A Hybrid Temporal Data Mining Method for Intelligent Train Braking Systems

WEN JING LIU, GUO CHUN WAN [ID], AND MEI SONG TONG [ID]
Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 201203, China
Corresponding author: Mei Song Tong (mstong@tongji.edu.cn)

ABSTRACT As big data mining technology penetrates into various fields, cross-domain topics driven by data predictive analysis have become important entry points for solving traditional problems. Due to the complex changes of the pressure sensor and the interaction of different grouped trains during the train braking process, the mechanism modeling is difficult, the data is highly temporalized, and the data distribution is not stable. Facing the development trend of long-grouped-heavy-duty train captains, if the braking analysis of the train by temporal data mining of small groups can be used for predictive analysis, it will make innovative progress in the entire train braking field. This paper focuses on combining latest technology such as machine learning, transfer learning and lifelong learning to construct the first predictive analysis research framework in the field of train braking systems. Based on the principle of train braking process and temporal data collected from intelligent experiment platform, a baseline has firstly been built to solve fixed-grouped and multi-grouped temporal prediction problems. Then a predictive algorithm for model verification and update for lifelong learning is established to automatically update model parameters over time. Finally, relying on the parameter transfer in transfer learning, a multi-grouped temporal data prediction analysis is performed. Through comparing the training results of the ''pre-trained'' model on the general domain, the ''tuned'' model on both general domain and the target domain, and the ''target only'' model on the target domain separately, multi-domain tuning results show their applicable scope and transfer conditions. In summary, this work can contribute to intelligently upgrading the semi-physical intelligent test platform for long-grouped-heavy-duty trains.

INDEX TERMS Train braking system, temporal data mining, lifelong learning, transfer learning.

I. INTRODUCTION

With the deepening of freight logistics cooperation all around the world, systematic and efficient research testing and onsite experiments have become the core steps in the development of all heavy-duty transportation technologies. Considering the developing and testing costs of new brakes, the research and development of an intelligent test platform for train braking systems based on the rapid upgrading of computer technology has been greatly improved in recent years. Almost all the research on simulation of heavy-duty train braking systems are concentrated on the analysis and calculation methods of fluid mechanics, then the modeling and predictive simulation of air braking systems are carried out. Classic studies include: the simulation prediction of the train

braking system characteristics of ABDW series valves in the United States [1], [2], performance prediction and parameter research based on vacuum brake system in India [3], and calculation method for solving the gas flow equation of the braking system considering actual thermal effects from Poland [4].

From 2000 to 2004, a research route that made outstanding contributions came from Wei Wei etl. from Dalian Railway Institute in China. This research combines the gas flow theory and the principle of 120-type braking valve to establish a train braking performance analysis simulation software which could predict any group within 150 trains [5]–[7]. In 2007, it further realized the calculation of real-time dynamic properties of the braking system, and used the system to predict braking characteristics of a 20,000-ton combined train [8]. As of 2012, the simulation system based on the basic theory of gas flow could already be combined

---

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Wei Tsai [ID].

with the air brake system for joint simulation, which has a significant impact on the simulation research of long and heavy trains [9]. Another analysis method focuses on the use of mature fluid simulation software AMESim, it uses graphic modeling to have good versatility. In 2015, the calculation model of jetting gas extinguishing agent from gas extinguisher vessel was constructed with AMESim and the jet performance of gas extinguishing agent was simulated by using the two-phase flow model [10].

Although the above research routes have achieved some results, they are currently unable to adapt to future replacement of braking valves and the ultra-long group of heavy-duty trains. On the one hand, there are errors between air-flow-theory-based modeling and actual circuit tests. And because of long research and development cycle, researchers are required to master the physical mechanical structure of air control valves. On the other hand, this kind of system simulation platform lacks the versatility and portability, which cannot adapt to the rapidly updated braking system hardware. Therefore, researchers in the field of train braking systems are paying attention to another kind of intelligent test technology——hardware-in-the-loop simulation (HILS) or semi-physical technology, which has been rapidly developed and applied in various fields of industry since the 1960s.

However, The HILS platforms all over the world come from large companies. These companies have already had a certain international monopoly in the corresponding application fields. For most researchers, whether implementing rapid control prototyping or semi-physical simulation, platforms such as the dSPACE real-time simulation system cooperated by German dSPACE company and American MathWorks company, the RT-LAB launched by Canada Opal-RT Technologies and real-time LabVIEW-RT developed by US National Instruments, have become their first choice. The above HILS technology inevitably has the problems of extremely high cost, poor versatility, and poor portability. Large companies have increased the generalization performance of their simulation systems due to their business breadth, so these platforms are difficult to use in some professional fields, especially the semi-physical simulation analysis of train braking systems that involves safety performance issues.

II. BACKGROUND AND PRELIMINARY
A. INTELLIGENT EXPERIMENT PLATFORM OF TRAIN BRAKING SYSTEM
In the field of rail transportation, research based on the concept of HILS has developed to a certain extent and has obtained corresponding research results.The Chinese Academy of Railway Sciences established the main circuit model of the auxiliary converter of the CRH3 EMU through MATLAB/Simulink [11]. In 2017, a semi-physical simulation test rig of urban EMU network control was tested, which could simulate actual train operation [12]. In 2019, a research was dedicated to the development and validation of a HILS test bench for virtual real-time testing of the GAZ

Group light commercial vehicles equipped with Electronic Stability Control(ESC) systems [13]. Though the application of modern real-time simulation technologies based on HILS allows to decrease the number of onsite tests, most of them are all focused on high-speed EMU.

For the vacancy of HILS technology research and development in the field of heavy-duty train braking system, a research project cooperated by Tongji University and China CRRC Qiqihar aims to solve this cutting-edge problem. By 2017, this research has finished modelling freight train brakes in various stages of braking process based on the fluid balance equation [14], and a semi-physical intelligent experiment platform of freight train braking system was initially established with respect to this model [15]. By 2019, a more detailed design and implementation of simulation modelling and interface parameter provided an important basis for the control part of entire semi-physical braking intelligent test system [16]. At the same time, a scheme for obtaining the braking performance of large-grouped trains by intelligent control of small-grouped trains was proposed [17]. In this work, the neural network algorithm is used to predict the data online and the train is corrected with the predicted value. Figure 1 illustrates the topology and on-site platform of the whole semi-physical intelligent system.

However, this semi-physical system still has some drawbacks. On the one hand, the modeling process based on fluid mechanics is not scalable and versatile, it is limited by the error between numerical and actual solution. On the other hand, although Back Propagation(BP) neural network has been used for performance prediction, it is only the predictive fitting of the abstract parameter of throttling coefficient. Test data of a small grouped train is not enough to ensure the generalization of this prediction method for long and large groups. Moreover, test data of train braking process is not independent. Considering different groups, it has an inseparable relationship with the position and time of the sensor. Classic neural network cannot achieve a good predictive model. Thus, it is necessary to combine diverse methods to analyze large temporal data collected from this intelligent system and help improve the control strategy in return.

