw}, X_{deriv}, X_{snv}]) \in \mathbb{R}^{N\times 3\times D} \tag{5}$$

3.3. Loss Function

This section will provide a detailed introduction to the task loss and the domain adaptation loss proposed in this paper. The overall loss function in this paper is the following.

$$\mathcal{L}_{total} = \mathcal{L}_{source} + \mathcal{L}_{shared} + \lambda_3\mathcal{L}_{domain} + \lambda_4\mathcal{L}_{align} \tag{6}$$

Among them, $\mathcal{L}_{source}$ and $\mathcal{L}_{shared}$ are task losses, while $\mathcal{L}_{domain}$ and $\mathcal{L}_{align}$ are domain adaptation losses. λ_3 and λ_4 are hyperparameters.

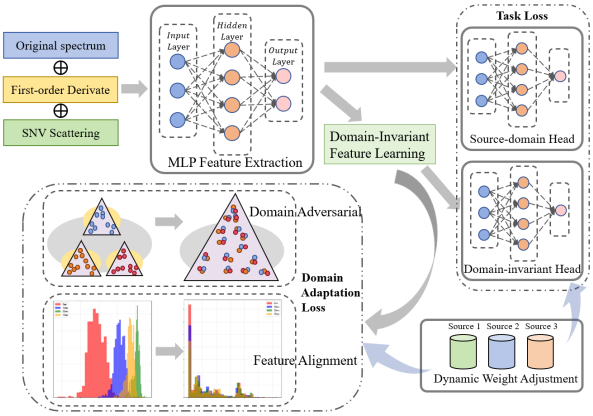

Fig. 1. Architecture of the multi-source domain adaptation model. Three-channel features are processed by an MLP, followed by dual regression heads for source fitting and target generalization. Domain-invariant features are then used for alignment and adversarial adaptation, with dynamic weight adjustment applied to the invariant regression head and adaptation loss

3.3.1. Task Loss

The task loss in this paper utilizes a dual-head regression architecture, which includes a source-domain-specific regression head and a domain-invariant regression head, each responsible for different learning objectives. The source-domain-specific regression head is solely used for calculating the loss and the final prediction result is entirely output by the domain-invariant regression head.

The source-domain-specific regression head makes predictions based on features learned by the network from processed raw data, ensuring the model's fitting capability on the source domain, with its loss defined as:

$$\begin{aligned}\mathcal{L}_{source} = &\frac{1}{K}\sum_{i=1}^{K}\frac{1}{N_i}\sum_{j=1}^{N_i}\Big[\lambda_1 \cdot MSE\left(f_{source}(g_{feat}(x_s^{(i,j)})), y_s^{(i,j)}\right)\\ &+ \lambda_2 \cdot MAE\left(f_{source}(g_{feat}(x_s^{(i,j)})), y_s^{(i,j)}\right)\Big]\end{aligned} \tag{7}$$

Among them, f_{source} is the head of the domain-specific regression of the source, g_{feat} is the feature extractor, MSE is the mean squared error, MAE is the mean absolute error.

The domain-invariant regression head learns domain-invariant features from data for prediction, enhancing cross-domain generalization capability, with its loss defined as:

$$\begin{aligned}\mathcal{L}_{shared} = &\frac{1}{K}\sum_{i=1}^{K} w_i \cdot \frac{1}{N_i}\sum_{j=1}^{N_i}\Big[\lambda_1 \cdot MSE\left(f_{shared}(g_{inv}(g_{feat}(x_s^{(i,j)}))), y_s^{(i,j)}\right)\\ &+ \lambda_2 \cdot MAE\left(f_{shared}(g_{inv}(g_{feat}(x_s^{(i,j)}))), y_s^{(i,j)}\right)\Big]\end{aligned} \tag{8}$$

Among them, f_{shared} is the head of domain-invariant regression, g_{inv} is the domain-invariant feature learning module, w_i is the dynamic weight of the i-th source domain.

The dynamic weight adjustment mechanism adaptively adjusts weights based on the actual performance of each source domain during training. It monitors the R^2 values at fixed epoch intervals and dynamically adjusts training weights according to cumulative performance gaps. This mechanism can automatically balance learning across source domains, significantly improving training effectiveness.

The performance of each source domain is as follows, measuring the predictive performance of the domain-invariant regression head.

$$R_s^{(i)} = R^2(f_{shared}(g_{inv}(g_{feat}(X_s^{(i)}))), Y_s^{(i)}) \tag{9}$$

Calculate the average performance of the last T epochs for each source domain. The averaging window T for performance calculation needs to balance weight adjustment stability and source-domain performance sensitivity. If T is too large, it causes a lag in weight adjustment and reduces responsiveness to challenging source domains. If T is too small, the calculation becomes highly sensitive to single-epoch noise, which may lead to weight oscillations.

$$\bar{R}^{(i)} = \frac{1}{T}\sum_{t=t_0}^{t_0+T-1} R_s^{(i,t)} \tag{10}$$

When the cumulative performance gap between the best-performing source domain and the worst-performing

source domain exceeds the threshold τ, weight adjustment is triggered, and the weight update rule is as follows.

