A multi-task model for mill load parameter prediction using physical information and domain adaptation

Validation with laboratory ball mill

Liu, Yiwen; Yan, Gaowei; Xiao, Shuyi; Wang, Fang; Li, Rong; Pang, Yusong

DOI
10.1016/j.mineng.2024.109148

Publication date
2025

Document Version
Final published version

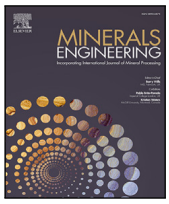
Published in
Minerals Engineering

Minerals Engineering 222 (2025) 109148

Contents lists available at ScienceDirect

Minerals Engineering

journal homepage: www.elsevier.com/locate/mineng



A multi-task model for mill load parameter prediction using physical information and domain adaptation: Validation with laboratory ball mill

Yiwen Liu [a], Gaowei Yan [a,*], Shuyi Xiao [a], Fang Wang [a], Rong Li [a], Yusong Pang [b]

[a] College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, 030024, China
[b] Section of Transport Engineering and Logistics, Delft University of Technology, Mekelweg 2, Netherlands

ARTICLE INFO

Keywords:
Physical information
Physical-data-driven hybrid model
Multi-task
Mill load parameter
Domain adaptation

ABSTRACT

Accurate prediction of mill load parameters is crucial to improving grinding efficiency and saving energy. Traditional prediction models have challenges such as poor interpretability, low prediction efficiency and differences in data distribution. This study innovatively proposed a multi-task prediction model that integrates physical information and domain adaptation. By constructing a physical-data-driven hybrid model, the physical relationship between mill load parameters is embedded into the model as prior knowledge to improve the prediction accuracy of the model. At the same time, multi-task learning is used to predict the material-to-ball volume ratio and the pulp density at the same time, which improves efficiency and reduces repetitive work. The domain adaptation method is introduced to ensure that the model maintains stable prediction performance when the data distribution changes. Laboratory ball mill data verification shows that the proposed model not only improves the prediction accuracy, but also adapts well to variable working conditions, showing significant superiority.

1. Introduction

The ball mill is an important mechanical equipment in the grinding process and is also a high-energy-consuming equipment (Lu et al., 2014; Yin et al., 2019). In the mineral processing industry, its energy consumption accounts for about 40%–70% of the mineral processing plant (Daniel, 2016). The grinding particle size, a key quality indicator of the grinding process, has a direct impact on product quality and subsequent processes (Dai and Chai, 2014; Zhou et al., 2009). Inappropriate grinding particle size will not only affect the efficiency of subsequent flotation operations, reduce the recovery rate and economic benefits of high-value minerals, but also cause environmental pollution (Dai et al., 2015; Pease et al., 2006). Mill load parameters such as material-to-ball volume ratio (MBVR) and pulp density (PD) are key parameters affecting grinding particle size and energy consumption (Tang et al., 2020c). Due to the limitations of physical conditions, it is usually difficult to achieve online measurement of mill load parameters in actual production (Zhuo et al., 2021). Operators can only perceive the operating status of the ball mill based on the particle size test value and by observing the mill current and mechanical signals. This cannot ensure that the ball mill is in the best working state for a long time, and it is easy to cause faults such as overload (Tang et al., 2018a, 2012). Therefore, establishing an accurate mill load parameter prediction model is of great significance to improving grinding efficiency, reducing energy consumption, and ensuring safety.

In existing research, scholars have modeled the grinding process of ball mills using methods such as finite element method (FEM) and discrete element method (DEM) (Jonsén et al., 2014; Mayank et al., 2015; Hilden et al., 2021). The mechanism model mainly relies on the understanding of the internal mechanisms of industrial processes to establish the model (Xia et al., 2024). With the development of artificial intelligence, data-driven models have also been gradually applied to the prediction of mill load parameters. Data-driven models mainly construct prediction models by mining the potential relationships between data without the need to understand detailed physical information (Habib et al., 2021). Commonly used data-driven mill load parameter prediction models include support vector machines (Liu et al., 2015; Cai et al., 2021), partial least squares (Tang et al., 2020b), neural networks (Yang and Cai, 2021), etc. Although pure data-driven models can capture the complex relationships between data to a certain extent, they have problems such as poor interpretability and unstable model prediction results (Frost Jr and Quinn, 2018; Zhang et al., 2024).

At present, more and more scholars are focusing on physical-data-driven hybrid models, hoping to overcome the shortcomings of physical

* Corresponding author.
E-mail address: yangaowei@tyut.edu.cn (G. Yan).

https://doi.org/10.1016/j.mineng.2024.109148
Received 13 September 2024; Received in revised form 30 November 2024; Accepted 3 December 2024
Available online 13 December 2024

models and data driven models (Sansana et al., 2021; Gao et al., 2024). Cao et al. (2024) constructed a physical-data-driven hybrid model and used it for the remaining life prediction of bearings, making the prediction results more in line with physical laws. Syauqi et al. (2024) combined the physical model with the data driven model to construct a hybrid model for solar photovoltaic power generation prediction, making the model more interpretable and accurate. Xin et al. (2024) used the constructed hybrid model for low calorific value fuel gas turbine load prediction to improve the model prediction accuracy. The above research results show that combining physical information with data-driven models can improve the prediction accuracy of the model and make the prediction results more consistent with physical laws (Alber et al., 2019). Although the physical-data-driven hybrid model has demonstrated its good prediction performance in many fields, there is still a lack of related work in the prediction of mill load parameters.

Summarizing the above models, we also found that the existing mill load parameter prediction models, whether mechanism models or data-driven models, are single-task (ST) models. When predicting mill load parameters, we usually need to complete multiple prediction tasks. In this case, the traditional single-task model needs to perform independent model training for each prediction task. This process not only increases the experimental cost, but also leads to a lot of repetitive work, affecting the overall prediction efficiency (Wang et al., 2020). The emergence of multi-task models provides a solution to the above problems. Multi-task models can learn multiple tasks at the same time, and the knowledge in one task can also be used by other tasks, thereby improving the generalization performance and prediction accuracy of the model (Zhang and Yang, 2021). Hard parameter sharing (HPS) is the most commonly used multi-task model, which mainly includes a shared layer and a task-specific output layer (Ruder, 2017). Due to the existence of the shared layer, HPS may cause the model prediction accuracy to decrease due to task conflicts, that is, the negative transfer problem (Tang et al., 2020a). The emergence of methods such as Multi-gate Mixture-of-Experts (MMoE) (Ma et al., 2018) and Customized Gate Control (CGC) (Tang et al., 2020a) provides solutions to avoid the above problems. The MMoE and CGC mainly add gating networks compared to HPS. The gating network controls the transmission of information and effectively alleviates the negative transfer problem.

The above multi-task model requires the same data distribution, that is, the data comes from the same working condition (Curreri et al., 2021). However, in the actual production process, due to different environments, set values, and operating conditions, the collected ball mill data are usually data from different working conditions (Gao-Wei et al., 2018; Liu et al., 2023). In this case, the traditional multi-task model will have the problem of reduced model prediction accuracy. Domain adaptation, as a type of transfer learning, is mainly applicable to situations where there are differences in data distribution between the source domain and the target domain, but the tasks of the two domains are the same (Pan and Yang, 2009). He et al. (2019) introduced manifold regularization to maintain the geometric structure of the data, and introduced a domain adaptation random weight neural network to propose a mill load parameter prediction model under variable working conditions. Zhang et al. (2022) aligned the marginal distribution and conditional distribution of historical working condition data and current working condition data to achieve soft measurement of mill load parameters under variable working conditions. Huang et al. (2022) proposed a joint discriminant high-order moment alignment network to solve the variable working condition problem. Liu et al. (2024) proposed a multi-source domain unsupervised domain adaptation model based on fusion features, which applied multi-source domains to the mill load parameter prediction model to improve the generalization and prediction accuracy of the model. Although the mill load prediction model based on domain adaptation has achieved initial results, the comprehensive solution involving both multi-task models and variable working condition adaptability requirements, that is, the construction of variable working condition multi-task models, still needs further exploration and research.

In response to the above problems, this paper proposes a multi-task mill load parameter prediction model based on physical information and domain adaptation (PIDAMT). The innovations and contributions of this study are as follows:

(1) A hybrid model integrating physical information and data-driven model is proposed to improve the prediction accuracy of mill load parameters.

(2) In view of the low efficiency and repeated training of traditional single-task models when predicting multiple mill load parameters, a multi-task mill load parameter prediction model is designed.

(3) Considering the impact of variable working conditions on the prediction accuracy of multi-task models, a multi-task model based on domain adaptation is constructed to solve the problem of reduced model prediction accuracy due to differences in data distribution.

2. Physical information

2.1. Grinding process

Fig. 1 shows a schematic diagram of the grinding process. The input of the ball mill includes raw ore, grinding media and water. The raw ore is sent to the ball mill through the ore bin and conveyor belt, and is ground together with steel balls and water. The steel balls grind the ore through collision, friction and extrusion to form slurry. The slurry enters the pump pool through the spiral classifier, is mixed with new water, and is pumped into the hydrocyclone to separate fine-grained overflow and coarse-grained sediment. The sediment is returned to the ball mill for re-grinding, and the overflow is further processed to become a grinding product.

As shown in Fig. 1, the ball load, water load, material density, and steel ball density at the mill inlet can be measured by installing detection instruments. The mill load parameters are related to material density, steel ball density, and other parameters. Therefore, we can use sensors to measure relevant information and construct an approximate physical relationship between mill load parameters.

2.2. Physical relationship between mill load parameters

There is a correlation between the mill load parameters PD and MBVR, which together reflect the operating status of the ball mill. Pulp density (PD) refers to the percentage of the material load in the ball mill to the sum of the material and water loads (Tang et al., 2010). The material changes dynamically during the grinding process and cannot be directly measured (Góralczyk et al., 2020). However, the material load is directly related to the volume and density. Therefore, the equation for PD can be rewritten as Eq. (1):

$$Y_{PD} = \frac{V_m \cdot \rho_m}{V_m \cdot \rho_m + L_w} \tag{1}$$

where Y_{PD} represents the value of PD, L_w represents the water load, V_m represents the volume of the material, and ρ_m represents the density of the material.