B. RELATED RESEARCHES AND TEMPORAL PREDICTION PROBLEMS
At present, the development of temporal data mining technology in the industrial field is still in its infancy, and the research methods for industrial sensor time-series prediction can be mainly divided into two categories. One is based on the classic models of statistics [18], such as the moving average model, exponential smoothing model, Autoregressive Integrated Moving Average (ARIMA) model and state space model, etc.. Because statistical models rely too much on assumptions of stability and so on, the data cannot always be suitable. The other is prediction model based on machine learning, such as K-Nearest Neighbor (KNN), Support Vector Machines (SVM) [19], BP neural network [20], [21] and Deep Neural Network (DNN). Among them, KNN, SVM,
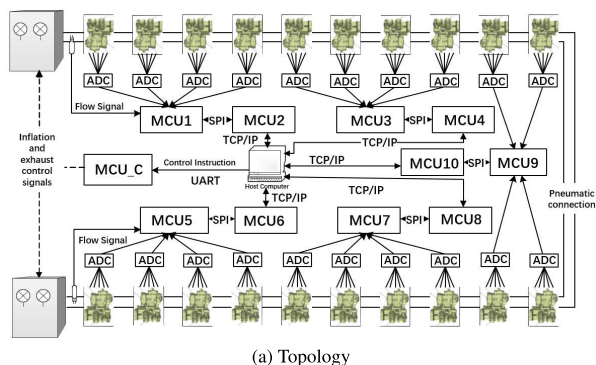
(a) Topology

(b) On-site Platform

FIGURE 1. Semi-physical intelligent experiment platform of train braking system.

BP neural network have simple structures and stable performance, but the prediction accuracy is limited. With the advent of the era of cloud computing and big data, the improvement of computing ability and the substantial increase of training data provide support for deep learning [22], and deep networks represented by Recurrent Neural Network (RNN) [23]. With its advantages of strong versatility and high prediction accuracy, time-series prediction has gradually become a popular research direction. In practical applications, the sensor data rules are mostly related to long-distance time-series data, and the gradient explosion or gradient dispersion of ordinary RNN as the cycle progresses makes the model only learn the short-term dependence [24]. To solve this problem, Long Short-Term Memory Neural Network (LSTM) is introduced [25]. The long-term and short-term memory unit of LSTM can control the accumulative speed of information, and shows superior prediction ability in predicting long-distance dependent time-series data, which is also the theoretical basis of the core network construction of this study. By 2021, some improved models for industrial application based on LSTM have been presented. A multi-output sequential learning model is proposed for Heating and Cooling load prediction [26]. The DB-Net incorporating a dilated convolutional neural network (DCNN) with bidirectional long

short-term memory (BiLSTM) to predict power consumption in integrated local energy systems [27]. The AB-Net incorporating an autoencoder (AE) with BiLSTM for Renewable Energy (RE) generation forecasting [28]. A comparative analysis of a variety of deep features with several sequential learning models is presented to select the optimized hybrid architecture for energy consumption prediction [29].

Based on the prediction and analysis of industrial sensor time series, the introduction of spatial location is the industrial application of temporal data mining. At present, the latest research related to industrial temporal data basically focuses on operation monitoring or equipment failure detection, and is rarely used for modelling and updating of working process. This is why most industrial applications only mention time series rather than temporal data. To apply data mining technology and train braking process in this study, we should consider sensors located in different numbers of trains, different groups and different positions in the same train. A research from JD Financial City Computing Business Department at the 2018 IJCAI Conference provide the basic solution for the problem [30]. It is oriented to urban computing and has realized the temporal sequence prediction of geographic sensors based on the multi-layer attention mechanism neural network. Recently, a multisource adaptation diagnosis network (MADN) method is proposed to transfer the diagnostic knowledge existed in multiple sources to the target [31].

Relying on the above research background and the latest progress, this work focuses on the current industry bottlenecks of train braking systems and the application of temporal data mining technology in this field, mainly to solve the following four problems:

1) How to collect braking data based on the intelligent train braking system with multi-sensors and construct a temporal data set suitable for temporal prediction analysis?
2) How to accurately predict and analyse a single fixed train group based on the temporal data of the train braking system and output the predictive model?
3) How to design a lifelong learning-oriented predictive model in the field of train braking according to the update of data set to achieve continuous model updating and stability verification?
4) How to carry out the transfer training of the model under variable condition of train groups, that is, multigroups, and discuss the suitable application range of transfer learning?

Relying on the experimental platform to build data set to solve Problem 1), a basic model based on LSTM network [32] and a more complicated model [33] have been established to solve Problem 2). The simulation results show that the extended model can predict the following time series data in accordance with the experimental results with high accuracy, which means that the model has good predictive performance in long-grouped train braking problems. With this preliminary, instead of improving the modelling process, this work keeps more focus on solving Problem 3) and 4).

III. MODEL UPDATE AND VERIFICATION OF TEMPORAL PREDICTION BASED ON LIFELONG LEARNING

A. NOTATIONS AND PROBLEM STATEMENT

Formal Problem I: Multi-sensor time series prediction in a single train

In fixed-grouped situations, given n time series collected from input sensors $X = (x^1, x^2, \cdots, x^n)^T = (x_1, x_2, \cdots, x_T) \in \mathbb{R}^{n \times T}$, where T is the length of time window size. We can use $x^k = (x_1^k, x_2^k, \cdots, x_T^k)^T \in \mathbb{R}^T$ to represent an input series of length T and employ $x_t = (x_t^1, x_t^2, \cdots, x_t^n)^T \in \mathbb{R}^n$ to denote a vector of n exogenous input series at time t. Thus, given the previous values of the target series $(y_1, y_2, \cdots, y_{T-1})$ with $y_t \in \mathbb{R}$, as well as the present and past values of n exogenous series $(x_1, x_2, \cdots, x)^T$ with $x_t \in \mathbb{R}^n$, the predictive model aims to learn a nonlinear mapping $F(\cdot)$ to the current value of the target series

$$\hat{y}_T = F(y_1, y_2, \cdots, y_{T-1}, x_1, x_2, \cdots, x_T) \quad (1)$$

Formal Problem II: Multi-sensor temporal prediction in long-grouped trains

In fixed-grouped situations, suppose there are G_g trains, each of which generates G_l kinds of time series to construct individual temporal data set. Among them, one kind of time series is specified as target series for making predictions, while others are features. Given a time window of length T, we use $Y = (y^1, y^2, \ldots, y^{G_g}) \in \mathbb{R}^{G_g \times T}$ to denote the readings of all target series during past T hours, where $y^i \in \mathbb{R}^T$ belongs to i^{th} train. We can use $X^i = (x^{i,1}, x^{i,2}, \cdots, x^{i,G_l})^T = (x_1^i, x_2^i, \cdots, x_T^i) \in \mathbb{R}^{G_l \times T}$ to represent inner features of train i. Among them, $x^{i,k} \in \mathbb{R}^T$ denotes the time series collected from k^{th} sensor in this train. Hence, $x_t^i = (x_t^{i,1}, x_t^{i,2}, \cdots, x_t^{i,G_l})^T \in \mathbb{R}^{G_l}$ represents all temporal data of ith train at time t. Therefore, this prediction problem can be stated as predicting the temporal data of ith train after τ time given the data of all sensors of each train. The predictive model aims to learn a nonlinear mapping $F(\cdot)$ to make:

$$\hat{y}^i = (\hat{y}_{T+1}^i, \hat{y}_{T+2}^i, \cdots, \hat{y}_{T+\tau}^i)^T \in \mathbb{R}^\tau$$
$$= F(Y, X^1, X^2, \cdots, X^{G_g}) \quad (2)$$