$$\tilde{w}_i = \begin{cases} \min(w_{max}, w_i + \beta_+ \cdot (\bar{R}_{overall} - \bar{R}^{(i)})) & \text{if } \bar{R}^{(i)} < \bar{R}_{overall} \\ \max(w_{min}, w_i - \beta_- \cdot (\bar{R}^{(i)} - \bar{R}_{overall})) & \text{if } \bar{R}^{(i)} \geq \bar{R}_{overall} \end{cases} \tag{11}$$

Among them, $\bar{R}_{overall} = \frac{1}{K}\sum_{i=1}^{K} \bar{R}^{(i)}$, β_+, β_-, w_{min} and w_{max} are hyperparameters. Such settings effectively prevent training bias and maintain balanced contributions among multiple source domains.

Finally, weight normalization is performed to ensure training stability.

$$w_i = \frac{\tilde{w}_i}{\sum_{j=1}^{K} \tilde{w}_j} \cdot K \tag{12}$$

The source domain-specific head ensures the fitting capability of the source domain and stabilizes the performance of regression tasks, without being compromised by domain adversarial training. The domain-invariant head learns domain-agnostic features, ensuring generalization capability in the target domain. The two regression heads complement each other, simultaneously minimizing both the overall generalization error and the per-domain regression error, making the model more flexible and robust than single-head models, thereby enhancing the overall performance of the model.

3.3.2. Domain Adaptation Loss

The domain adaptation loss in this paper consists of feature alignment loss and domain adversarial loss. The purpose of feature alignment loss is to reduce the feature distribution differences between different source domains, while domain adversarial loss learns domain-invariant features by introducing an adversarial game mechanism.

The feature alignment loss is formed by a weighted sum of three sub-losses, which measure and narrow the feature distribution between source domain i and source domain j from different perspectives, promoting the alignment of feature distributions across source domains. The w_i in the following equation refers to the dynamic weight adjustment mechanism mentioned above.

$$\mathcal{L}_{align} = \frac{1}{K(K-1)} \sum_{i=1}^{K} \sum_{j \neq i} w_i \cdot \mathcal{L}_{align}^{(i,j)} \tag{13}$$

$$\mathcal{L}_{align}^{(i,j)} = \alpha \mathcal{L}_{mean}^{(i,j)} + \beta \mathcal{L}_{std}^{(i,j)} + \gamma \mathcal{L}_{mmd}^{(i,j)} \tag{14}$$

Among it, α, β and γ all are hyperparameters.

The mean alignment loss aligns the first-order statistics of features from two domains, namely their "mean" or "center", ensuring that features from different domains are generally located in the same place.

$$\mathcal{L}_{mean}^{(i,j)} = MSE(\mu_i, \mu_j) \tag{15}$$

where $\mu_i = \frac{1}{N_i}\sum_{k=1}^{N_i} g_{inv}(g_{feat}(x_s^{(i,k)}))$.

The standard deviation alignment loss aims to align the second-order statistics of features from two domains, namely their variance or standard deviation. It not only requires the feature centers of the two domains to coincide but also demands that their distribution shapes be similar.

$$\mathcal{L}_{std}^{(i,j)} = MSE(\sigma_i, \sigma_j) \tag{16}$$

Among it, $\sigma_i = \sqrt{\frac{1}{N_i}\sum_{k=1}^{N_i}(g_{inv}(g_{feat}(x_s^{(i,k)})) - \mu_i)^2}$.

The alignment combining first-order and second-order statistics essentially constrains the two distributions to share identical means and variances, a method commonly referred to as CORrelation ALignment (CORAL).

MMD is a more powerful and versatile non-parametric metric used to determine whether the feature distributions of two source domains are identical. It achieves this by calculating the distance between the embeddings of two distributions in a reproducing kernel Hilbert space.

$$\mathcal{L}_{mmd}^{(i,j)} = \text{MMD}(g_{inv}(g_{feat}(X_s^{(i)})), g_{inv}(g_{feat}(X_s^{(j)}))) \tag{17}$$

MMD loss calculation is

$$\begin{aligned} \text{MMD}(X, Y) =& \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} k(x_i, x_j) + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} k(y_i, y_j) \\ & - \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} k(x_i, y_j) \end{aligned} \tag{18}$$

Among them, the kernel function $k(x, y)$ is used to calculate the similarity between samples x and y in the high-dimensional feature space. If the two distributions are consistent, the mean of the within-domain similarity should be equal to the mean of the between-domain similarity, so the MMD value will be close to 0.

$$k(x, y) = \sum_{l=1}^{L} \exp\left(-\frac{\|x - y\|^2}{2\sigma_l^2}\right) \tag{19}$$

Among them, $\sigma_l = \sigma_0 \cdot 2^l$, where σ_0 is the adaptive bandwidth parameter. Using multi-scale Gaussian kernels (i.e., employing multiple bandwidths) can capture distribution features at different scales. Large bandwidths capture the macroscopic overall distribution structure, while small bandwidths capture microscopic local detail distributions.