The material-to-ball volume ratio (MBVR) refers to the ratio between the material volume and the pore volume of the steel ball (Tang et al., 2010). The steel ball pore volume refers to the volume of space between steel balls and between steel balls and the inner wall of the ball mill that is not occupied by steel balls. This variable cannot be measured directly, but is directly related to the porosity of the medium and the volume of the steel ball. The volume of the steel balls can be approximated by the density of the steel balls and the load of the steel balls. Therefore, the equation for MBVR can be written as Eq. (2):

$$Y_{MBVR} = \frac{V_m \cdot (1-\mu) \cdot \rho_b}{\mu \cdot L_b} \tag{2}$$

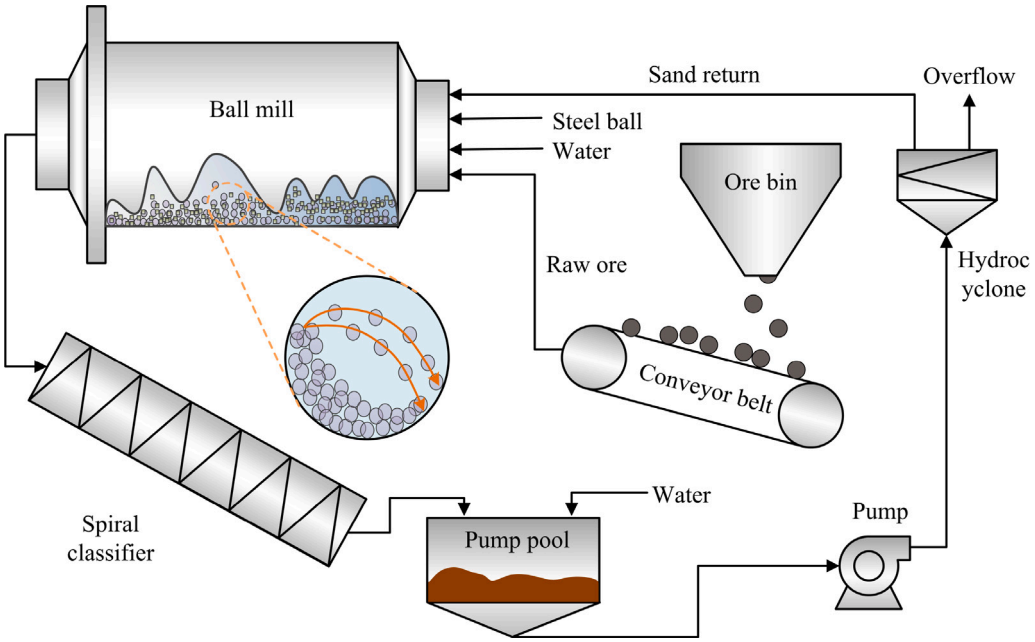


Fig. 1. Schematic diagram of grinding process.

where Y_{MBVR} represents the value of MBVR, μ is the porosity of the medium, L_b represents the load of the steel ball, and ρ_b is the density of the steel ball.

Combining Eq. (1) and Eq. (2), we can get the relationship between PD and MBVR, as shown in Eq. (3):

$$Y_{PD} = \frac{Y_{MBVR} \cdot \mu \cdot L_b \cdot \rho_m}{Y_{MBVR} \cdot \mu \cdot L_b \cdot \rho_m + L_w \cdot (1-\mu) \cdot \rho_b} \tag{3}$$

3. Physics-data-driven hybrid model

3.1. PIDAMT

The source domain data is defined as $X^s = \{x_i^s\} \in \mathrm{R}^{n^s \times m}$, where n^s represents the number of source domain data samples and m is the feature dimension. The number of tasks is defined as N, and the source domain data has a continuous label $Y_j^s = \left\{ y_{i,j}^s \right\} \in \mathrm{R}^{n^s \times 1}, j = 1, \ldots, N$. The target domain data is defined as $X^t = \{x_i^t\} \in \mathrm{R}^{n^t \times m}$, and n^t represents the number of samples to be predicted in the target domain.

In order to accurately predict the mill load parameters, a multi-task mill load parameter prediction model based on physical information and domain adaptation (PIDAMT) is proposed. The structure of the proposed multi-task model includes three parts: feature extractor, gate network and tower network. The loss function includes prediction loss, domain adaptation loss and physical loss. The PIDAMT framework diagram is shown in Fig. 2.

3.2. Multi-task data-driven model

The framework diagram of the multi-task data-driven model proposed in this paper is similar to CGC, which includes three parts: feature extraction block, gating network and tower network. However, in order to simplify the model structure and reduce the number of model parameters, this paper does not add an expert system to the feature extraction block.

The feature extractor of the multi-task data-driven model consists of two parts: common feature extraction and special feature extraction. Common feature extraction is responsible for capturing common information between multiple tasks, which is crucial to improving the generalization ability of the model. First, the data is input into the common feature extractor to extract common features of different tasks. The common feature extractor consists of two fully connected layers. Define the common feature extractor as $F_c(\cdot)$, then the common features of the source domain and target domain data can be expressed as Eq. (4):

$$\begin{cases} Z_c^s = F_c(X^s) \\ Z_c^t = F_c(X^t) \end{cases} \tag{4}$$

where Z_c^s represents the common features of the source domain data, and Z_c^t represents the common features of the target domain data.

However, there are often differences between tasks in actual situations, which requires the model to mine special features unique to each task while extracting common features. The task-specific special feature extraction block can achieve this goal. Therefore, the data is simultaneously input into the task-specific special feature extractor to mine a more refined feature representation unique to each task. The network structure of the special feature extractor is also two layers of fully connected layers, so the special features of the source domain and target domain data can be expressed as Eq. (5):

$$\begin{cases} Z_p^s = \left\{ Z_{p,1}^s, \ldots, Z_{p,j}^s, \ldots, Z_{p,N}^s \right\} \\ \quad = \left\{ F_{p,1}(X^s), \ldots, F_{p,j}(X^s), \ldots, F_{p,N}(X^s) \right\} \\ Z_p^t = \left\{ Z_{p,1}^t, \ldots, Z_{p,j}^t, \ldots, Z_{p,N}^t \right\} \\ \quad = \left\{ F_{p,1}(X^t), \ldots, F_{p,j}(X^t), \ldots, F_{p,N}(X^t) \right\} \end{cases} \tag{5}$$

where Z_p^s and Z_p^t represent the special features of the source domain and the target domain respectively, $Z_{p,j}^s$ and $Z_{p,j}^t$ represent the special features of the jth task of the source domain and target domain data respectively. represents the special feature extractor of the jth task.

Given that common features and special features contribute differently to the task, we introduce a gating network (Jacobs et al., 1991). The main task of the gating network is to further process and filter the extracted feature vectors, controlling which information should be retained and passed to the tower network, and which information should be suppressed or ignored. The gating network acts as an information filter, and the network includes a fully connected layer and a Softmax layer.

In the gated network, the output of the Softmax layer is used as a "gate" switch to control the degree of information passing. Specifically,

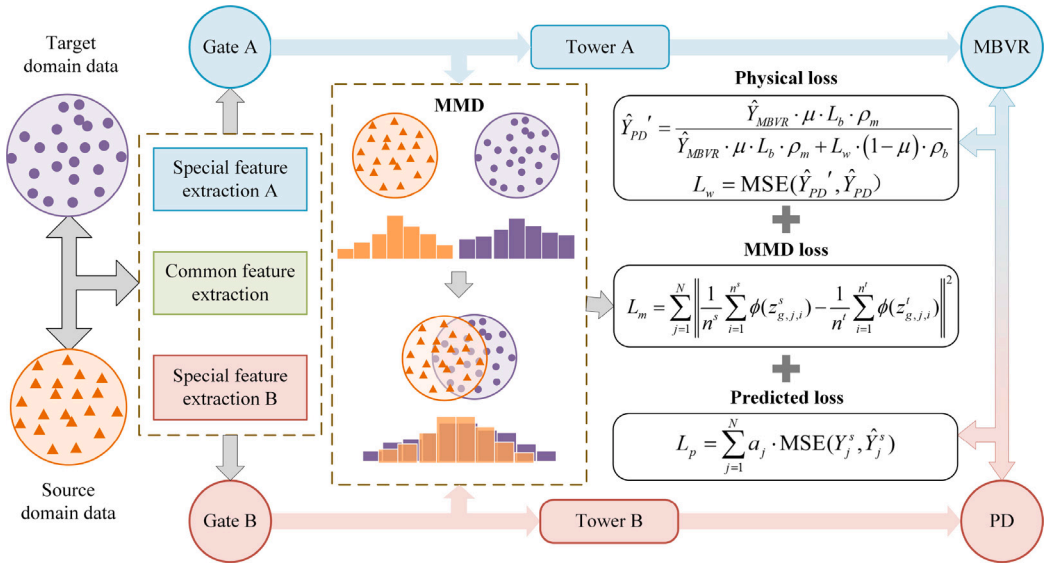


Fig. 2. PIDAMT framework diagram.

when the Softmax layer assigns a higher weight to a feature, the information of the feature will be retained to a greater extent. On the contrary, if the Softmax output value of the feature is low, the information will be appropriately suppressed during the transmission process. In this way, the gated network can effectively filter out key information and pass it to the subsequent parts of the model for further processing. The output of the gated network can be defined as Eq. (6):

$$\begin{cases} Z_g^s = \left\{ Z_{g,1}^s, \ldots, Z_{g,j}^s, \ldots, Z_{g,N}^s \right\} \\ \quad = \left\{ G_1(X^s) \otimes \left(Z_c^s, Z_{p,1}^s \right), \ldots, G_j(X^s) \otimes \left(Z_c^s, Z_{p,j}^s \right), \ldots, \right. \\ \quad \left. G_1(X^s) \otimes \left(Z_c^s, Z_{p,j}^s \right) \right\} \\ Z_g^t = \left\{ Z_{g,1}^t, \ldots, Z_{g,j}^t, \ldots, Z_{g,N}^t \right\} \\ \quad = \left\{ G_1(X^t) \otimes \left(Z_c^t, Z_{p,1}^t \right), \ldots, G_j(X^t) \otimes \left(Z_c^t, Z_{p,j}^t \right), \ldots, \right. \\ \quad \left. G_1(X^t) \otimes \left(Z_c^t, Z_{p,j}^t \right) \right\} \end{cases} \tag{6}$$

where Z_g^s and Z_g^t represent the fusion features of the common features and special features of the source domain and target domain respectively. $G_j(\cdot)$ represents the gating network of the jth task. $Z_{g,j}^s$ and $Z_{g,j}^t$ represent the gating outputs of the source domain and target domain data of the jth task, respectively. $\otimes$ represents matrix dot multiplication.