B. BASELINE FOR TRAIN BRAKING PREDICTION

To improve and generalize the network in [32], [33] to build a baseline for further research. A two-layer LSTM-based model with an input attention layer added is established, the structure is shown in Figure 2. Given the input sequence of kth external feature as $x^k = (x_1^k, x_2^k, \cdots, x_T^k)^T \in \mathbb{R}^T$, that is, the sensor input feature of the previous trains. We can construct the input attention mechanism through the deterministic attention model, namely multilayer perceptron. The method is to use formula (3) and formula (4) to encode the previously hidden state h_{t-1} and cell state C_{t-1} into the first layer of LSTM cell unit.

$$e_t^k = v_e^T \tanh(W_e \cdot [h_{t-1}; C_{t-1}] + U_e x^k) \quad (3)$$

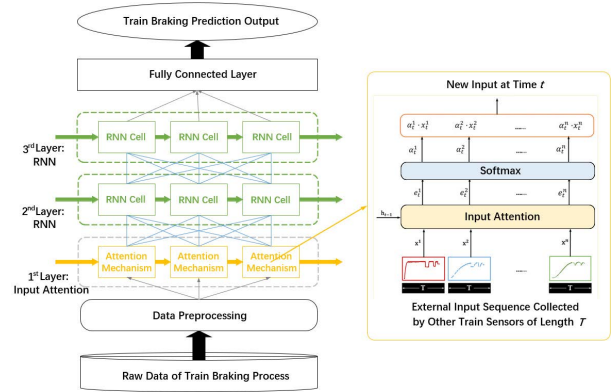$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^n \exp(e_t^i)} \quad (4)$$

FIGURE 2. Predictive network with input attention mechanism added.

A classic LSTM cell unit is composed of three gates, forget gate f_t can be illustrated by formula (5), input gate i_t is shown by formula (6) to (8), output gate o_t is shown by formula (9) to (10). If $h_t \in \mathbb{R}^m$, $v_e \in \mathcal{R}^T$, then $W_e \in \mathcal{R}^{T \times 2m}$ and $U_e \in \mathcal{R}^{T \times T}$ are all trainable parameters. For more conciseness, the bias term in (3) can be ignored. α_t^k is the attention weight which is used to measure the importance of the input sequence of kth external feature at time t. The Softmax function applied after e_t^k is to ensure that the sum of all attention weights is 1.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (7)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (8)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t * \tanh(C_t) \quad (10)$$

The input attention mechanism is a feed-forward network that can be jointly trained with other structural cells evolved from RNN. With these attention weights, we can use formula (11) to automatically extract the features of external time series. Then the state of the hidden layer at time t can be updated with the cell state by formula (12).

$$\tilde{x}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \cdots, \alpha_t^n x_t^n)^T \quad (11)$$

$$h_t = f_1(h_{t-1}, \tilde{x}_t) \quad (12)$$

where f_1 is an LSTM cell unit operating according to (5) to (10), except that x_t is replaced with $\tilde{x}_t$. Through this input attention mechanism, the encoder can selectively focus on certain external input feature sequences without having to treat all input features equally.

C. LIFELONG LEARNING BASED ON DATA SET UPDATE

Lifelong machine learning [34]–[36] (or lifelong learning, LML) is to imitate the learning process and ability of humans. Since the affairs around us are closely related and interconnected, this way of learning is very natural. There are three basic elements of lifelong learning: retention of learned knowledge, selective transfer of previous knowledge when

learning new tasks, and systematic methods to ensure the effectiveness of retention and transfer of knowledge. It is through these three elements that lifelong machine learning has demonstrated a powerful model update effect and self-learning ability. Since lifelong machine learning is a continuous learning, compared to tuning through a single model, the core is to continuously update the learned model and to overcome "catastrophic forgetting" at the same time. Considering all data sets collected from the same system, the stability of data source and application for the same formal problem ensure that it is not a lifelong learning scenario that easily leads to catastrophic forgetting. In the actual problem, the input data set is DS in a certain time period, and the current learned model is M, then the output is the updated prediction model M'. If the data set collected in an experiment is $DS1$, there is a function $F(\cdot)$ such that

$$M(DS1, F(\cdot)) = Y_1 \qquad (13)$$

Then, keeping the test conditions unchanged and collecting another data set $DS2$. There is another function $F(\cdot)$ such that

$$M'(DS2, F'(\cdot)) = Y_2 \qquad (14)$$

If $F(\cdot) = F'(\cdot)$, the model remains unchanged; if $F(\cdot) \neq F'(\cdot)$, the model needs to be updated through design judgement conditions. If it needs to be updated, let $M = M'$. Based on the above basic mathematical problems, the key lies in determining whether the model needs to be updated through the new data set. In normal circumstances, judging method mostly adopts loss function. Considering the continuous accumulation of experimental data in the future, the construction of a lifelong learning-oriented prediction method based on the similarity measurement and loss function of the data. Since this research focuses on the long-term regularity between trains and is limited by experimental conditions, it is not suitable for directly using the method of updating the data stream, so the overall mode of the data set is updating the whole data set. The lifelong learning prediction framework based on data set updating is shown in Figure 3. The updating model is divided into two steps in detail. One is the similarity measurement, which evaluates the degree of change between two data sets. The other is the prediction loss deviation, that is, by designing the loss rate function to determine whether to keep, discard or update the model. Firstly, since the data in this study are all numerical, the similarity measurement method uses the similarity coefficient, and the formula uses the Pearson Correlation Coefficient for calculation. However, considering that the data sets obtained under the same test conditions at different times have a certain degree of similarity, the standard Pearson Correlation Coefficient is taken as the absolute value to ignore the positive and negative correlations. The calculation formula is as (15).

$$
\begin{aligned}
sim(DS1, DS2) &= \left| \frac{cov(DS1, DS2)}{\sigma_{DS1}\sigma_{DS2}} \right| \\
&= \left| \frac{E(DS1 - \mu_{DS1}, DS2 - \mu_{DS2})}{\sigma_{DS1}\sigma_{DS2}} \right| \quad (15)
\end{aligned}
$$

By setting the similarity threshold SIM, when the calculated similarity is less than the threshold, the prediction loss deviation is calculated and evaluated. The value range of SIM is $(0, 1)$. For the predicted loss deviation, it is mainly to judge whether the knowledge base is updated or discarded, and the evaluation method of formula (16) is usually adopted.

$$\theta^* = argmin \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i; \theta_i)) + \lambda \Phi(\theta) \qquad (16)$$

In this model, L is the loss function and $\Phi(\theta)$ is the penalty function. For the problem of train braking process as continuous air pressure prediction, Mean Square Error(MSE) or Mean Absolute Percentage Error(MAPE) shown in equation (17) or (18) can be directly used for evaluation.

$$MSE = \frac{1}{Total} \sum_{t=1}^{Total} (obs_t - pre_t)^2 \qquad (17)$$

$$MAPE = \frac{1}{Total} \sum_{t=1}^{Total} \left| \frac{obs_t - pre_t}{obs_t} \right| \times 100\% \qquad (18)$$

A percentage of predicted loss L_p is defined here to determine whether the knowledge base is updated and discarded. If the prediction loss of the current model M is L_M, and the loss obtained by training the new data set is $L_{M'}$, then

$$L_p = \left| \frac{L_M - L_{M'}}{L_M} \right| \qquad (19)$$

Two thresholds are set here as L_{low} and L_{high}, when $L_p < L_{low}$, the original model is kept, when $L_p > L_{high}$, the model is discarded, when $L_{low} < L_p < L_{high}$, the model is updated.