The feature alignment loss progressively narrows the representation gap between source domains by aligning first-order, second-order, and higher-order kernel statistics, mitigating distribution shifts. Subsequently, the domain adversarial loss maximizes domain confusion from the discriminator's perspective while minimizing domain separability from the representation's viewpoint, approximating domain-invariant representations.

Domain adversarial loss achieves domain-invariant feature learning through a gradient reversal layer.

$$\mathcal{L}_{domain} = \frac{1}{K}\sum_{i=1}^{K} w_i \cdot \frac{1}{N_i}\sum_{j=1}^{N_i} \text{CrossEntropy}(D(\text{GRL}(g_{inv}(g_{feat}(x_s^{(i,j)})), \alpha)), d_i) \tag{20}$$

Among them, D is the domain classifier; GRL is the gradient reversal layer; α is the gradient reversal intensity parameter, which controls the intensity of adversarial training by adopting a sigmoid schedule; d_i is the label of the i-th domain, $d_i \in \{1, 2, 3\}$.

First, forward propagation is performed through the gradient reversal layer, which is $\text{GRL}(x, \alpha) = x$, then multiplied by the negative coefficient α, and backpropagation is carried out using $\frac{\partial \text{GRL}(x,\alpha)}{\partial x} = -\alpha \cdot \frac{\partial \mathcal{L}}{\partial x}$. When optimizing the domain classifier D, gradients propagate normally, allowing the parameters of the domain classifier D to be updated for better classification. However, when optimizing the feature extractor, the GRL reverses the gradient sign, meaning the feature extractor updates its parameters in the direction that increases the domain classification loss, effectively striving to make the features harder for the domain classifier to distinguish.

4. Experiments

4.1. Experimental Details

The experiments in this paper were conducted on a Linux operating system, with hardware environment being a CUDA-enabled GPU and development environment being Python 3.10. The main software libraries relied on for the experiments include PyTorch for building and training deep learning models, as well as NumPy and Pandas for data preprocessing and management. During training, the Adam optimizer is used with a fixed learning rate of 1×10^{-3}, without using a scheduler, and set weight decay to 1×10^{-3}.

This training consists of a total of 1,000 epochs, with the following hyperparameter settings: in the dual regression head loss, the MSE weight λ_1 is set to 0.7, the MAE weight λ_2 is set to 0.3, the domain adversarial loss weight λ_3 is 0.5 × progress, and the feature alignment loss weight λ_4 is 0.2 × progress. Here, progress is the current training epoch divided by the total number of epochs, representing the training progress. This means the domain adversarial loss weight increases from 0 to 0.5 as training progresses, and the feature alignment loss weight increases from 0 to 0.2. The weights of the three source domains are initialized to 1, with a weight update threshold τ of 0.1. The gradient reversal intensity parameter α adopts a sigmoid schedule.

$$\alpha = \frac{2}{1 + \exp(-10 \cdot progress)} - 1 \tag{21}$$

The dataset used in this study was collected using the MicroNIR Pro handheld near-infrared spectrometer. The instrument operates over a spectral range of 908.1–1676.2 nm with a sampling interval of 6.24 nm, yielding 125 wavelength points (corresponding to 125 spectral bands). Spectral measurements of each coal sample were acquired at four sampling distances of 5, 10, 15, and 20 mm. After preprocessing and quality screening, we obtained spectral feature data for 781 samples, each paired with its corresponding air-dried moisture (Mad) content label. The experiment involves training models on multiple source domain datasets and then transferring them to the target domain for testing and evaluation, thereby validating the effectiveness of the proposed multi-source domain transfer method.

4.2. Comparison Model Introduction

To comprehensively evaluate the effectiveness of the proposed method, this paper compares it with various typical domain adaptation approaches, covering different strategies such as adversarial learning and distribution alignment. The following is a brief introduction to the basic ideas of each method.

- DANN [13] reduces the differences in distribution between domains by introducing a gradient reversal layer and a domain classifier, allowing the feature extractor to learn domain-invariant features.
- MMD [8] achieves domain alignment by minimizing the distribution discrepancy between the source and target domains in the feature space.
- CORAL [9] achieves distribution calibration by matching second-order statistics (covariance matrix) of the characteristics of the source and target domains.
- CDAN [29] introduces DANN-based conditional distribution adversarial learning, using the joint distribution of features and predicted labels for discrimination, thereby enhancing the discriminative ability of transfer between domains.
- MDD [32] method measures the discrepancy between the source and target domains by constructing a loss of inconsistency based on margins and minimizes this gap during optimization.