The final task prediction is implemented by task-specific tower networks. The main function of the tower network is to achieve accurate predictions for each task by integrating different feature information based on the features extracted and screened by the above modules. To achieve this goal, the tower network structure adopts three fully connected layers. In addition, to further improve the prediction performance, the tower network also integrates dropout and ReLU activation function layers. The dropout layer randomly discards the output of a part of the neurons during the training process, which can prevent the network from overfitting. The ReLU activation function enhances the network's expression ability through its nonlinear characteristics, enabling the model to capture the nonlinear relationship of the input data and improve the accuracy and generalization of the prediction.

Define the tower network as $T_j(\cdot)$, then the calculation equation of the tower network can be expressed as Eq. (7):

$$\begin{cases} \hat{Y}^s = \left\{ \hat{Y}_1^s, \ldots, \hat{Y}_j^s, \ldots, \hat{Y}_N^s \right\} \\ \quad = \left\{ T_1(Z_{g,1}^s), \ldots, T_j(Z_{g,j}^s), \ldots, T_N(Z_{g,N}^s) \right\} \\ \hat{Y}^t = \left\{ \hat{Y}_1^t, \ldots, \hat{Y}_j^t, \ldots, \hat{Y}_N^t \right\} \\ \quad = \left\{ T_1(Z_{g,1}^t), \ldots, T_j(Z_{g,j}^t), \ldots, T_N(Z_{g,N}^t) \right\} \end{cases} \tag{7}$$

where $\hat{Y}^s$ and $\hat{Y}^t$ represent the prediction results of the source domain and target domain respectively. $\hat{Y}_j^s$ and $\hat{Y}_j^t$ represent the prediction results of the jth task of the source domain and target domain data respectively.

3.3. Loss function

In order to improve the prediction effect of the model, three loss functions are constructed: prediction loss, domain adaptation loss, and physical loss. The overall training process of the model is shown in Fig. 2. The purpose of prediction loss is to quantify the difference between the model prediction value and the actual observation value, so as to guide the model to continuously optimize during the training process, thereby improving the accuracy of model prediction. In the mill load parameter prediction task, mean squared error (MSE) is often used as prediction loss. The calculation equation of prediction loss is shown in Eq. (8):

$$\begin{aligned} L_p &= \sum_{j=1}^{N} a_j \cdot \mathrm{MSE}(Y_j^s, \hat{Y}_j^s) = a_1 \cdot \frac{1}{n^s} \sum_{i=1}^{n^s} (y_{1,i}^s - \hat{y}_{1,i}^s) + \cdots \\ &+ a_j \cdot \frac{1}{n^s} \sum_{i=1}^{n^s} (y_{j,i}^s - \hat{y}_{j,i}^s) + \cdots + a_N \cdot \frac{1}{n^s} \sum_{i=1}^{n^s} (y_{N,i}^s - \hat{y}_{N,i}^s) \end{aligned} \tag{8}$$

where L_p represents the prediction loss and a_j is the weight coefficient.

The purpose of domain adaptation loss is to solve the problem of model performance degradation and reduced prediction accuracy caused by data distribution differences. To achieve this goal, domain adaptation loss usually adopts various technical means, such as adversarial training and distance metric. Adversarial training introduces a domain discriminator to distinguish the data of the source domain and the target domain, and at the same time trains the model to deceive the domain discriminator so that the features learned by the model cannot

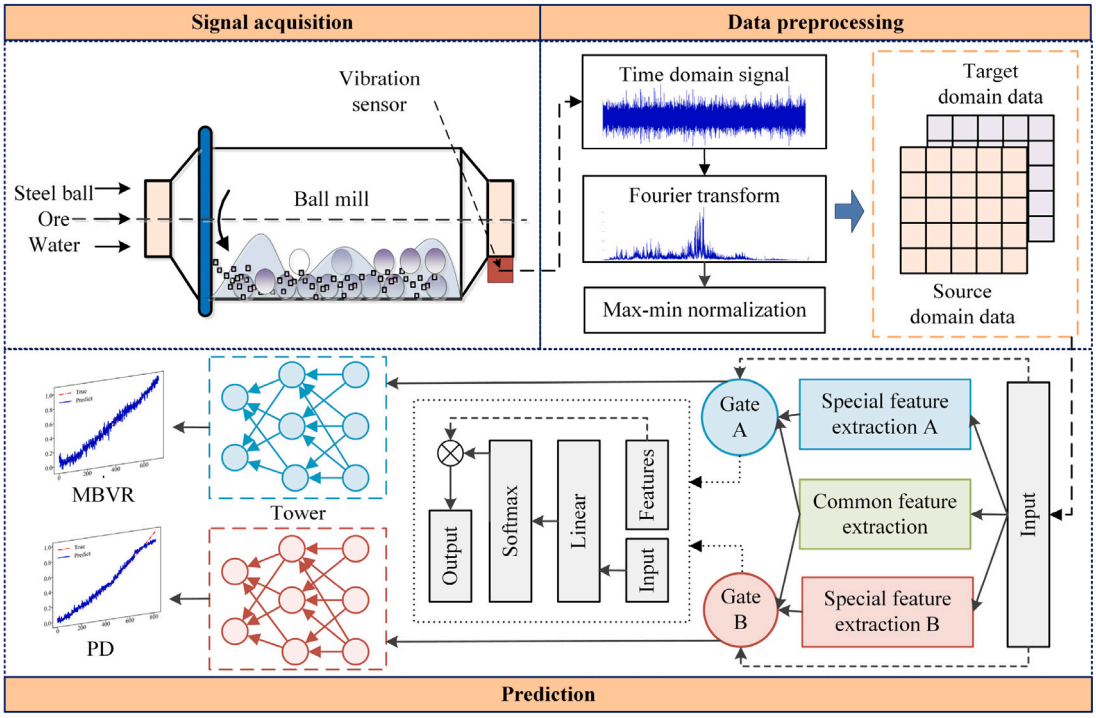


Fig. 3. Model framework diagram.

be distinguished by the domain discriminator, thereby achieving cross-domain feature alignment. The distance metric calculates the distance between the source domain and the target domain and optimizes it as part of the loss function to reduce the difference between the domains. Since adversarial training is often used for classification tasks, this paper adopts the maximum mean discrepancy (MMD) (Lin et al., 2018), a commonly used distance metric, to achieve domain adaptation. The calculation method of MMD is shown in Eq. (9):

$$
\begin{aligned}
L_m &= \sum_{j=1}^{N} \left\| \frac{1}{n^s} \sum_{i=1}^{n^s} \phi(z^s_{g,j,i}) - \frac{1}{n^t} \sum_{i=1}^{n^t} \phi(z^t_{g,j,i}) \right\|^2 \\
&= \sum_{j=1}^{N} \left[\frac{1}{(n^s)^2} \sum_{i=1}^{n^s} \sum_{i=1}^{n^s} \phi(z^s_{g,j,i})^{\mathrm{T}} \phi(z^s_{g,j,i}) - \frac{1}{n^s \cdot n^t} \sum_{i=1}^{n^s} \sum_{i=1}^{n^t} \phi(z^s_{g,j,i})^{\mathrm{T}} \phi(z^t_{g,j,i}) \right. \\
&\left. + \frac{1}{(n^t)^2} \sum_{i=1}^{n^t} \sum_{i=1}^{n^t} \phi(z^t_{g,j,i})^{\mathrm{T}} \phi(z^t_{g,j,i}) \right] \\
&= \sum_{j=1}^{N} \left[\frac{1}{(n^s)^2} \sum_{i=1}^{n^s} \sum_{i=1}^{n^s} k(z^s_{g,j,i}, z^s_{g,j,i}) - \frac{1}{n^s \cdot n^t} \sum_{i=1}^{n^s} \sum_{i=1}^{n^t} k(z^s_{g,j,i}, z^t_{g,j,i}) \right. \\
&\left. + \frac{1}{(n^t)^2} \sum_{i=1}^{n^t} \sum_{i=1}^{n^t} k(z^t_{g,j,i}, z^t_{g,j,i}) \right]
\end{aligned}
\tag{9}
$$

where L_m represents domain adaptation loss, ϕ represents feature map, and $\left\{z^s_{g,j,i}\right\} \in Z^s_{g,j}$, $\left\{z^t_{g,j,i}\right\} \in Z^t_{g,j}$, and $k(\cdot,\cdot)$ represent kernel functions.

In addition to the above losses, we also introduced physical losses to further constrain and optimize the model. The physical loss is constructed based on the physical laws and prior knowledge of the mill operation process, and takes into account the physical relationship between the mill load parameters. By incorporating physical laws into the loss function, we can guide the model to follow these physical principles during the prediction process, thereby improving the rationality and accuracy of the prediction results.

The physical relationship between MBVR and PD is shown in Eq. (3). According to Eq. (7), the prediction results of MBVR and PD can be obtained. Substituting the prediction results of MBVR into Eq. (3), the prediction value of PD calculated according to the physical Equation is obtained. The calculation Equation is shown in Eq. (10):

$$
\hat{Y}'_{PD} = \frac{\hat{Y}_{MBVR} \cdot \mu \cdot L_b \cdot \rho_m}{\hat{Y}_{MBVR} \cdot \mu \cdot L_b \cdot \rho_m + L_w \cdot (1-\mu) \cdot \rho_b} \tag{10}
$$

where $\hat{Y}_{MBVR}$ represents the MBVR prediction value obtained according to the data-driven model, and $\hat{Y}'_{PD}$ represents the PD prediction value calculated according to the physical equation.

The constructed physical loss is shown in Eq. (11):

$$
\begin{aligned}
L_w &= \mathrm{MSE}(\hat{Y}_{PD}{}', \hat{Y}_{PD}) \\
&= \frac{1}{n^s} \sum_{i=1}^{n^s} \left(\frac{\hat{y}_{MBVR,i} \cdot \mu \cdot L_b \cdot \rho_m}{\hat{y}_{MBVR,i} \cdot \mu \cdot L_b \cdot \rho_m + L_w \cdot (1-\mu) \cdot \rho_b} - \hat{y}_{PD,i} \right)
\end{aligned}
\tag{11}
$$

where L_w is the physical loss, $\hat{Y}_{PD} = \left\{\hat{y}_{PD,i}\right\}$ is the predicted value of slurry concentration obtained according to the data-driven model, and $\left\{\hat{y}_{MBVR,i}\right\} \in \hat{Y}_{MBVR}$ is.