IV. MULTI-GROUED TEMPORAL PREDICTION BASED ON TRANSFER LEARNING

A. NOTATIONS AND PROBLEM STATEMENT

According to the most authoritative transfer learning review article written by the team of Professor Yang Qiang of the Hong Kong University of Science and Technology [37], transfer learning methods can be divided into four categories according to the learning method: Instance based Transfer Learning, Feature based Transfer Learning, Model based Transfer Learning, Relation based Transfer Learning. The focus here is mainly on the third type of Model-based transfer method, which is also called Parameter based Transfer Learning, which refers to the method of finding the parameter information shared between them from the source domain and the target domain to realize the transfer. The assumption required by this method is that the data in source domain and target domain can share some model parameters. Representative work mainly includes [38], [39]. An algorithm known as TransEMDT (Transfer learning EMbedded Decision Tree) integrates a decision tree and the k-means clustering algorithm for personalized activity-recognition model adaptation [38]. A new dimensionality reduction method is proposed to find a latent space, which minimizes the distance between distributions of the data in different domains
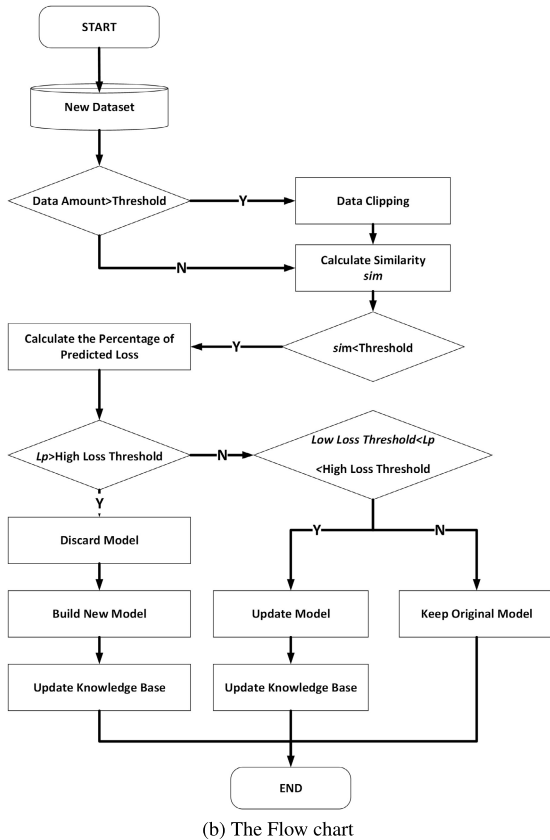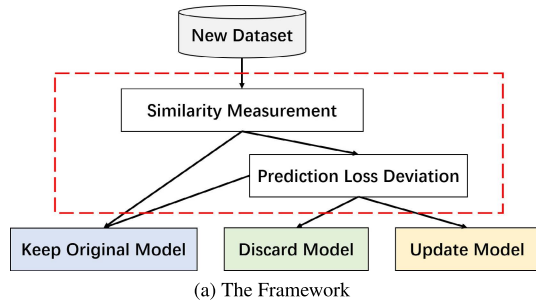
(a) The Framework

(b) The Flow chart

FIGURE 3. The framework and flow chart of lifelong learning prediction based on data set updating.

in a latent space [39]. By 2021, analyzing related operating parameters and designing MLP's structure adjustment strategies can help knowledge transfer among domains for predicting the energy consumption of industrial robots [40].

Through the investigation of existing work, it can be found that most of current model-based transfer learning methods are combined with deep neural networks. By taking advantages of both deep learning and optimal two-sample matching, a unified deep adaptation framework for jointly learning transferable representation and classifier is proposed to enable scalable domain adaptation [41]. The joint adaptation networks(JAN), which learn a transfer network by aligning the joint distributions of multiple domain-specific layers across domains based on a joint maximum mean discrepancy (JMMD) criterion, is presented in [42]. A new CNN architecture in [43] is to exploit unlabeled and sparsely

labeled target domain data. These methods modify some existing neural network structures, add a domain adaptation layer to the network, and then conduct joint training. Therefore, these methods can also be regarded as a combination of methods based on models and features. A standard transfer learning process based on the baseline in Section III.B can be realized by parameter transfer and model fine-tuning. The model training of the data set grouped by G' can be assisted by the learned model grouped by G through the parameter transfer under two different groups. However, it should be noted that the two domains transferred by transfer learning will not have a direct and continuous relationship with each other, which is different from the concept of model verification and update for lifelong learning in Section III.B. Now formalize the temporal prediction problem of multiple groups, and set the time series length as t. G is a group variable. If there are m kinds of groups, then $G \in \{G_1, G_2, \cdots, G_m\}$. Suppose N is the number of train number variables under this group, there are n number variables in total and $n \leq m$, $N \in N_1, N_2, \cdots, N_n$. Let P be the total number of sensor variables in the train number of the group, $P \in \{P_1, P_2, \cdots, P_p\}$. $d_{N,P}^G \in \mathcal{R}^t$ denotes the Pth sensor data sequence collected by the Nth train when the group variable takes G. Since the number of sensors is fixed during actual collection, it can be simplified by omitting the variable P, then $d_N^G \in \mathcal{R}^{p \times t}$ denotes the temporal data sequence collected by the Nth train when the group variable takes G. All train data sequences d_N^G under this grouping can form $d^G \in \mathcal{R}^{l \times p \times t}$. Furthermore, all the groupings finally constitute the cross-grouping temporal data set $D^G \in \mathcal{R}^{m \times l \times p \times t}$. According to the definition of the transfer learning domain, a data set D^G with this nature is a domain. Such a data domain can be arbitrarily used as the source domain D_s or the target domain D_t. Given another definition: When a domain is superimposed as the union of the source domain D_s and the target domain D_t, it is called a general domain, denoted by D_{Gen}.

B. TEMPORAL DATA MINING BASED ON PARAMETER TRANSFER

According to the formal problem in Section IV.A, the entire model/parameter-based braking prediction transfer method is designed. The goal is to clarify the help degree of transfer learning for training process and training result performance. The overall comparison design can be divided into two steps.