Table 1. Comparative experimental results of dual-source domain adaptation (10mm+20mm->15mm)

Methods	DANN	CDAN	ADDA	MMD	CORAL	MDD	SHOT	DIRT-T	ours
R^2	0.2439	0.3181	0.2634	0.3022	0.2905	0.2558	0.2593	0.2651	**0.6457**
RMSE	2.7971	2.6563	2.7607	2.6870	2.7095	2.7749	2.7685	2.7576	**1.9145**
MAE	1.9864	1.8667	2.0311	1.9519	1.9042	2.0266	1.9609	1.9021	**1.2976**

Table 2. Comparative experimental results of dual-source domain adaptation (5mm+10mm->15mm)

Methods	DANN	CDAN	ADDA	MMD	CORAL	MDD	SHOT	DIRT-T	ours
R^2	-0.0876	-0.0539	-0.2469	-0.3072	-0.0735	-0.2729	-0.0091	-0.1626	**0.5193**
RMSE	3.3546	3.3022	3.5919	3.6777	3.3328	3.6291	3.2313	3.4683	**2.1987**
MAE	2.2451	2.1802	2.5882	2.4581	2.2085	2.6732	2.3408	2.2737	**1.6153**

- SHOT [33] does not require access to source domain data, but instead performs adaptive optimization on the target domain based on pretrained model parameters from the source domain and a pseudo-label generation mechanism.
- ADDA [34] employs adversarial learning to achieve the alignment of the feature space by learning a discriminator between the source and target domains.
- DIRT-T [35] enhances the effect of self-training by iteratively refining the decision boundaries in the target domain, thus improving the discriminative capacity of the model in the target domain.

4.3. Evaluation Metrics

In the experiment, coefficient of determination (R^2), root mean square error ($RMSE$), and mean absolute error(MAE) were used to evaluate model performance.

The coefficient of determination (R^2) reflects the extent to which the model explains the variance of the dependent variable, with values closer to 1 indicating a better fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{22}$$

Among them, y_i represents the true value, $\hat{y}_i$ represents the predicted value, $\bar{y}$ is the mean of the true values, n is the number of samples.

Root Mean Square Error ($RMSE$) measures the deviation between predicted and actual values, with smaller values indicating higher prediction accuracy.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{23}$$

Mean Absolute Error (MAE) represents the average absolute deviation between predicted results and actual values, providing an intuitive reflection of the error magnitude.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{24}$$

4.4. Comparative Experiment

To further validate the effectiveness of the proposed method, this paper conducts experiments under two different settings (dual-source domain and triple-source domain) and compares with multiple baseline approaches. It is worth noting that all baseline models included in the comparison were implemented using the same network architecture and training configurations to ensure a fair and consistent evaluation. By comparing the predictive performance of different methods on the same task, it intuitively demonstrates the advantages of the proposed model in cross-domain regression tasks.

4.4.1. Dual-source Domain Adaptation

In this section, two out of the three available source domains (5 mm, 10 mm, and 20 mm) are selected for training in each experiment, while the remaining domain (15 mm) serves as the target domain for testing. These distance values represent the sampling distances between the probe and the sample. Tables 1 to 3 summarize the performance of different methods under various source-domain settings in terms of R^2, $RMSE$, and MAE. The bold results indicate the best performance, and the underlined ones denote the second-best results.

Table 1 presents the results obtained when using the 10 mm and 20 mm domains as sources and the 15 mm domain as the target. Since the selected source domains are close to the target domain and jointly provide complementary information from both near and far sampling distances, they effectively cover the target domain's feature distribution. Consequently, the learned domain-invariant representations are more comprehensive and transferable. Under this

Table 3. Comparative experimental results of dual-source domain adaptation (5mm+20mm->15mm)

Methods	DANN	CDAN	ADDA	MMD	CORAL	MDD	SHOT	DIRT-T	ours
R^2	0.1508	0.1198	0.2286	0.2195	0.2202	0.2619	0.1562	0.0775	**0.6062**
RMSE	2.9643	3.0178	2.8252	2.8418	2.8406	2.7636	2.9548	3.0895	**2.0185**
MAE	1.9864	1.8667	2.0311	1.9519	1.9042	2.0266	1.9609	1.9021	**1.3475**

Table 4. Comparative experimental results of triple-source domain adaptation

Methods	DANN	CDAN	ADDA	MMD	CORAL	MDD	SHOT	DIRT-T	ours
R^2	0.3371	0.2473	0.3992	0.3569	0.3749	0.3942	0.3573	0.2920	**0.7336**
RMSE	2.6191	2.7906	2.4932	2.5795	2.5431	2.5036	2.5787	2.7066	**1.6601**
MAE	1.8109	2.0455	1.6888	1.8116	1.7601	1.7116	1.8020	1.9420	**1.0987**

dual-source configuration, the proposed model achieves the best overall performance (R^2 = 0.6457, $RMSE$ = 1.9145, MAE = 1.2976), outperforming all baseline methods.

In contrast, Table 2 shows the results for the 5 mm and 10 mm source-domain setting. Because both source domains have smaller probe–sample distances than the target domain (15 mm), their spectral distributions are relatively similar and lack information from farther distances. As a result, the learned domain-invariant features exhibit weaker generalization capability, leading to slightly lower performance (R^2 = 0.5193, $RMSE$ = 2.1987, MAE = 1.6153). Nevertheless, the proposed framework still significantly surpasses all baseline approaches, demonstrating that it can capture stable and task-relevant representations even under limited domain diversity.