According to the above content, the total loss function of the model can be obtained as shown in Eq. (12):

$$
L = \min \left\{ L_p + bL_w + cL_m \right\} \tag{12}
$$

where L represents the total loss function of the model, b and c are weight coefficients.

3.4. Hybrid model training process

The overall training process of the hybrid model is shown in Fig. 3, which includes three parts: signal acquisition, data processing, and model prediction.

(1) Signal acquisition: The vibration signals of the laboratory ball mill bearings under different working conditions are collected, and the sampling rate of the signal is 51.2 kHz.

(2) Data processing: Perform fast fourier transform on the collected vibration signals to obtain the representation of the vibration signals in the frequency domain. After that, perform maximum and minimum

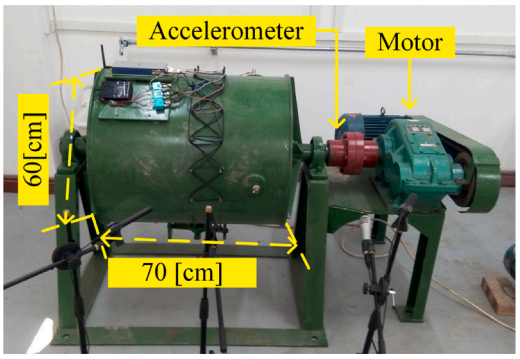


Fig. 4. Laboratory small ball mill.

normalization on the data under different working conditions to obtain the source domain data and target domain data.

(3) Prediction: Use the source domain data to train the model. Input the data to be predicted into the trained model to obtain the prediction results of different tasks.

The specific training steps of the model are as follows:

Step 1: Input the processed source domain data $X^s = \{x_i^s\}$ and target domain data $X^t = \{x_i^t\}$ into Eq. (4) to obtain common features of different tasks.

Step 2: At the same time, input source domain data $X^s = \{x_i^s\}$ and target domain data $X^t = \{x_i^t\}$ into Eq. (5) to obtain special features of different tasks.

Step 3: Input source domain data, target domain data, common features, and special features into Eq. (6) to calculate the output of the gating network.

Step 4: Input the output of the gating network into Eq. (7) to obtain the predicted values of the data-driven models MBVR and PD.

Step 5: Input the label value of the source domain data and the predicted value of the mill load parameter obtained from Step 4 into Eq. (8) to calculate the prediction loss.

Step 6: Input the output of the source domain and target domain data gating network into Eq. (9) to calculate the domain adaptation loss of the model.

Step 7: Input the predicted values of steel ball load L_b, steel ball density ρ_b, material density ρ_m, water load L_w and MBVR of the working condition of the source domain data into Eq. (10) to obtain the PD predicted value calculated according to the physical equation.

Step 8: Input the PD prediction value obtained by the physical equation and the data-driven model into Eq. (11) to obtain the physical loss.

Step 9: Input the loss and weight coefficients a_i, b and c obtained from Step 5, Step 6 and Step 8 into Eq. (12) to obtain the total loss function of the model. Through back propagation, the trained model is obtained.

Step 10: Input the target domain data $X^t = \{x_i^t\}$ into the trained model to obtain the predicted value of the mill load parameter of the target domain data.

4. Experimental description

4.1. Dataset introduction

The data set used in the experiment is the data collected from a small ball mill in the laboratory. The volume of the mill is 200 L, the diameter is 60 cm, and the length is 70 cm. The rotation speed of the mill cylinder is 42 r/min. The experimental material is iron ore, and the grinding medium is steel balls with a diameter of about 30 mm. The small ball mill in the laboratory is shown in Fig. 4.

Table 1
Experimental parameters.

Dataset	Ball charge volume ratio	L_b	L_w	Number of training sets	Number of test sets
1	0.3	292	35	2780	1112
2	0.35	340.69	40	2060	824
3	0.4	389.36	40	1760	704
4	0.45	438.03	35	1900	760
5	0.5	486.7	40	2040	816

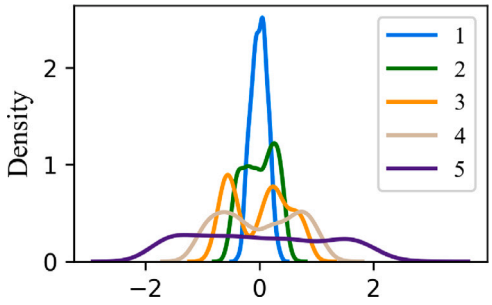


Fig. 5. Dataset distribution.

The experiment collected data from five working conditions. In actual industry, the ball charge volume ratio (BCVR) of the mill fluctuates between 0.3 and 0.5. Therefore, when designing the experiment, a fixed medium filling rate, that is, a certain steel ball mass, was used to change the load parameters such as MBVR and PD inside the mill by continuously adding materials. The experimental parameters of the five working conditions are shown in Table 1.

The probability density curves of the five working conditions are plotted as shown in Fig. 5. Among them, the five data sets are defined as 1–5, the ordinate represents the probability density, and the abscissa represents the value of the sample after dimensionality reduction. The probability density curve is used to describe the data distribution. As can be seen from Fig. 5, the distribution of the five sets of data collected is inconsistent.

4.2. Model parameters and hyperparameters

This section details the network structure, specific parameters, and hyperparameters set during network training. The network structure and specific parameters are shown in Table 2. The structures of the common feature extractor and the special feature extractor are both three-layer fully connected layers. In order to make the model nonlinear, the ReLU activation function is introduced. Dropout is added to prevent the model from overfitting. Since the network structure and input of the common feature extractor and the special feature extractor are consistent, the network structure and parameters of different feature extractors are uniformly represented by feature extractors in Table 2. Similarly, the parameters and input dimensions of the gated network and tower network for different tasks are consistent, so the gated and tower network structures and parameters are uniformly represented. The gated network learns the weights between input features through the fully connected layer, and the Softmax layer is designed to convert the output of the fully connected layer into a probability value. In this way, each input feature or feature combination has a corresponding probability value, indicating the degree to which the feature should be retained. The structure of the tower network is also three-layer fully connected layers, two layers of ReLU activation function layers and two layers of Dropout layers, which are used to enhance the nonlinear ability of the model while preventing the model from overfitting.

Table 2
Network structure and parameters.

Structure	Basic block	Input size	Output size
Feature Extractor	Linear	(1,512)	(1,128)
	ReLU	(1, 128)	(1,128)
	Dropout	(1, 128)	(1,128)
	Linear	(1, 128)	(1,64)
	ReLU	(1, 64)	(1,64)
	Dropout	(1, 64)	(1,64)
Gated Network	Linear	(1,512)	(1,2)
	Softmax	(1, 2)	(1,2)
Tower Network	Linear	64	64
	ReLU	64	64
	Dropout	64	64
	Linear	64	32
	ReLU	32	32
	Dropout	32	32
	Linear	32	1

Table 3
Network hyperparameters.

Parameter	Value	Parameter	Value
Batch size	64	lr	0.01
epoch	100	iteration	100
a_1	4	a_2	6
b	0.8	c	1
optimizer	Adam	device	Cpu

The model is implemented using Python in the Pytorch deep learning framework. Intel(R) Core(TM) i5-7200U CPU is used to train and evaluate the models. The medium porosity parameter in the physical information is set to a general value of 0.38 in the actual industrial process (Tang et al., 2018b). The hyperparameters in the data-driven model are optimized using the grid search method, and the obtained hyperparameter values are shown in Table 3.

5. Results and discussion

5.1. Comparative experiment

In order to verify the effectiveness of the multi-task method proposed in this paper, the experimental results of ST and three commonly used multi-task methods are compared. The compared multi-task methods are HPS, MMoE and CGC. The number of experts of MMoE and CGC is set to 2. The network parameters are consistent with Table 2.

The experimental results of different methods for the MBVR are shown in Table 4, and the prediction results for the slurry concentration are shown in Table 5. The evaluation indicators used are the root mean square error (RMSE) and the R-square (R^2). The experimental results show that the experimental results of the multi-task method proposed in this paper are the best among the five methods. In order to more clearly compare the experimental results of different methods, we use box plots and bar graphs to plot the R^2 values and RMSE values in Tables 4 and 5. Fig. 6(a) and (b) are box plots of the R^2 values of the prediction results of different methods for the MBVR and PD. The closer the R^2 value is to 1, the higher the fit of the model. Fig. 6(c) and (d) are bar graphs of the RMSE values of the prediction results of different methods for the MBVR and PD.

As can be seen from Fig. 6(a) and Fig. 6(b), when performing multi-task prediction, the R^2 value of PIDAMT is closer to 1, indicating that PIDAMT can predict data more accurately. In contrast, the R^2 values of the other four models (ST, HPS, MMoE and CGC) are relatively low and the data distribution is more dispersed. This shows that the model is more unstable and it is difficult to ensure good prediction performance in different experiments. Among the five models, ST has the smallest median in the prediction results of the two tasks, indicating that the overall prediction performance of the model on this task is poor. The prediction accuracy of HPS is slightly improved compared with that of the ST model, indicating that the multi-task learning model can effectively enhance the prediction ability of the model by sharing information between tasks. MMoE adds a gating network compared to the HPS model. By comparing the experimental results of the two, it can be found that MMoE performs better in the MBVR prediction task, but the model performance decreases slightly in the PD prediction task. However, after adding a special feature extraction block to the model, that is, the CGC model, the model accuracy is improved and the prediction results are more stable. However, combining the experimental results of the two tasks, it can be found that PIDAMT shows good prediction performance in both tasks, proving the effectiveness of the multi-task method proposed in this paper.

Fig. 6(c) and Fig. 6(d) show the RMSE values of the five methods when predicting MBVR and PD. RMSE is an indicator to measure the prediction accuracy of the model. The smaller the value, the better the prediction performance of the model. Combining Fig. 6(c) and (d), it can be seen that the RMSE value of the ST model on the two tasks is the highest, indicating that the multi-task model has higher prediction accuracy than the single-task model. The RMSE values of HPS, MMoE and CGC are between the two, but show a trend of gradually decreasing. This change shows that in the multi-task model, by introducing the gating network (such as MMoE) and the special feature extraction block (such as CGC), the prediction performance of the model can be effectively improved. By analyzing the RMSE values of different methods predicting the two tasks, it can be found that the PIDAMT model has the lowest RMSE value, indicating that the model has high prediction accuracy on both tasks, proving the effectiveness of the method proposed in this paper.