Step 1:Build a model on the source data set D_s (as the source domain), and then transfer the weight parameters after training to be used as the initialization parameters (standard transfer learning) of the model to be trained on the target data set D_t (target domain).

Step 2:Based on the combined data from two sources (source domain D_s and target domain D_t), build the model into a general domain D_{Gen}, and pre-train the model built on the source data set D_s in the general domain data. The transfer model of a general domain is defined as the "Pre-tuned Model." Then put it on the target domain D_t to obtain tuning, and the output
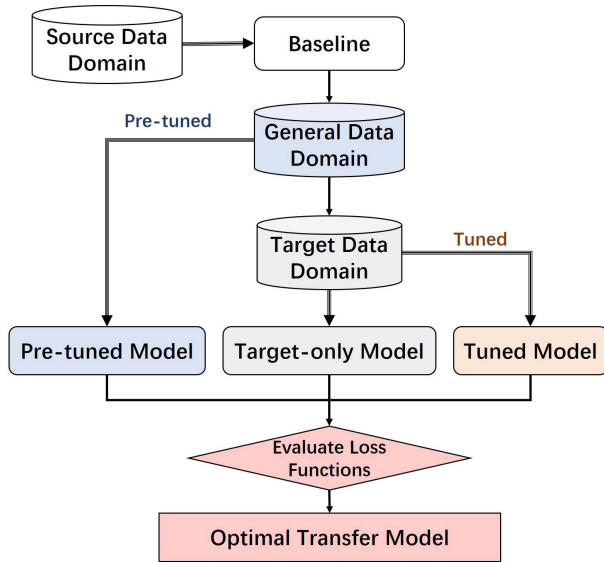
FIGURE 4. The comparative design of transfer methods based on parameter transfer.

multi-domain transfer model is defined as the "Tuned Model." The performance of the two models obtained in this way is compared with the performance of the "Target-only" Model trained only on the target data D_t.

Figure 4 is the comparative design of the entire model/ parameter-based transfer method for this problem. In cross-train transfer training, different target domains are the data domains of different trains in the same group, such as transferring the predicted parameters of the 1st train to the 8th car in the 8-grouped. In cross-group transfer training, different target domains are the data domains of the same train in different groups, such as transferring from the prediction parameters of the 1st train in the 8-grouped to that in the 1st train in the 20-grouped. For a specific problem, loss functions are caluated to evaulate which is the optimal trasfer model.

V. EXPERIMENTS AND RESULTS
A. DATA SETS
Table 1 lists all variables related to the train braking system in the process of multi-sensor data collection. These variables are the basis for the construction of temporal prediction experiments. According to the analysis of train braking principle and braking conditions in Table 1. There are a total of 10 values for the working condition variable S. The specific input and output characteristics are shown in Table 2. In actual test process, there could be mixed working conditions. For multi-working mixed data analysis, S itself is only used for working condition explanation without participating in the predicting process. Table 3 lists part of the data sets collected by multi-sensors of the train braking system. The data sets gather all the characteristics of the data collected by the intelligent test platform, so they are also the main objects for mining and analyzing in this research. All data dimensions and volumes in the characteristic parameters come from the original raw data. Most of the dimensions here refer to the number of sensor groups. Since there are 5 air pressure sensors on a train, the data dimension is generally the number of Groups × 5. The specific prediction problem also involves the relevant variables in Table 1.

B. PERFORMANCE OF IMPROVED BASELINE FOR EXPERIMENTS
As illustrated in Section III, A two-layer LSTM-based model with an input attention layer added is the improved baseline. To compare the result with the model in [33], they are trained with the same hyper-parameters. The setting of hyper-parameters can be referred to [33]. Table 4 shows the comparative results,where Learning Rate initialized as 0.002 with Adam algorithm, Dropout = 0.5, Time Step(T) = 50, Iteration = 1000, Batch Size = 128, Unit = 32, Epoch = 10.

Compared with the model in [33], the evaluation parameters MSE and MAPE of the improved model perform better. MAPE can be increased by 77.87%, and the training time does not increase much.

C. EXPERIMENT I: MODEL UPDATE AND VERIFICATION
1) MODEL UPDATE TUNING FOR FORMAL PROBLEM I
The lifelong learning prediction framework based on the data set update has two main parameters, which are the similarity coefficient and the percentage of prediction loss. The choice of which will affect the stability of the entire prediction. When optimizing similarity for industrial big data, too high or too low thresholds of these two parameters will affect the update result of the model. When performing predictive analysis based on temporal data sets $DS1$ (Single Train) and $DS2$ (5-grouped), a better-performing baseline with an attention mechanism is used here. The data sets of different working conditions under the model are divided separately, and the similar threshold test of multi-sensor in a single train is oriented to the lifelong learning framework through experiments. We first fix $L_{low} = 0.4$ and $L_{high} = 1$ (simplified as 0.4/1), and set the similarity threshold to 0.3, 0.5, and 0.7 for experimentation. Set a flag parameter $FLAG$ here to display the update status of the model. When $FLAG = 1$, the model is updated, when $FLAG = 0$, the model remains. As shown in Figure 5, the test results show that when the similarity threshold $SIM = 0.5$, the update of the model is relatively stable. When $SIM = 0.3$, the algorithm update frequency is low, and when $SIM = 0.7$, it is too high. Therefore, $SIM = 0.5$ is selected.

After determining the similarity threshold, further experiments are carried out on the threshold of the predicted loss percentage, which are respectively set to 0.4/1, 0.4/0.9, 0.4/0.8, 0.3/0.8, 0.3/0.9, 0.3/1. At this time, the flag parameter $FLAG$ has three values. When $FLAG = 2$, the model is discarded. When $FLAG = 1$, the model is updated. When $FLAG = 0$, the model remains. As shown in the experimental results of Figure 6, when the predicted loss threshold

TABLE 1. Related variables of train braking system.

Variables	Name	Type	Notes
G	Group	Numerical	Integer
N	Train Number	Numerical	Integer, $N \leq G$
T	Time Interval	Numerical	/s, the sampling interval is not fixed before preprocessing
P1	Train Pipe Input	Numerical	/KPa, barometric data
P2	Secondary Air Cylinder	Numerical	/KPa, barometric data
P3	Braking Cylinder	Numerical	/KPa, barometric data
P4	Train Pipe Output	Numerical	/KPa, barometric data
P5	Accelerated Release Air Cylinder	Numerical	/KPa, barometric data
ID	Data Label	Nominal	Unique label
S	Working Condition	Nominal	S0-S9, more details are shown in Table 2. The working conditions may be mixed.

TABLE 2. Different input features and labels for different working situations.

Working Condition S	Working Process	Input Features	Output Features(Labels)
S0	Inflation	P1 P4 P5(P3)	P4/P2
S1-S7	Service Braking(7 Stages)	P1 P2 P3(P5)	P4
S8	Emergency Braking	P1 P2 P3(P5)	P4
S9	Re-inflation and Release	P1 P2 P4 P5	P4/P2

TABLE 3. The characteristic parameters of the data sets collected by the train braking system.