For the Table 3 setting with 5 mm and 20 mm as source domains, the domain discrepancy is the largest, representing an extreme condition of distance shift. Although this setup covers a broader measurement range, the substantial spectral differences between the 5 mm and 20 mm datasets increase the difficulty of domain alignment. Consequently, compared with the (10 mm + 20 mm) setting, the model achieves robust but slightly inferior performance (R^2 = 0.6062, $RMSE$ = 2.0185, MAE = 1.3475). This observation suggests that excessive domain discrepancy may hinder effective domain-invariant feature learning, although the proposed dynamic weighting mechanism still manages to balance the contributions of multiple source domains effectively.

4.4.2. Triple-source Domain Adaptation

Table 4 presents the performance of different methods in terms of R^2, $RMSE$, and MAE. When extended to the three-source scenario, the proposed model achieves the best overall performance, with R^2 increasing to 0.7336 and $RMSE$ and MAE decreasing to 1.6601 and 1.0987, respectively. Integrating all three source domains enables the model to fully exploit complementary spectral information, expand the coverage of domain characteristics, and enhance its generalization capability to the target domain.

4.4.3. Results and Visual Analysis

The performance advantages can be attributed to three main aspects: (1) The multi-head regression architecture effectively enhances the generalization capability in the target domain while ensuring source domain fitting. (2) The combination of domain adversarial training and feature alignment strategies, through gradient reversal, distribution statistical matching, and multi-scale MMD, significantly reduces inter-domain discrepancies. (3) The dynamic weight adjustment mechanism enables flexible allocation of learning resources when multi-source distribution differences are substantial, further improving adaptability. In summary, both dual-source and triple-source experiments validate the superior performance and robustness of the proposed method in cross-domain regression tasks.

A comprehensive analysis of the dual-source and triple-source domain experimental results suggests that the contribution differences among source domains are primarily determined by their domain shift degree relative to the target domain and their respective data distribution characteristics. The source domain closest to the target domain in sampling distance and showing the highest spectral distribution overlap exhibits the weakest domain shift effect. This domain provides the model with fundamental transferable features and ensures baseline accuracy in cross-domain predictions. In contrast, source domains with larger domain shifts cover a wider range of measurement scenarios and supplement the model with additional spectral characteristics from both near and far measurement distances. This complementary information effectively enhances the model's adaptability to variations in domain shift.

Fig. 2 illustrates the low-dimensional distribution of

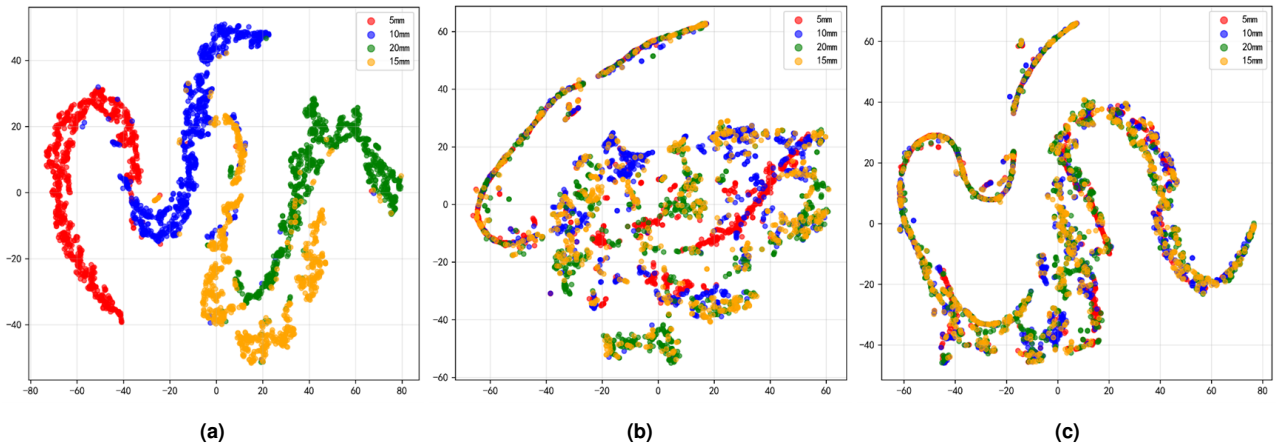

Fig. 2. t-SNE feature visualization. From left to right are the original features, MLP features, and domain-invariant features

three feature representations during model training. As shown in Fig. 2a, the original spectral features from different sampling distances are clearly separated, exhibiting distinct domain-specific clusters. This reflects substantial inter-domain discrepancies, making direct modeling prone to poor target-domain generalization. After nonlinear feature extraction (Fig. 2b), the boundaries between domains become less pronounced, with partial overlap emerging in the feature space, indicating improved cross-domain consistency. However, notable domain differences remain and full alignment is not yet achieved. In contrast, Fig. 2c shows that the final domain-invariant features form a highly mixed distribution with substantial overlap across domains, demonstrating that the proposed adversarial training and alignment strategy effectively reduces inter-domain discrepancies and enables the model to learn robust domain-invariant representations.