In order to more clearly compare the experimental results of different methods, the experimental results of dataset 1 predicting dataset 5 are visualized. Curve graphs between model prediction values and true values under different methods are drawn to evaluate the performance of each model. Fig. 7 is a curve graph of MBVR predicted by dataset 1 for dataset 5. Fig. 8 is a curve graph of PD predicted by dataset 1 for dataset 5. In Figs. 7 and 8, sub-figures (a)–(e) are the prediction result curve graphs of ST, HPS, MMoE, CGC and PIDAMT, respectively. By analyzing these charts, it can be observed that the prediction effect of the method PIDAMT proposed in this paper is the best.

5.2. The impact of physical information and domain adaptation on model performance

In order to explore the influence of physical information and domain adaptation methods on the prediction effect of mill load parameters, we also conducted an ablation experiment on PIDAMT. Different versions of PIDAMT are shown in Table 6.

The experimental results of the ablation experiment are shown in Tables 7 and 8. Table 7 shows the R^2 values of different tasks predicted by the ablation experiment, and Table 8 shows the RMSE values of different tasks predicted by the ablation experiment. The bold in the table indicates the best result.

PIMT is a hybrid model without domain adaptation, and DAMT is a multi-task model without physical information. When physical information is incorporated into the multi-task model, that is, when it is transformed from DAMT to PIDAMT, the prediction accuracy of the model is improved. This result shows that the introduction of physical information can improve the accuracy of model prediction. By comparing PIMT and PIDAMT, that is, introducing domain adaptation methods in the model, the prediction accuracy of the model under variable operating conditions is improved. In most cases, the prediction accuracy of the model will be greatly improved after adding domain adaptation methods.

In order to further compare the comprehensive performance of the model under different conditions, we visualized Tables 7 and 8, and the visualization results are shown in Fig. 9. Fig. 9(a) uses a

Table 4
MBVR prediction results.

Source domain	Target domain	ST		HPS		MMoE		CGC		PIDAMT	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
1	2	0.930	0.078	0.963	0.057	0.957	0.061	0.952	0.064	0.965	0.055
	3	0.825	0.124	0.895	0.096	0.916	0.086	0.920	0.083	0.942	0.071
	4	0.788	0.135	0.919	0.084	0.953	0.063	0.943	0.070	0.959	0.060
	5	0.916	0.086	0.910	0.089	0.926	0.080	0.921	0.083	0.952	0.065
2	1	0.873	0.105	0.819	0.125	0.764	0.143	0.947	0.067	0.950	0.066
	3	0.955	0.063	0.979	0.042	0.971	0.051	0.978	0.044	0.984	0.037
	4	0.962	0.057	0.969	0.052	0.943	0.070	0.971	0.050	0.974	0.047
	5	0.929	0.079	0.940	0.072	0.961	0.058	0.967	0.053	0.986	0.035
3	1	0.755	0.145	0.703	0.160	0.858	0.110	0.842	0.117	0.932	0.076
	2	0.872	0.105	0.903	0.092	0.906	0.090	0.902	0.092	0.980	0.042
	4	0.866	0.107	0.979	0.042	0.979	0.042	0.981	0.041	0.987	0.033
	5	0.936	0.075	0.938	0.074	0.948	0.067	0.923	0.082	0.990	0.030
4	1	0.651	0.173	0.816	0.126	0.846	0.115	0.833	0.120	0.929	0.078
	2	0.851	0.113	0.881	0.102	0.886	0.099	0.877	0.103	0.962	0.058
	3	0.905	0.091	0.918	0.084	0.974	0.048	0.963	0.057	0.976	0.045
	5	0.809	0.129	0.934	0.076	0.976	0.046	0.936	0.075	0.989	0.032
5	1	0.317	0.243	0.507	0.206	0.439	0.220	0.544	0.198	0.922	0.082
	2	0.337	0.240	0.670	0.169	0.545	0.198	0.629	0.179	0.944	0.070
	3	0.739	0.151	0.762	0.144	0.853	0.113	0.803	0.131	0.946	0.069
	4	0.851	0.113	0.900	0.093	0.857	0.111	0.838	0.118	0.959	0.060

Table 5
PD prediction results.

Source domain	Target domain	ST		HPS		MMoE		CGC		PIDAMT	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
1	2	0.886	0.097	0.914	0.084	0.912	0.085	0.914	0.084	0.925	0.079
	3	0.730	0.151	0.739	0.149	0.726	0.152	0.844	0.115	0.920	0.082
	4	0.872	0.102	0.935	0.073	0.947	0.065	0.950	0.064	0.963	0.055
	5	0.906	0.090	0.904	0.091	0.969	0.051	0.961	0.058	0.972	0.049
2	1	0.577	0.188	0.654	0.170	0.700	0.158	0.790	0.132	0.913	0.085
	3	0.895	0.094	0.933	0.075	0.952	0.064	0.925	0.080	0.958	0.059
	4	0.932	0.074	0.969	0.050	0.969	0.050	0.923	0.079	0.974	0.046
	5	0.879	0.102	0.957	0.060	0.961	0.058	0.954	0.063	0.962	0.057
3	1	0.295	0.242	0.483	0.207	0.468	0.210	0.418	0.220	0.875	0.102
	2	0.711	0.154	0.784	0.134	0.835	0.117	0.815	0.124	0.936	0.073
	4	0.726	0.150	0.851	0.110	0.908	0.087	0.816	0.123	0.956	0.060
	5	0.924	0.081	0.930	0.077	0.874	0.104	0.924	0.081	0.944	0.069
4	1	0.693	0.160	0.799	0.129	0.675	0.164	0.759	0.142	0.867	0.105
	2	0.809	0.126	0.896	0.093	0.827	0.119	0.889	0.096	0.901	0.091
	3	0.871	0.104	0.876	0.103	0.853	0.112	0.897	0.094	0.927	0.078
	5	0.914	0.086	0.872	0.104	0.724	0.153	0.925	0.080	0.965	0.054
5	1	0.422	0.219	0.547	0.194	0.624	0.177	0.575	0.188	0.840	0.115
	2	0.655	0.169	0.851	0.111	0.792	0.131	0.824	0.121	0.929	0.077
	3	0.784	0.135	0.899	0.093	0.917	0.084	0.926	0.079	0.937	0.073
	4	0.853	0.110	0.888	0.096	0.923	0.080	0.910	0.086	0.970	0.049

Table 6
Different versions of PIDAMT for ablation studies.

Method	Physical information	Domain adaptation
PIMT	✓	×
DAMT	×	✓
PIDAMT	✓	✓

box plot to visualize the R^2 value of the ablation experiment, and Fig. 9(b) uses a bar chart to visualize the RMSE value of the ablation experiment. It can be seen from the figure that after adding physical information, the prediction accuracy of the model is improved, indicating the effectiveness of building a physical-data-driven hybrid model. After adding the domain adaptation method, the prediction accuracy of the model under variable working conditions is greatly improved, indicating that adding the domain adaptation method can improve the model prediction performance.

5.3. Model performance comparison

To further illustrate the effectiveness of the proposed method, we use a scatter plot to visualize the R^2 value of the experimental results, as shown in Fig. 10. The horizontal axis represents the R^2 value of the MBVR prediction result, and the vertical axis represents the R^2 value of the PD prediction result. The closer the scatter point is to the upper right, the higher the R^2 value of the two task prediction results, and the better the performance. Taking the experimental results of the single-task model as the benchmark, the experimental results can be roughly divided into four cases.

(1) Case 1 (Fig. 10(a), Dataset 1 predicts Dataset 2): The multi-task learning model performs significantly better than the corresponding single-task model in both MBVR and PD prediction tasks.

(2) Case 2 (Fig. 10(b), Dataset 1 predicts Dataset 5): The prediction effect of some multi-task models on both tasks is not as good as that of single-task models, that is, negative transfer problem occurs.

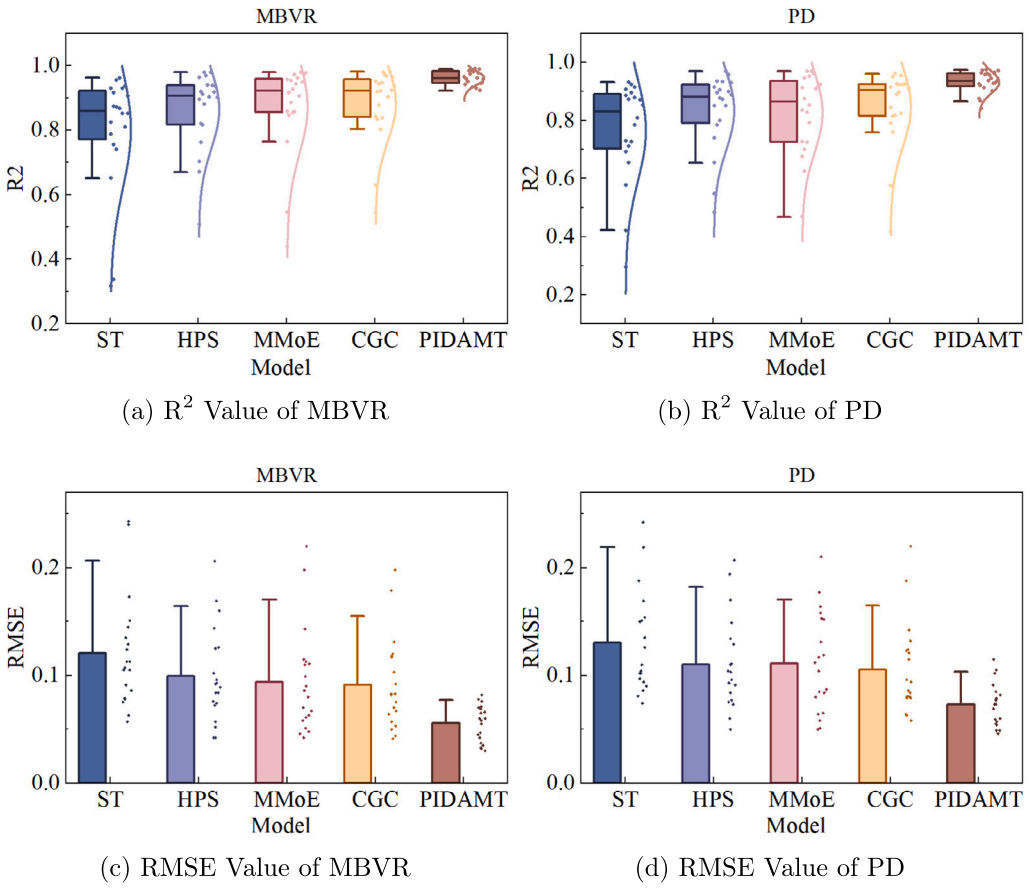


Fig. 6. Visualization of different methods for predicting MBVR and PD experimental indicators.