Data set ID	Name	Working Condition	Dimension	Raw Data Volume
DS1	Single Train	10 single working conditions	5*10	587
DS2	5-grouped	10 single working conditions	25*10	11262
DS3	8-grouped	All working conditions are mixed except emergency braking	25	26378
DS4	10-grouped	Inflation + Braking + Release	50	27357
DS5	15-grouped	Inflation + Braking + Release	75	12226
DS6	20-grouped I	All working conditions are mixed except emergency braking	100	34974
DS7	20-grouped II	All working conditions are mixed except emergency braking	100	32579

TABLE 4. Comparative results of improved baseline and [33].

Evaluation Parameters	Improved Baseline	Model in [33]
Training_time/s	2539.69±0.13	2417.25±0.28
MAPE($\times 10^{-2}$%)	1.29±0.32	5.83±0.67
MSE($\times 10^{-2}$%)	0.01±0.002	0.02±0.003

is 0.3/0.8, the overall update frequency is relatively stable. Based on the above experimental results, when $SIM = 0.5$, $L_{low} = 0.3$ and $L_{high} = 0.8$, the lifelong learning model update framework combined with the multi-sensor model in a single train can be relatively fixed.

2) MODEL VERIFICATION FOR FORMAL PROBLEM II

Although the prediction network with the input attention mechanism can be used to optimize the prediction of Formal Problem I and Formal Problem II, the actual test data between long trains is not as large as the data collected in a single train. Here we use $DS4$ (10-grouped) and $DS5$ (15-grouped) to verify the baseline model. Table 5 shows the changes in the loss functions MSE and MAPE. The results show that under the average of 10 Epochs, because the data volume of $DS3$ and $DS5$ are equivalent, the MSE of $DS5$ on the baseline is relatively stable, and the MAPE and training time are smaller, indicating that the model has better robustness under the same test conditions.

D. EXPERIMENT II: STANDARD TRANSFER OF CROSS-TRAIN TRANSFER LEARNING

During the migration training test, three data sets were mainly analyzed, namely $DS3$ (8-grouped), $DS6$ (20-grouped I: all working conditions mixed except braking) And $DS7$ (20-grouped II: all working conditions mixed except emergency braking). It should be noted here that although both $DS6$ and $DS7$ are 20-grouped temporal data sets, their are collected under two completely different test conditions. The two sets of temporal data show different temporal characteristics. The working conditions of mixed trains are also different, so it is valuable to do comparative research. However, they both come from the same set of sensor acquisition system, so they meet the universal conditions of transfer learning. The following types of independent tests are carried out on the basis of a learned model. Firstly, a source data domain $DS1$ ($G = 8, N = 1$, 7-stage braking) is defined:

1) Set a target domain $DS1$ ($G = 8, N = 5$, 7-stage braking), train 2 models (with and without transfer learning) on last Q data points of part of the target domain, and predict certain advance value.
2) Train 2 models (with and without transfer learning) on the entire target domain, and predict a certain leading value.
3) Switch to the next target domain $DS1$ ($G = 8, N = 8$, 7-stage braking) until all forward and reverse parameter transfer in $DS1$ are completed.
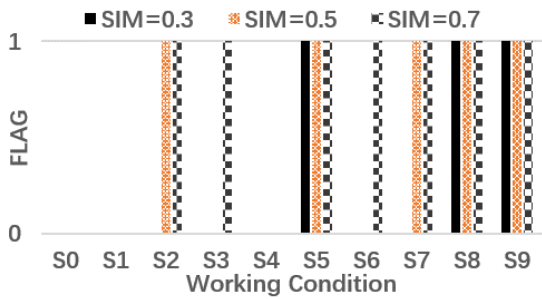
TABLE 5. Comparative verification test on different data sets of baseline.

Evaluation Parameters	DS3(8-grouped,the data set for baseline training)	DS4(10-grouped)	DS5(15-grouped)
Training_time/s	2539.69±0.13	2870.44±1.98	2043.16±0.62
MAPE($\times10^{-2}$%)	1.29±0.32	2.33±0.52	1.03±0.29
MSE($\times10^{-2}$%)	0.01±0.002	0.02±0.003	0.01±0.002
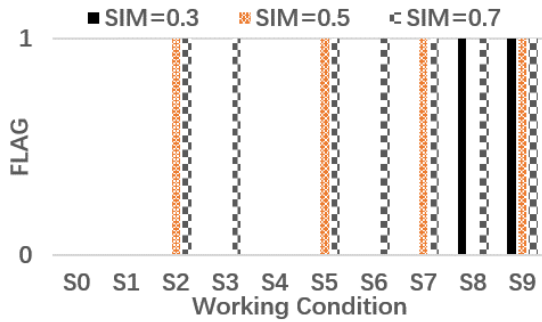
(a) Single Train

(b) 5-grouped

FIGURE 5. Prediction Loss Tuning for Formal Problem I.
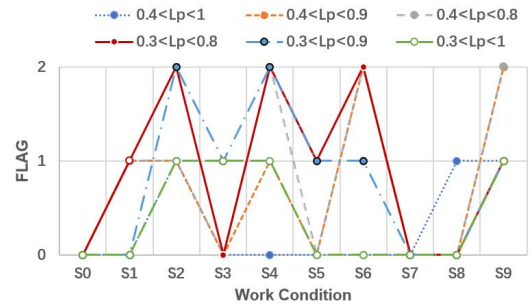

(a) Single Train

(b) 5-grouped

FIGURE 6. Prediction Loss Tuning for Formal Problem II.

4) Switch to the next source data domain $DS6$ ($G = 20$, $N = 1$, inflation and release), and repeat the above process.
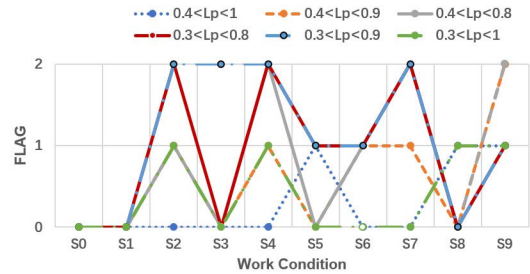
The above experimental design is mainly for two goals. One is to test the impact of transfer learning on prediction performance in the absence of data, and the other is to study how parameter transfer is effective or better for cross-train transfer training problem. Table 6 is a comparison of the Root Mean Square Error(RMSE) results of all test loss functions that are transferred across train parameters.

Analyzing Table 6, it can be seen that transfer learning does not work in all situations. In general, for cross-train parameter transfer training, the following conclusions can be drawn from the RMSE comparison results:

1) Transfer learning is not suitable for parameter transfer under mixed conditions of seven-level braking.
2) The inflatable release cascade test is suitable for reverse parameter migration from $N = 20$.
3) In the case of the train charging release cascade and the charging brake cascade, even if more than half of the

data is missing, the parameter transfer can still greatly improve the prediction performance.

E. EXPERIMENT III: MULTI-DOMAIN TUNING OF CROSS-GROUP TRANSFER LEARNING

Based on the model/parameter transfer method design, further multi-domain experiments are carried out. For the data sets of $DS1$, $DS6$ and $DS7$ across different grouped trains, a multi-domain tuning comparison experiment is carried out by constructing a general domain. The "Pre-tuned" model, "Tuned" model and "Target only" model are trained separately. In general, the Tuned model performs well in cross-group transfer training, and the loss fluctuates relatively stable under various test conditions. This model is especially outstanding in data sets with similar transfer, as shown in Table 7, Figure 7 and Figure 8.