Fig. 3. The scatter plot of predicted versus true Mad values

Fig. 3 illustrates the relationship between the predicted and true Mad values in the target domain. Overall, most data points are densely distributed along the ideal fit line ($y = x$), indicating a strong linear consistency and correlation between the predicted and actual values, which reflects the model's excellent regression performance. In general, the model exhibits high predictive accuracy in the target domain, with the majority of samples showing errors within a reasonable range. However, it can be observed that prediction errors tend to increase as the Mad value becomes larger. This phenomenon may result from enhanced spectral nonlinearity and reduced signal-to-noise ratio under high-moisture conditions, which increase the model's prediction uncertainty in that range.

Figs. 4 and 5 provide an intuitive assessment of the model's predictive performance and error distribution in the target domain. As shown in Fig. 4, the predicted values closely follow the ground truth across a wide range of moisture levels, reflecting strong fitting capability and generally small deviations. Noticeable fluctuations appear only in the high-moisture region, indicating reduced model stability under elevated humidity conditions, likely due to increased spectral nonlinearity and a diminished signal-to-noise ratio. Fig. 5 further illustrates that most samples fall within the ±MAE interval (green-shaded region), demonstrating low and uniformly distributed errors overall. Only a small number of samples exceed twice the MAE, primarily concentrated in the high-moisture range, which further confirms the presence of systematic deviations under such conditions. Overall, the model exhibits strong predictive stability and generalization in low and medium moisture ranges, while performance under high-moisture conditions remains an area for potential improvement.

4.5. Ablation Study

To validate the effectiveness of each module in the model, ablation experiments were conducted under the three-source-domain condition. By sequentially removing key modules from the model—including the source-specific regression head $\mathcal{L}_{source}$, domain-invariant regression head

Table 5. Ablation experiment results

$\mathcal{L}_{source}$	$\mathcal{L}_{shared}$	$\mathcal{L}_{domain}$	$\mathcal{L}_{align}$	$\mathcal{W}$	R^2	*RMSE*	*MAE*
	✓	✓	✓	✓	0.6831	1.8108	1.2350
✓		✓	✓	✓	0.6723	1.8413	1.2820
✓	✓		✓	✓	0.6687	1.8515	1.2743
✓	✓	✓		✓	0.6589	1.8787	1.2464
✓	✓	✓	✓		0.6718	1.8428	1.2644
✓	✓	✓	✓	✓	**0.7336**	**1.6601**	**1.0987**

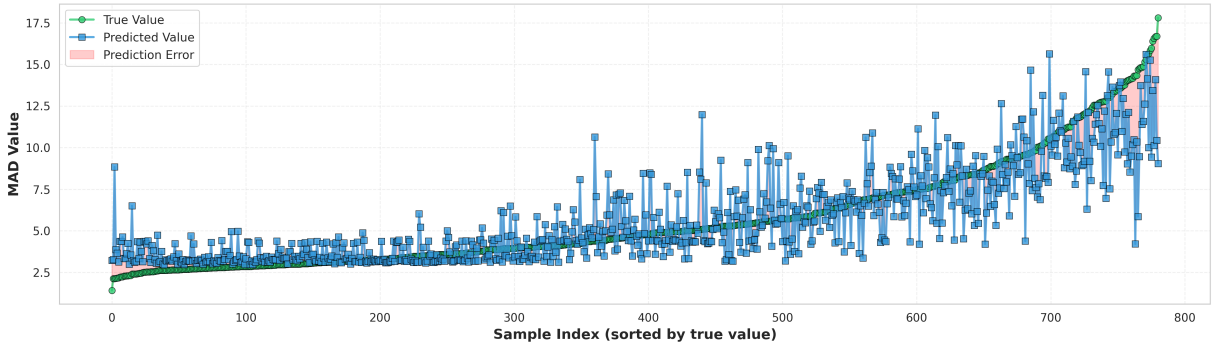

Fig. 4. Comparison of model-predicted Mad values with true values across the sample sequence (sorted by true labels)

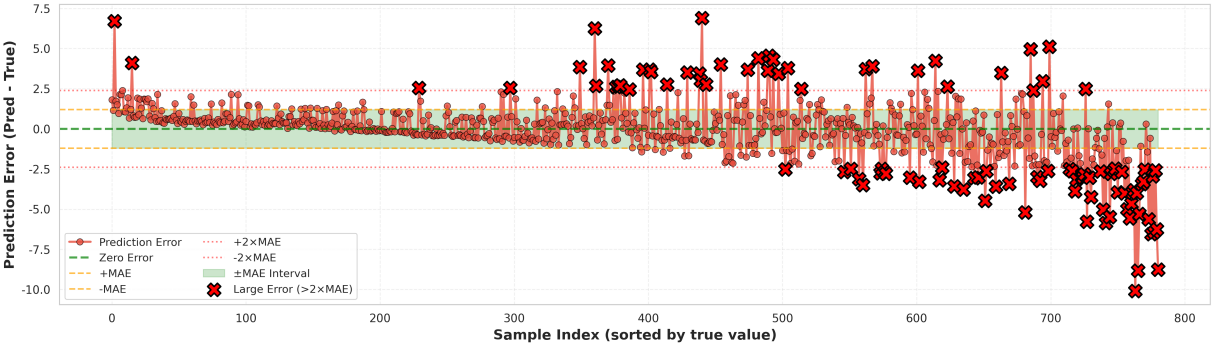

Fig. 5. Variation of prediction error relative to the average error across the sample sequence (sorted by true labels)

$\mathcal{L}_{shared}$, domain adversarial training $\mathcal{L}_{domain}$, feature alignment $\mathcal{L}_{align}$, and dynamic weight adjustment strategy $\mathcal{W}$. The performance variations under different module omissions were compared. Table 5 presents the ablation experiment results, where ✓ indicates the use of the module, a blank space indicates its absence and bold indicates the optimal result.