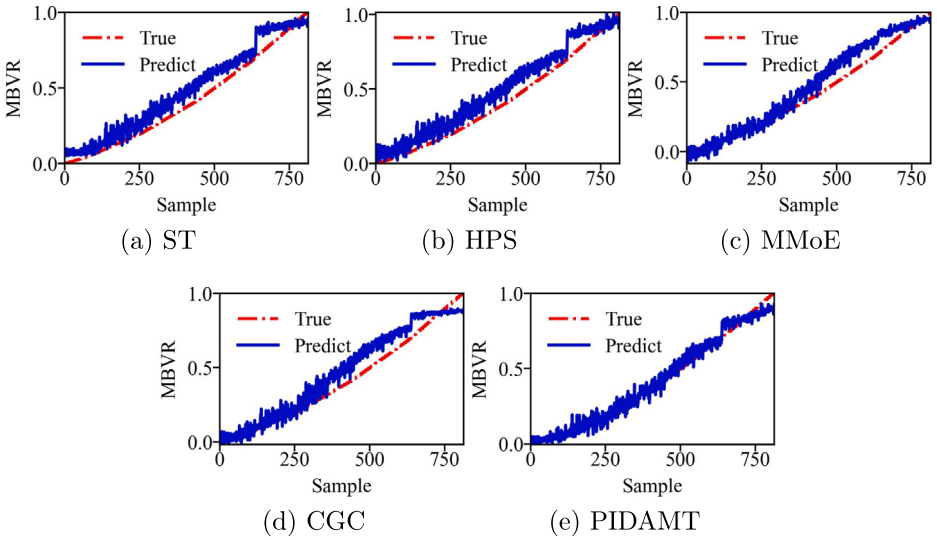


Fig. 7. Experimental results of dataset 1 predicting MBVR of dataset 5.

(3) Case 3 (Fig. 10(c), Dataset 2 predicts Dataset 1): Certain multi-task models show higher accuracy in predicting PD, but the accuracy decreases when predicting MBVR.

(4) Case 4 (Fig. 10(d), Dataset 3 predicts Dataset 5): In contrast to Case 3, the accuracy of some multi-task models in predicting MBVR is improved, while the accuracy in predicting PD is decreased. Cases 3 and 4 are both seesaw phenomena (Tang et al., 2020a).

As can be seen from Fig. 10, the HPS model has unstable predictions in the cross-operating mill load parameter prediction task, which is specifically manifested in the proneness of negative transfer problems and seesaw phenomena. After adding a gated network, that is, the MMoE model, the negative transfer problem can be reduced, but the seesaw phenomenon still exists. The CGC model adds a special feature extractor on the basis of the MMoE model, which can alleviate the seesaw phenomenon to a certain extent, but the accuracy of some tasks will still decrease under variable operating conditions. The model proposed in this paper not only effectively avoids the occurrence of negative migration and seesaw phenomenon by integrating physical

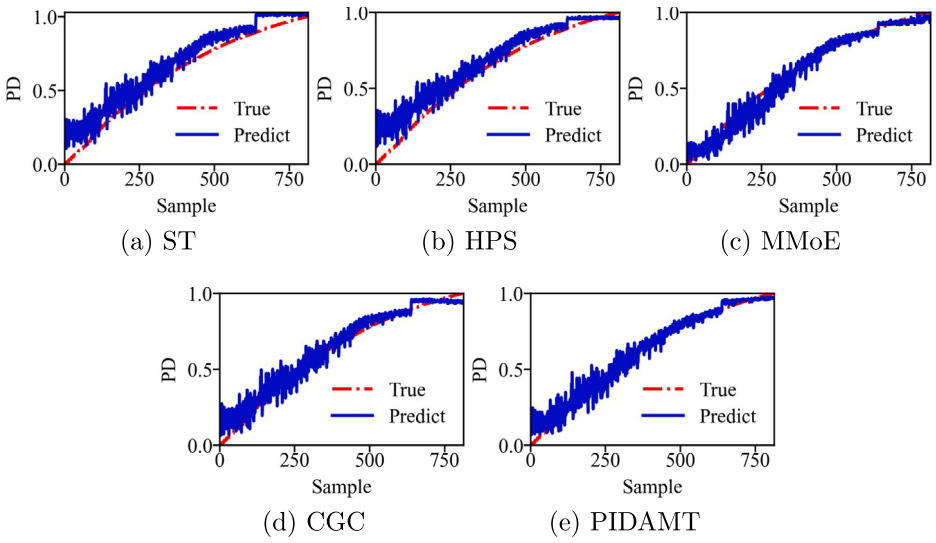


Fig. 8. Experimental results of dataset 1 predicting PD of dataset 5.

Table 7
Ablation experiment R^2 value.

Source domain	Target domain	PIMT		DAMT		PIDAMT	
		MBVR	PD	MBVR	PD	MBVR	PD
1	2	0.953	0.920	0.963	0.922	0.965	0.925
	3	0.939	0.830	0.927	0.876	0.942	0.920
	4	0.916	0.962	0.952	0.945	0.959	0.963
	5	0.941	0.967	0.945	0.958	0.952	0.972
2	1	0.897	0.770	0.943	0.912	0.950	0.913
	3	0.978	0.948	0.955	0.950	0.984	0.958
	4	0.965	0.968	0.957	0.965	0.974	0.974
	5	0.970	0.958	0.946	0.951	0.986	0.962
3	1	0.739	0.458	0.925	0.823	0.932	0.875
	2	0.897	0.816	0.976	0.920	0.980	0.936
	4	0.962	0.854	0.985	0.954	0.987	0.956
	5	0.932	0.910	0.982	0.920	0.990	0.944
4	1	0.775	0.718	0.920	0.838	0.929	0.867
	2	0.901	0.842	0.944	0.892	0.962	0.901
	3	0.942	0.909	0.975	0.916	0.976	0.927
	5	0.855	0.937	0.982	0.858	0.989	0.965
5	1	0.752	0.701	0.785	0.845	0.804	0.861
	2	0.534	0.812	0.928	0.897	0.944	0.929
	3	0.719	0.873	0.936	0.936	0.946	0.937
	4	0.860	0.872	0.940	0.962	0.959	0.970

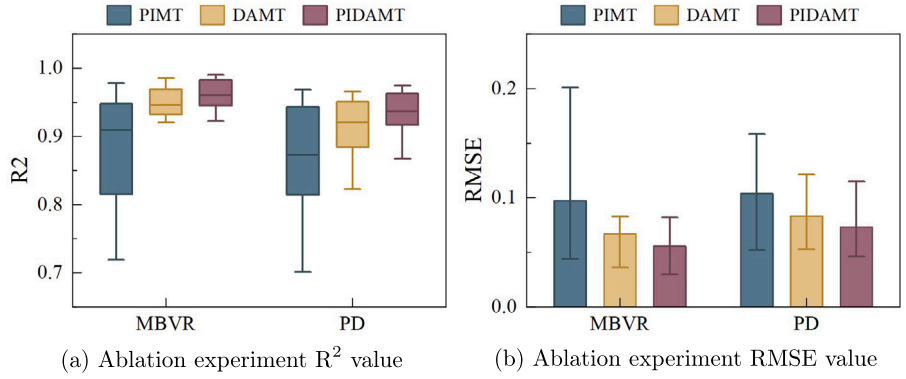


Fig. 9. Visualization of ablation experiment results.

Table 8
Ablation experiment RMSE value.

Source domain	Target domain	PIMT		DAMT		PIDAMT	
		MBVR	PD	MBVR	PD	MBVR	PD
1	2	0.064	0.081	0.057	0.080	0.055	0.079
	3	0.073	0.120	0.080	0.102	0.071	0.082
	4	0.085	0.056	0.065	0.067	0.060	0.055
	5	0.072	0.053	0.070	0.060	0.065	0.049
2	1	0.094	0.138	0.070	0.086	0.066	0.085
	3	0.044	0.066	0.063	0.065	0.037	0.059
	4	0.055	0.052	0.061	0.053	0.047	0.046
	5	0.051	0.060	0.069	0.064	0.035	0.057
3	1	0.150	0.212	0.081	0.121	0.076	0.102
	2	0.094	0.123	0.045	0.082	0.042	0.073
	4	0.058	0.109	0.036	0.061	0.033	0.060
	5	0.077	0.087	0.039	0.082	0.030	0.069
4	1	0.139	0.153	0.083	0.116	0.078	0.105
	2	0.093	0.114	0.070	0.095	0.058	0.091
	3	0.071	0.088	0.047	0.084	0.045	0.078
	5	0.112	0.073	0.040	0.110	0.032	0.054
5	1	0.146	0.158	0.136	0.114	0.130	0.108
	2	0.201	0.125	0.079	0.092	0.070	0.077
	3	0.156	0.104	0.075	0.074	0.069	0.073
	4	0.110	0.102	0.072	0.055	0.060	0.049