F. RESULTS ANALYSIS

In order to concatenate and apply all the best fit experimental results, we conclude and elaborate all models and quantifiable analysis results as a methodology which lays foundation for future work. Firstly, for multi-sensor time series prediction

**TABLE 6.** Results comparison of *RMSE* for Cross-train transfer learning.

Experimental conditions	DS1(G=8)		DS6(G=20)		DS7(G=20)	
Target domain data volume Q	1500	3563	1600	9209	300	3000
None-transfer learning	1.17	2.33	87.44	103.24	104.22	165.98
N=1 transfer to N=5	3.65	5.78	/	/	/	/
N=1 transfer to N=8	6.33	18.23	16.21	25.72	34.83	132.88
N=1 transfer to N=20	/	/	18.32	35.43	18.72	63.20
N=20 transfer to N=1	/	/	1.33	2.94	2.29	4.33
N=20 transfer to N=8	/	/	2.83	1.02	3.21	8.02

TABLE 7. Results comparison of RMSE for cross-group transfer learning.

Experimental Conditions	DS1+DS6(N=1)	DS1+DS7(N=1)	DS6+DS7(N=1)	DS1+DS6(N=8)	DS1+DS7(N=8)	DS6+DS7(N=20)
Test number	(a)	(b)	(c)	(d)	(e)	(f)
Pre-tuned	3.1958	**4.7913**	8.3810	14.2841	5.1048	33.3270
Tuned	5.9273	**6.5300**	9.6590	7.2134	4.1235	12.7632
Target-only	9.1742	**5.1812**	39.213	10.2141	9.2481	42.260
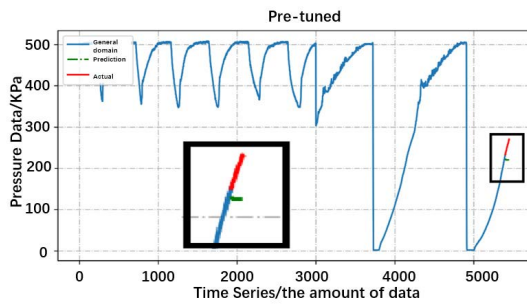
(a) DS6+DS7(N=20) Pre-tuned
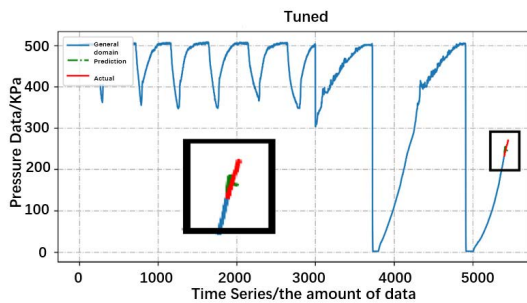
(b) DS6+DS7(N=20) Tuned
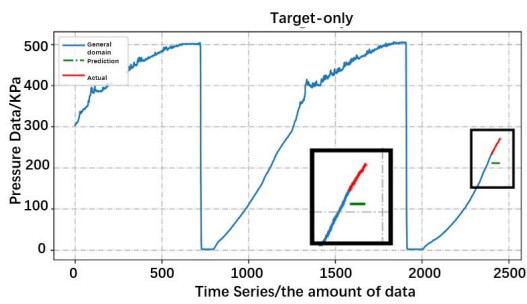
(c) DS6+DS7(N=20) Target-only

FIGURE 7. Multi-domain tuning for *DS*6 + *DS*7($N = 20$).

FIGURE 8. RMSE comparison chart of cross-group parameter transfer experiments.

problem in long-grouped trains, a two-layer LSTM-based model with an input attention layer added in Figure 2 can be a baseline, where Learning 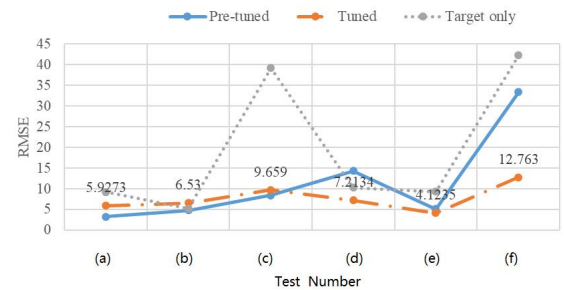Rate initialized as 0.002 with Adam algorithm, Dropout = 0.5, Time Step(T) = 50, Iteration = 1000, Batch Size = 128, Unit = 32, Epoch = 10. Secondly, the framework in Figure 3 can be applied to make full use of a steady stream of experimental data, three core parameters are set as $SIM = 0.5$, $L_{low} = 0.3$ and $L_{high} = 0.8$. Last but not the least, as shown in Figure 4, parameter-based transfer learning could be a good method to benefit training performance. The multi-domain tuning of the "Tuned" model is more robust in cross-group transfer training, and the "Pre-tuned" model performs better on the head-end sensor.

VI. CONCLUSION

This work is the further study on the basis of an intelligent experiment platform of train braking system which can provide large amount of temporal data. We design a lifelong learning-oriented predictive model in the field of train braking according to the continuous model updating and stability verification. In addition, due to the limitations of experimental conditions, we were unable to obtain enough data to carry out the model updating test between long trains. However, this work has clarified the feasibility of this research route. Further study can be carried out in the following two aspects: 1) Lifelong learning prediction based on data set updating between long trains; 2) Transfer learning prediction for lifelong learning.

REFERENCES

[1] K. S. Abdol-Hamid, D. E. Limbert, and R. G. Gauthier, "Simulation of freight train air brake system," in *Proc. ASME Winter Annu. Meeting*, New York, NY, USA, 1986, pp. 5–9.

[2] M. R. Johnson, G. F. Booth, and D. W. Mattoom, "Development of practical techniques for the simulation of train air brake operation," in *Proc. ASME Winter Annu. Meeting*, New York, NY, USA, 1986, pp. 1–4.

[3] S. Bharath, B. C. Nakra, and K. N. Gupta, "Modeling and analysis of railway brake system," *J. Inst. Eng.*, vol. 70, no. 5, pp. 12–17, 1989.

[4] T. Piechowiak, "Verification of pneumatic railway brake models," *Veh. Syst. Dyn.*, vol. 48, no. 3, pp. 283–299, Mar. 2010.

[5] W. Wei and H. B. Li, "Computer simulation of 120 valve test process—Emergency valve model and simulation results," *J. Dalian Railway Univ.*, vol. 21, no. 3, pp. 73–77, 2000.

[6] W. Wei and H. B. Li, "Computer simulation of 120 valve test process—Main valve model and simulation results," *J. Dalian Railway Univ.*, vol. 21, no. 2, pp. 18–24, 2000.

[7] W. Wei and H. B. Li, "Introduction to the simulation program for the air brake performance of the 120-valve train," *Railway Veh.*, vol. 42, no. 7, pp. 34–35, 2004.