As can be seen from Table 5, the removal of any module leads to a decline in model performance, indicating that each module plays a positive role in improving overall performance. Among them, the removal of the feature alignment loss results in the most significant performance drop, demonstrating that distribution alignment is crucial for reducing inter-domain differences and enhancing generalization capabilities in cross-domain tasks. Meanwhile, other modules also contribute significantly to improving model stability and prediction accuracy. In summary, only when all modules work together can the model achieve optimal performance, which fully validates the rationality and necessity of the proposed overall architecture design.

5. Conclusions

This paper proposes a cross-domain modeling framework for multi-source domain regression tasks, with its core contributions reflected in the following aspects. Firstly, a multi-head regression structure is introduced, combining source-domain-specific regression heads with domain-invariant regression heads to achieve an effective balance between source-domain fitting and target-domain generalization. Domain adversarial training in domain adaptation progres-

sively reduces distribution discrepancies between different domains, enhancing the domain invariance of features. Simultaneously, feature alignment strategies, through mean-variance alignment and multi-scale MMD loss, further narrow the differences between source domains at both statistical and distributional levels, ensuring consistency in the feature space. The proposed dynamic weight adjustment mechanism adaptively allocates training weights based on the performance of each source domain, thereby preventing any single source domain from dominating the model training and enhancing overall robustness.

In the comparison experiments between dual-source and triple-source scenarios, the proposed method significantly outperformed various typical domain adaptation approaches across all metrics, and ablation studies further validated the effectiveness of each module's design. Future work could explore more efficient alignment methods and lightweight model architectures based on this foundation to promote the application of multi-source domain adaptation in more complex scenarios.

6. Acknowledgement

This work was financially supported by the R&D Project of China Certification & Inspection Group Hebei Co., Ltd. (2025ZJHBYF004-1) and the Scientific Research Project of the General Administration of Customs (2023HK113).

References

[1] Q. Wang, D. Zhang, Y. Dou, Q. Wang, and Y. Wang, (2025) *"Advanced Multi-Parameter Prediction of Coal Quality Using LIBS and Ensemble Machine Learning Techniques"* **ACS Omega *10***(33): 37574–37582.

[2] L. Zou, J. Qiao, X. Yu, X. Chen, and M. Lei, (2023) *"Intelligent proximate analysis of coal based on near-infrared spectroscopy and multioutput deep learning"* **IEEE Trans. Artif. Intell. *5***(3): 1398–1410.

[3] J. Li, R. Gao, Y. Zhang, L. Zhang, L. Dong, W. Ma, W. Yin, and S. Jia, (2025) *"Research on accurate analysis of coal quality using NIRS-XRF fusion spectroscopy in complex coal type scenarios"* **Opt. Laser Technol. *181***: 111734.

[4] J. Tian, M. Li, X. Zhang, M. Lei, L. Ke, and L. Zou, (2024) *"Enhancing Moisture Detection in Coal Gravels: A Deep Learning-Based Adaptive Microwave Spectra Fusion Method"* **Spectrochim. Acta Part A Mol. Biomol. Spectrosc. *313***: 124147.

[5] L. Zou, S. Kou, Z. Xu, Z. Tan, and M. Lei, (2025) *"Simultaneous prediction of bauxite quality parameters using TC-Unet and near-infrared spectroscopy"* **Memetic Computing *17***(2): 25.

[6] B. Hu and J. Wang, (2022) *"A weighted multi-source domain adaptation approach for surface defect detection"* **IET Image Process. *16***(8): 2210–2218.

[7] A. J. Parrott, A. C. McIntyre, M. Holden, G. Colquhoun, Z.-P. Chen, D. Littlejohn, and A. Nordon, (2022) *"Calibration model transfer in mid-infrared process analysis with in situ attenuated total reflectance immersion probes"* **Anal. Methods *14***(19): 1889–1896.

[8] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, (2012) *"A kernel two-sample test"* **J. Mach. Learn. Res. *13***(1): 723–773.

[9] B. Sun and K. Saenko. "Deep coral: Correlation alignment for deep domain adaptation". In: *Eur. Conf. Comput. Vis.* 2016, 443–450.

[10] M. Hassan Pour Zonoozi, V. Seydi, and M. Deypir, (2025) *"An unsupervised adversarial domain adaptation based on variational auto-encoder"* **Mach. Learn. *114***(5): 1–26.