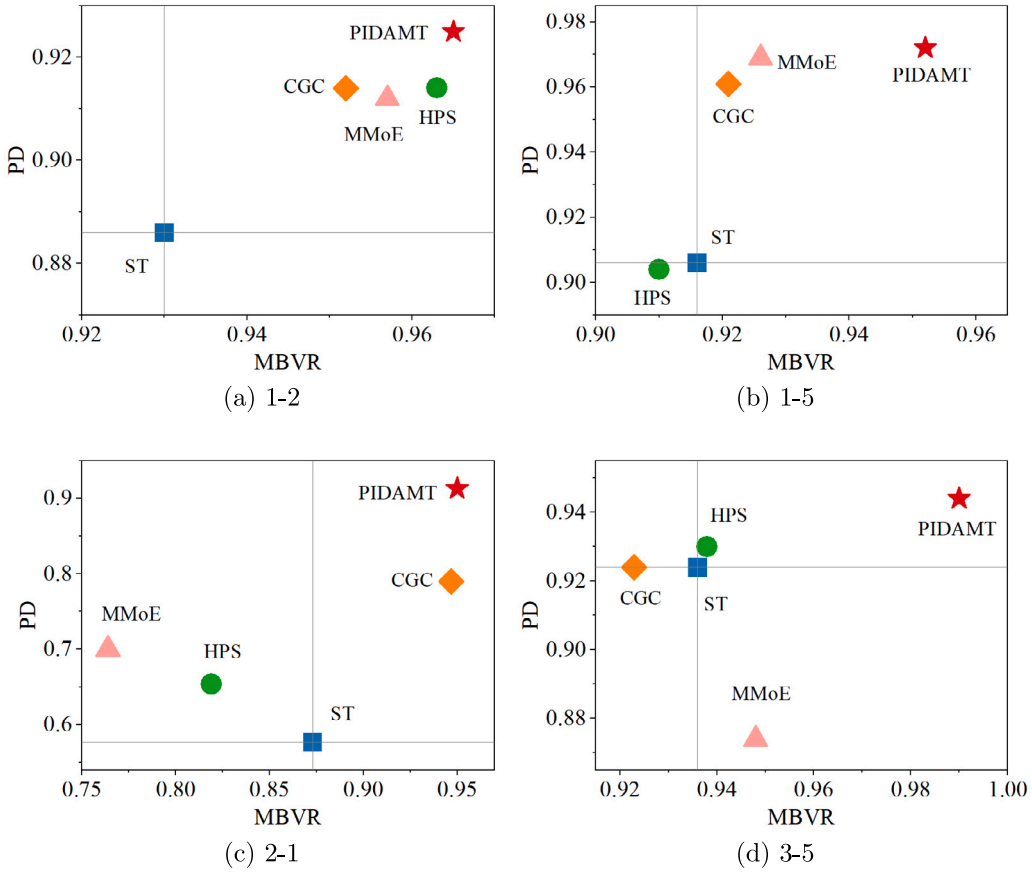
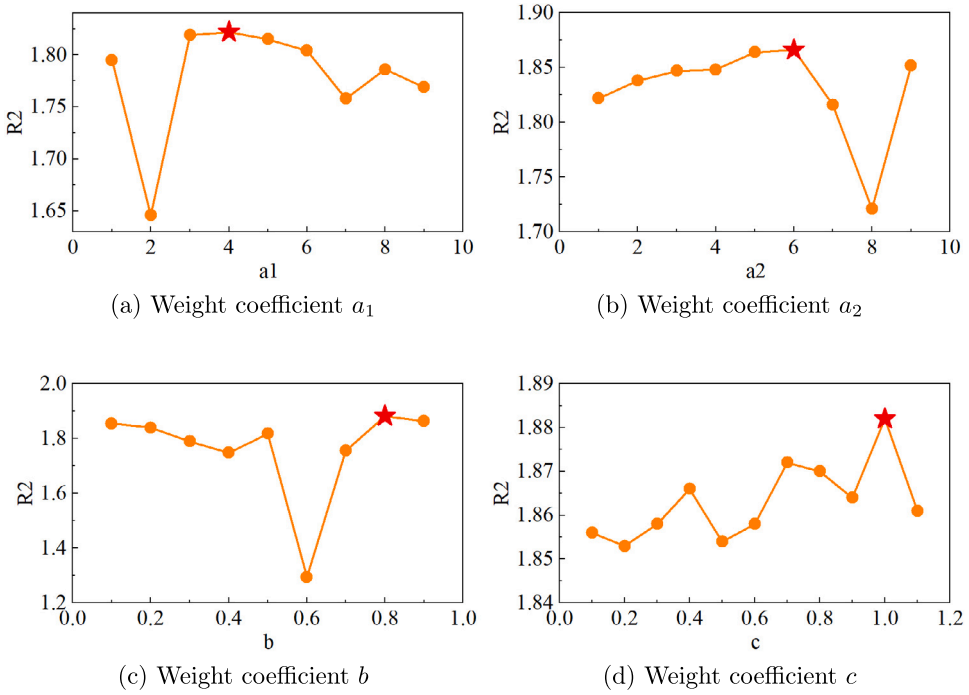


Fig. 10. Task model performance analysis.

information and domain adaptation methods in the multi-task model, but also has a further improvement in prediction accuracy (as shown in Fig. 10, located in the second quadrant and closer to the upper right corner) compared with other models. Through the high-precision mill load prediction results, operators can understand the current MBVR and PD without waiting for measurement results. According to the predicted mill load results, the feed and water supply in the grinding process can be adjusted in time to keep the grinding process in the best operating state, thereby improving grinding efficiency and reducing mill energy consumption.

5.4. Loss weight coefficient performance analysis

The loss function in the model proposed in this paper consists of four parts, namely the prediction loss of the MBVR prediction task, the prediction loss of the PD prediction task, the physical information loss,



(a) Weight coefficient a_1 (b) Weight coefficient a_2

(c) Weight coefficient b (d) Weight coefficient c

Fig. 11. Impact of loss weight on model performances.

and the domain adaptation loss. Different losses contribute differently to the model training process, so different weight coefficients need to be assigned to each loss term. Fig. 11 shows the impact of different weight coefficients on model performance. Since the model includes two tasks, the evaluation index is the sum of the R^2 values of the two tasks. In the figure, the horizontal axis is the value of the weight coefficient, the vertical axis is the sum of the R^2 values of the two tasks, and the red point is the optimal value. The larger the sum of the R^2 values, the better the model performance.

Fig. 11(a) shows the result of the sum of the R^2 values of the two tasks when the MBVR prediction loss weight is [1,9]. It can be clearly seen from the figure that when the value is 4, the sum of the R^2 values is the highest. Fig. 11(b) shows the result of the sum of the R^2 values of the two tasks when the PD prediction loss weight is [1,9]. As the value increases, the value of the evaluation index gradually increases. When the sum of the R^2 values of the two tasks is the highest. As the value continues to increase, the model performance decreases. Fig. 11(c) shows the experimental results when the physical information loss weight is [0.1,0.9]. The experimental results show that when the value is 0.8, the prediction accuracy of the model is the highest. Fig. 11(d) shows the experimental results when the domain adaptation loss weight is [0.1,1.1]. It can be seen from the figure that when is 1, the sum of the R^2 values is the highest.

6. Conclusions

Accurately predicting mill load parameters is of great significance for improving grinding efficiency and reducing mill energy consumption. This paper proposes a multi-task mill load parameter prediction model based on physical information and domain adaptation for the prediction of multiple mill load parameters. By deriving the physical relationship between mill load parameters and embedding it into the data-driven model as a physical constraint, a physical-data-driven hybrid model is constructed. A multi-task learning framework is adopted to achieve simultaneous prediction of multiple mill load parameters, improve prediction efficiency and avoid repeated training of traditional single-task models. In view of the significant distribution difference of ball mill data under different working conditions, the maximum mean discrepancy is introduced to improve the adaptability of the model to variable working conditions. The effectiveness of the proposed model is verified by conducting experiments on laboratory small ball mill data. The experimental results show that compared with the single-task model and the conventional multi-task model, the proposed model shows higher prediction accuracy under cross-working conditions. In addition, in order to clarify the contribution of physical information and domain adaptation methods to the improvement of model performance, this paper conducts ablation experiments. The experimental results show that the addition of physical information and domain adaptation methods improves the prediction accuracy of the model.

CRediT authorship contribution statement

Yiwen Liu: Writing – original draft, Visualization, Methodology. **Gaowei Yan:** Writing – review & editing, Funding acquisition, Data curation. **Shuyi Xiao:** Funding acquisition. **Fang Wang:** Visualization. **Rong Li:** Funding acquisition. **Yusong Pang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61973226), the Shanxi Province Major Special Program of Science and Technology (202201090301013), the Shanxi Province Science Foundation for Youths (202203021222101), and the Shanxi Province Graduate Research Innovation Project (2023KY209).

Data availability

The data that has been used is confidential.

References

Alber, M., Buganza Tepole, A., Cannon, W.R., De, S., Dura-Bernal, S., Garikipati, K., Karniadakis, G., Lytton, W.W., Perdikaris, P., Petzold, L., et al., 2019. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ Digit. Med. 2 (1), 115. http://dx.doi.org/10.1038/s41746-019-0193-y.

Cai, J., Yang, L., Zeng, C., Chen, Y., 2021. Integrated approach for ball mill load forecasting based on improved EWT, refined composite multi-scale dispersion entropy and fireworks algorithm optimized SVM. Adv. Mech. Eng. 13 (2), 1687814021991264, URL https://api.semanticscholar.org/CorpusID:231731040.

Cao, H., Xiao, W., Sun, J., Gan, M.G., Wang, G., 2024. A hybrid data-and model-driven learning framework for remaining useful life prognostics. Eng. Appl. Artif. Intell. 135, 108557. http://dx.doi.org/10.1016/j.engappai.2024.108557.

Curreri, F., Patanè, L., Xibilia, M.G., 2021. Soft sensor transferability: A survey. Appl. Sci. 11 (16), 7710. http://dx.doi.org/10.3390/app11167710.

Dai, W., Chai, T., 2014. Data-driven optimal operational control of complex grinding process. Acta Automat. Sinica 40 (9), 2005–2014, https://api.semanticscholar.org/CorpusID:124268429.

Dai, W., Liu, Q., Chai, T., 2015. Particle size estimate of grinding processes using random vector functional link networks with improved robustness. Neurocomputing 169, 361–372. http://dx.doi.org/10.1016/j.neucom.2014.08.098.

Daniel, M.J., 2016. Energy use in comminution in a global context. In: Sustainability in the Mineral and Energy Sectors. CRC Press, pp. 51–72. http://dx.doi.org/10.1201/9781315369853-5.

Frost Jr, E., Quinn, T., 2018. Data preconditioning for predictive and interpretative algorithms: Importance in data-driven analytics and methods for application. Petrophysics 59 (06), 873–890. http://dx.doi.org/10.30632/PJV59N6-2018a11.