[8] W. Wei, "Braking characteristics of 20,000-ton combined train," *J. Traffic Transp. Eng.*, vol. 7, no. 6, pp. 12–16, 2007.

[9] W. Wei, X. B. Zhao, Y. Jiang, and J. Zhang, "The integrated model of train air brake and longitudinal dynamics," *J. China Railway Soc.*, vol. 34, no. 4, pp. 39–46, 2012.

[10] J.-M. Jin, F.-L. An, Y. Shou, R.-M. Pan, Y. Xuan, and Q.-W. Li, "Simulation on release characteristics of the gas extinguishing agent in fire extinguisher vessel with different filling conditions based on AMESim," *Proc. Eng.*, vol. 211, pp. 315–324, Jan. 2018.

[11] S. Q. Song, Y. X. Wang, and N. Yang, "Semi-physical simulation system for auxiliary converter of CRH3 EMU," *China Railway Sci.*, vol. 33, no. 2, pp. 138–143, 2012.

[12] J. J. Zhang, J. Y. Zhao, P. Zhang, J. Yun, and F. Y. Guo, "Construction of semi-physical simulation test rig for urban EMU network control," *Railway Locomotive Motor Car*, vol. 3, no. 4, pp. 5–8, 2017.

[13] A. V. Tumasov, A. S. Vashurin, Y. P. Trusov, E. I. Toropov, P. S. Moshkov, V. S. Kryaskov, and A. S. Vasilyev, "The application of hardware-in-the-loop (HIL) simulation for evaluation of active safety of vehicles equipped with electronic stability control (ESC) systems," *Proc. Comput. Sci.*, vol. 150, pp. 309–315, 2019.

[14] Z. D. Ying, W. H. Gao, and L. Y. Gu, "Analysis of freight car braking performance based on fluid equation of pneumatic system," *J. China Railway Soc.*, vol. 39, no. 11, p. 5965, 2017.

[15] C. Gao, G. C. Wan, and M. S. Tong, "A hardware-in-the-loop model and simulation for the braking system of freight trains," in *Proc. Prog. Electromagn. Res. Symp. (PIERS)*, Shanghai, China, 2016, pp. 3671–3675.

[16] Z. D. Ying, G. C. Wan, W. J. Liu, and M. S. Tong, "Simulation modeling and interface parameter design of the semi-physical braking intelligent test system," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 32, p. e2602, May 2019.

[17] Y. K. Kuang, G. C. Wan, W. J. Liu, and M. S. Tong, "Intelligent control of freight train braking system based on hardware-in-the-loop simulation platform," in *Proc. Photon. Electromagn. Res. Symp.*, Rome, Italy, 2019, pp. 1542–1546.

[18] G. Janacek, "Time series analysis forecasting and control," *J. Time Ser. Anal.*, vol. 31, no. 4, p. 303, Dec. 2009.

[19] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah, "Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique," *Energy Buildings*, vol. 126, pp. 94–103, Aug. 2016.

[20] F. S. Wong, "Time series forecasting using backpropagation neural networks," *Neurocomputing*, vol. 2, no. 4, pp. 147–159, 1991.

[21] L. Wang, Y. Zeng, and T. Chen, "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting," *Expert Syst. Appl.*, vol. 42, no. 2, pp. 855–863, 2015.

[22] X. Q. Cheng, X. L. Jin, and Y. Z. Wang, "Overview of big data system and analysis technology," *J. Softw.*, vol. 25, no. 9, pp. 1889–1908, 2014.

[23] J. Connor and L. Atlas, "Recurrent neural networks and time series prediction," in *Proc. Seattle Int. Joint Conf. Neural Netw.*, vol. 1, Jul. 1991, pp. 301–306.

[24] L. C. Jiao, J. Zhao, and S. Y. Yang, *Deep Learning, Optimization and Recognition*. Beijing, China: Tsinghua Univ. Press, 2017.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] M. Sajjad, S. U. Khan, N. Khan, I. U. Haq, A. Ullah, M. Y. Lee, and S. W. Baik, "Towards efficient building designing: Heating and cooling load prediction via multi-output model," *Sensors*, vol. 20, no. 22, p. 6419, Nov. 2020.

[27] N. Khan, I. U. Haq, S. U. Khan, S. Rho, M. Y. Lee, and S. W. Baik, "DB-Net: A novel dilated CNN based multi-step forecasting model for power consumption in integrated local energy systems," *Int. J. Electr. Power Energy Syst.*, vol. 133, Dec. 2021, Art. no. 107023.

[28] N. Khan, F. U. M. Ullah, I. U. Haq, S. U. Khan, M. Y. Lee, and S. W. Baik, "AB-Net: A novel deep learning assisted framework for renewable energy generation forecasting," *Mathematics*, vol. 9, no. 19, p. 2456, Oct. 2021.

[29] I. U. Haq, A. Ullah, S. U. Khan, N. Khan, M. Y. Lee, S. Rho, and S. W. Baik, "Sequential learning-based energy consumption prediction model for residential and commercial sectors," *Mathematics*, vol. 9, no. 6, p. 605, Mar. 2021.

[30] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3428–3434.

[31] Z. Chai and C. Zhao, "Deep transfer learning based multisource adaptation fault diagnosis network for industrial processes," *IFAC-PapersOnLine*, vol. 54, no. 3, pp. 49–54, 2021.

[32] W. J. Liu, G. C. Wan, and M. S. Tong, "Multi-sensor data mining of single train braking system based on long short-term memory network," in *Proc. Photon. Electromagn. Res. Symp.*, Rome, Italy, 2019, pp. 1537–1541.

[33] W. J. Liu, M. X. Zhang, G. C. Wan, and M. S. Tong, "Multi-sensor data mining of long-grouped train braking system based on an LSTM network," in *Proc. Photon. Electromagn. Res. Symp.*, Xiamen, China, 2019, pp. 2264–2268.

[34] D. L. Silver, Q. Yang, and L. Li, "Lifelong machine learning systems: Beyond learning algorithms," in *Proc. AAAI Spring Symp. Ser.*, 2013, pp. 1–7.

[35] B. Liu, "Lifelong machine learning: A paradigm for continuous learning," *Frontiers Comput. Sci.*, vol. 11, no. 3, pp. 359–361, Jun. 2017.

[36] Z. Chen and B. Liu, "Lifelong machine learning," *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 10, no. 3, pp. 1–145, 2016.

[37] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[38] Z. Zhao, Y. Chen, J. Liu, Z. Shen, and M. Liu, "Cross-people mobile-phone based activity recognition," in *Proc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 11, 2011, pp. 2545–2550.

[39] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 8, 2008, pp. 677–682.

[40] J. Yan and M. Zhang, "A transfer-learning based energy consumption modeling method for industrial robots," *J. Cleaner Prod.*, vol. 325, Nov. 2021, Art. no. 129299.

[41] M. Long, J. Wang, Y. Cao, J. Sun, and S. Y. Philip, "Deep learning of transferable representation for scalable domain adaptation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 8, pp. 2027–2040, Aug. 2016.

[42] M. Long, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. ICML*, 2017, pp. 2208–2217.

[43] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4068–4076.

● ● ●