[11] S. Zhao, G. Wang, S. Zhang, Y. Gu, Y. Li, Z. Song, P. Xu, R. Hu, H. Chai, and K. Keutzer. "Multi-source distilling domain adaptation". In: *Proc. AAAI Conf. Artif. Intell.* ***34***. 07. 2020, 12975–12983.

[12] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang. "Moment matching for multi-source domain adaptation". In: *Proc. IEEE/CVF Int. Conf. Comput. Vis.* 2019, 1406–1415.

[13] H. Zhao, S. Zhang, G. Wu, J. M. F. Moura, J. P. Costeira, and G. J. Gordon. "Adversarial multiple source domain adaptation". In: *Adv. Neural Inf. Process. Syst.* ***31***. 2018.

[14] L. Yang, Y. Balaji, S.-N. Lim, and A. Shrivastava. "Curriculum manager for source selection in multi-source domain adaptation". In: *Eur. Conf. Comput. Vis.* 2020, 608–624.

[15] L. Zhao, X. Wang, Z. Liu, Z. Wang, and Z. Chen, (2024) *"Learnable Graph Guided Deep Multi-view Representation Learning via Information Bottleneck"* **IEEE Trans. Circuits Syst. Video Technol. *35***(4): 3303–3314.

[16] L. Zhao, Q. Xie, Z. Li, S. Wu, and Y. Yang, (2024) *"Dynamic Graph Guided Progressive Partial View-Aligned Clustering"* **IEEE Trans. Neural Netw. Learn. Syst. *36***(5): 9370–9382.

[17] S. Ma, L. Zhao, M. Lu, Y. Guo, and B. Xu. *Consistency-Aware Padding for Incomplete Multi-Modal Alignment Clustering Based on Self-Repellent Greedy Anchor Search*. [Preprint]. 2025.

[18] L. Zhao, P. Huang, T. Chen, C. Fu, Q. Hu, and Y. Zhang, (2023) *"Multi-sentence complementarily generation for text-to-image synthesis"* **IEEE Trans. Multimed. *26***: 8323–8332.

[19] C. Cortes and M. Mohri. "Domain adaptation in regression". In: *Int. Conf. Algorithmic Learn. Theory*. 2011, 308–323.

[20] I. Nejjar, Q. Wang, and O. Fink. "Dare-gram: Unsupervised domain adaptation regression by aligning inverse gram matrices". In: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 2023, 11744–11754.

[21] H.-R. Yang, C.-X. Ren, and Y.-W. Luo. "Cod: Learning conditional invariant representation for domain adaptation regression". In: *Eur. Conf. Comput. Vis.* 2024, 108–125.

[22] Y. Wu, G. Parmigiani, and B. Ren. *Multi-source domain adaptation for regression*. [Preprint]. 2023.

[23] Y. Zhang, G. Yan, G. Zhao, S. Ma, Z. Liu, and G. Zhao, (2024) *"A Dynamic Domain Adaptation Regression Method for Multiple Working Conditions Based on Continual Learning"* **Ind. Eng. Chem. Res. *63***(40): 17252–17265.

[24] Y. Mansour, M. Mohri, and A. Rostamizadeh. "Domain adaptation with multiple sources". In: *Adv. Neural Inf. Process. Syst.* ***21***. 2008.

[25] R. Xu, Z. Chen, W. Zuo, J. Yan, and L. Lin. "Deep cocktail network: Multi-source unsupervised domain adaptation with category shift". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2018, 3964–3973.

[26] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada. "Maximum classifier discrepancy for unsupervised domain adaptation". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2018, 3723–3732.

[27] Y. Zuo, H. Yao, and C. Xu, (2021) *"Attention-based multi-source domain adaptation"* **IEEE Trans. Image Process. *30***: 3793–3803.

[28] Y. Ganin and V. Lempitsky. "Unsupervised domain adaptation by backpropagation". In: *Int. Conf. Mach. Learn.* 2015, 1180–1189.

[29] M. Long, Z. Cao, J. Wang, and M. I. Jordan. "Conditional adversarial domain adaptation". In: *Adv. Neural Inf. Process. Syst.* ***31***. 2018.

[30] Z. Pei, Z. Cao, M. Long, and J. Wang. "Multi-adversarial domain adaptation". In: *Proc. AAAI Conf. Artif. Intell.* ***32***. 1. 2018.

[31] R. Xu, G. Li, J. Yang, and L. Lin. "Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation". In: *Proc. IEEE/CVF Int. Conf. Comput. Vis.* 2019, 1426–1435.

[32] Y. Zhang, T. Liu, M. Long, and M. I. Jordan. "Bridging theory and algorithm for domain adaptation". In: *Int. Conf. Mach. Learn.* 2019, 7404–7413.

[33] J. Liang, D. Hu, and J. Feng. "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation". In: *Int. Conf. Mach. Learn.* 2020, 6028–6039.

[34] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. "Adversarial discriminative domain adaptation". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2017, 7167–7176.

[35] R. Shu, H. H. Bui, H. Narui, and S. Ermon. *A dirt-t approach to unsupervised domain adaptation*. [Preprint]. 2018.