Gao, T., Zhu, H., Wu, J., Lu, Z., Zhang, S., 2024. Hybrid physics data-driven model-based fusion framework for machining tool wear prediction. Int. J. Adv. Manuf. Technol. 132 (3), 1481–1496. http://dx.doi.org/10.1007/s00170-024-13365-6.

Gao-Wei, Y., Min, H., Jian, T., Dong-Sheng, H., 2018. Soft sensor of wet ball mill load based on maximum mean discrepancy multi-source domain transfer learning. Control Decis. 33 (10), 1795–1800. http://dx.doi.org/10.13195/j.kzyjc.2017.0636.

Góralczyk, M., Krot, P., Zimroz, R., Ogonowski, S., 2020. Increasing energy efficiency and productivity of the comminution process in tumbling mills by indirect measurements of internal dynamics—an overview. Energies 13 (24), 6735. http://dx.doi.org/10.3390/en13246735.

Habib, M.K., Ayankoso, S.A., Nagata, F., 2021. Data-driven modeling: concept, techniques, challenges and a case study. In: 2021 IEEE International Conference on Mechatronics and Automation. ICMA, IEEE, pp. 1000–1007. http://dx.doi.org/10.1109/ICMA52036.2021.9512658.

He, M., Tang, J., Guo, X., Yan, G., 2019. Soft sensor for ball mill load using DAMRRWNN model. Acta Automat. Sinica 45 (2), 398–407, URL https://api.semanticscholar.org/CorpusID:221213404.

Hilden, M.M., Powell, M.S., Yahyaei, M., 2021. An improved method for grinding mill filling measurement and the estimation of load volume and mass. Minerals Eng. 160, 106638. http://dx.doi.org/10.1016/j.mineng.2020.106638.

Huang, P., Guo, J., Sang, G., Miao, Q., Jia, M., 2022. Soft measurement of ball mill load under variable working conditions based on deep transfer learning. Meas. Sci. Technol. 33 (7), 075009. http://dx.doi.org/10.1088/1361-6501/ac5c92.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E., 1991. Adaptive mixtures of local experts. Neural Comput. 3 (1), 79–87. http://dx.doi.org/10.1162/neco.1991.3.1.79.

Jonsén, P., Pålsson, B.I., Stener, J.F., Häggblad, H.-Å., 2014. A novel method for modelling of interactions between pulp, charge and mill structure in tumbling mills. Minerals Eng. 63, 65–72. http://dx.doi.org/10.1016/j.mineng.2013.12.005.

Lin, W., Mak, M., Chien, J.T., 2018. Multisource I-Vectors domain adaptation using maximum mean discrepancy based autoencoders. IEEE/ACM Trans. Audio Speech Lang. Process. (TASLP) http://dx.doi.org/10.1109/TASLP.2018.2866707.

Liu, Z., Chai, T., Yu, W., Tang, J., 2015. Multi-frequency signal modeling using empirical mode decomposition and PCA with application to mill load estimation. Neurocomputing 169, 392–402. http://dx.doi.org/10.1016/j.neucom.2014.08.087.

Liu, Y., Yan, G., Li, R., Pang, Y., Qiao, T., 2023. Multi-source domain adaptation method of mill load based on common and special characteristics. In: 2023 42nd Chinese Control Conference. CCC, IEEE, pp. 6682–6688. http://dx.doi.org/10.23919/CCC58697.2023.10240161.

Liu, Y., Yan, G., Li, R., Xiao, S., Ren, M., Cheng, L., 2024. Multi-source unsupervised domain adaptive mill load forecasting method based on deep learning and fusion features. Minerals Eng. 209, 108650. http://dx.doi.org/10.1016/j.mineng.2024.108650.

Lu, S., Zhou, P., Chai, T., Dai, W., 2014. Modeling and simulation of whole ball mill grinding plant for integrated control. IEEE Trans. Autom. Sci. Eng. 11 (4), 1004–1019. http://dx.doi.org/10.1109/TASE.2013.2296309.

Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., Chi, E.H., 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1930–1939. http://dx.doi.org/10.1145/3219819.3220007.

Mayank, K., Malahe, M., Govender, I., Mangadoddy, N., 2015. Coupled DEM-CFD model to predict the tumbling mill dynamics. Procedia IUTAM 15, 139–149. http://dx.doi.org/10.1016/j.piutam.2015.04.020.

Pan, S.J., Yang, Q., 2009. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. http://dx.doi.org/10.1109/TKDE.2009.191.

Pease, J., Curry, D., Young, M., 2006. Designing flotation circuits for high fines recovery. Minerals Eng. 19 (6–8), 831–840. http://dx.doi.org/10.1016/j.mineng.2005.09.056.

Ruder, S., 2017. An overview of multi-task learning in deep neural networks. http://dx.doi.org/10.48550/arXiv.1706.05098, arXiv preprint arXiv:1706.05098.

Sansana, J., Joswiak, M.N., Castillo, I., Wang, Z., Rendall, R., Chiang, L.H., Reis, M.S., 2021. Recent trends on hybrid modeling for Industry 4.0. Comput. Chem. Eng. 151, 107365. http://dx.doi.org/10.1016/j.compchemeng.2021.107365.

Syauqi, A., Eldi, G.P., Andika, R., Lim, H., 2024. Reducing data requirement for accurate photovoltaic power prediction using hybrid machine learning-physical model on diverse dataset. Sol. Energy 279, 112814. http://dx.doi.org/10.1016/j.solener.2024.112814.

Tang, J., Chai, T., Yu, W., Zhao, L., 2012. Modeling load parameters of ball mill in grinding process based on selective ensemble multisensor information. IEEE Trans. Autom. Sci. Eng. 10 (3), 726–740. http://dx.doi.org/10.1109/TASE.2012.2225142.

Tang, H., Liu, J., Zhao, M., Gong, X., 2020a. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In: Proceedings of the 14th ACM Conference on Recommender Systems. pp. 269–278. http://dx.doi.org/10.1145/3383313.3412236.

Tang, J., Qiao, J., Liu, Z., Sheng, N., Yu, W., Yu, G., 2020b. Dual-layer optimized selective information fusion using multi-source multi-component mechanical signals for mill load parameters forecasting. Mech. Syst. Signal Process. 135, 106371. http://dx.doi.org/10.1016/j.ymssp.2019.106371.

Tang, J., Qiao, J., Liu, Z., Zhou, X., Yu, G., Zhao, J., 2018a. Mechanism characteristic analysis and soft measuring method review for ball mill load based on mechanical vibration and acoustic signals in the grinding process. Minerals Eng. 128, 294–311. http://dx.doi.org/10.1016/j.mineng.2018.09.006.

Tang, J., Qiao, J., Liu, Z., Zhou, X., Yu, G., Zhao, J., 2018b. Mechanism characteristic analysis and soft measuring method review for ball mill load based on mechanical vibration and acoustic signals in the grinding process. Minerals Eng. 128, 294–311. http://dx.doi.org/10.1016/j.mineng.2018.09.006.

Tang, J., Yan, G., Liu, Z., Liu, Y., Yu, G., Sheng, N., 2020c. Experimental analysis of wet mill load parameter based on multiple channel mechanical signals under multiple grinding conditions. Minerals Eng. 159, 106609. http://dx.doi.org/10.1016/j.mineng.2020.106609.

Tang, J., Zhao, L.j., Zhou, J.w., Yue, H., Chai, T.y., 2010. Experimental analysis of wet mill load based on vibration signals of laboratory-scale ball mill shell. Minerals Eng. 23 (9), 720–730. http://dx.doi.org/10.1016/j.mineng.2010.05.001.

Wang, Y., Qin, B., Liu, K., Shen, M., Niu, M., Han, L., 2020. A new multitask learning method for tool wear condition and part surface quality prediction. IEEE Trans. Ind. Inform. 17 (9), 6023–6033. http://dx.doi.org/10.1109/TII.2020.3040285.

Xia, H., Tang, J., Yu, W., Qiao, J., 2024. Hybrid simulator-based mechanism and data-driven for multidemand dioxin emissions intelligent prediction in the MSWI process. IEEE Trans. Ind. Electron. http://dx.doi.org/10.1109/TIE.2024.3357896.

Xin, X., Chen, P., Liu, H., Sa, G., Hou, M., Liu, Z., Tan, J., 2024. Research on load prediction of low-calorific fuel fired gas turbine based on data and knowledge hybrid model. Appl. Therm. Eng. 253, 123762. http://dx.doi.org/10.1016/j.applthermaleng.2024.123762.

Yang, L., Cai, J., 2021. A method to identify wet ball mill's load based on CEEMDAN, RCMDE and SRNN classification. Minerals Eng. 165, 106852. http://dx.doi.org/10.1016/j.mineng.2021.106852.

Yin, Z., Peng, Y., Zhu, Z., Ma, C., Yu, Z., Wu, G., 2019. Effect of mill speed and slurry filling on the charge dynamics by an instrumented ball. Adv. Powder Technol. 30 (8), 1611–1616. http://dx.doi.org/10.1016/J.APT.2019.05.009.

Zhang, B., Wang, P., Liu, G., Ma, Z., Zhao, T., 2024. AHU sensor fault diagnosis in various operating conditions based on a hybrid data-driven model combined energy consumption. J. Build. Eng. 87, 109028. http://dx.doi.org/10.1016/j.jobe.2024.109028.

Zhang, Z., Yan, G., Qiao, T., Fang, Y., Pang, Y., 2022. Multi-source unsupervised soft sensor based on joint distribution alignment and mapping structure preservation. J. Process Control 109, 44–59. http://dx.doi.org/10.1016/j.jprocont.2021.11.009.

Zhang, Y., Yang, Q., 2021. A survey on multi-task learning. IEEE Trans. Knowl. Data Eng. 34 (12), 5586–5609. http://dx.doi.org/10.1109/TKDE.2021.3070203.

Zhou, P., Chai, T., Wang, H., 2009. Intelligent optimal-setting control for grinding circuits of mineral processing process. IEEE Trans. Autom. Sci. Eng. 6 (4), 730–743. http://dx.doi.org/10.1109/TASE.2008.2011562.

Zhuo, L., Jian, T., Tian-You, C., Wen, Y., 2021. Selective ensemble modeling approach for mill load para-meter forecasting based on multi-modal feature sub-sets. Acta Automat. Sinica 47 (8), 1921–1931, URL https://api.semanticscholar.org/CorpusID:271326087.