Intelligent Automation
& Soft Computing

Tech Science Press

DOI: 10.32604/iasc.2023.040124
Article

State Accurate Representation and Performance Prediction Algorithm Optimization for Industrial Equipment Based on Digital Twin

Ying Bai[1],*, Xiaoti Ren[2] and Hong Li[1]

[1]School of Artificial Intelligence and Big Data, Hefei University, Hefei, 230601, China
[2]Anhui Imagine Industrial Technology Co., Ltd., Hefei, 230088, China
*Corresponding Author: Ying Bai. Email: baiying@hfuu.edu.cn

Received: 06 March 2023; Accepted: 12 June 2023; Published: 11 September 2023

Abstract: The combination of the Industrial Internet of Things (IIoT) and digital twin (DT) technology makes it possible for the DT model to realize the dynamic perception of equipment status and performance. However, conventional digital modeling is weak in the fusion and adjustment ability between virtual and real information. The performance prediction based on experience greatly reduces the inclusiveness and accuracy of the model. In this paper, a DT-IIoT optimization model is proposed to improve the real-time representation and prediction ability of the key equipment state. Firstly, a global real-time feedback and the dynamic adjustment mechanism is established by combining DT-IIoT with algorithm optimization. Secondly, a strong screening dual-model optimization (SSDO) prediction method based on Stacking integration and fusion is proposed in the dynamic regulation mechanism. Lightweight screening and multi-round optimization are used to improve the prediction accuracy of the evolution model. Finally, taking the boiler performance of a power plant in Shanxi as an example, the accurate representation and evolution prediction of boiler steam quantity is realized. The results show that the real-time state representation and life cycle performance prediction of large key equipment is optimized through these methods. The self-lifting ability of the Stacking integration and fusion-based SSDO prediction method is 15.85% on average, and the optimal self-lifting ability is 18.16%. The optimization model reduces the MSE loss from the initial 0.318 to the optimal 0.1074, and increases R^2 from the initial 0.731 to the optimal 0.9092. The adaptability and reliability of the model are comprehensively improved, and better prediction and analysis results are achieved. This ensures the stable operation of core equipment, and is of great significance to comprehensively understanding the equipment status and performance.

Keywords: Digital twin (DT); digital representation; transfer learning; dual model optimization; information fusion

1 Introduction

With the development of the Industrial Internet of Things (IIoT), to understand the future operating state of equipment, it is necessary to accurately capture the state characteristics of industrial equipment and accurately predict the operating performance of equipment. Competencies in collecting data about the equipment and predicting its abilities are key in progressing from passive maintenance of the system to active prevention of complications. Conventional manual spot inspection methods often lead to difficulties in quickly capturing the degradation characteristics of large industrial equipment. Once failure or serious performance degradation occurs, a high amount of human, material and financial losses can be incurred. The combination of DT technology and the IIoT enables sensor-based information transfer to be linked with digital-driven state representation and conventional device-based performance trend prediction.

Through the IIoT network, the equipment status of industrial systems can be remotely monitored, and subject to intelligent perception as well as intelligent recognition before data-driven dynamic representation of the equipment's entire process is obtained. A machine learning-based optimization algorithm is used to identify and predict the equipment's health status; then, the lightweight digital prediction and evaluation system is derived. These algorithms are of great significance in ensuring the operation safety as well as the reliability and effectiveness of complex equipment.

Conventional statistical fusion of sensor data has limitations in establishing an effective correlation between historical data and equipment health status. Common digital data-driven methods are mainly based on data modeling and analysis, which often make representing the state information of equipment in real-time difficult [1].

The DT system can fully and accurately represent the changing states of large, complicated equipment throughout its life cycle. Thanks to the emergence of DT technology, more data on the core equipment in the industrial system can be visually represented [2]. Within this system, bidirectional transmission and regulation of virtual-real model data are particularly significant. The combination of IIoT, sensor output, and big data analysis is adapted to update real-time data based on drive data [3]. The results of the simulation control and virtual computation are applied to the physical entity to form a closed-loop control [4]. The abilities of DT systems include simulation, monitoring, regulation, and advanced prediction. Based on real data, the virtual system updates the representation indicators in real-time and provides better predictive decision support for the physical system [5].

However, dynamic representation based on DT also faces some challenges when combined with large-scale industrial equipment. In the past, researchers focused on the role of big data sample analysis on equipment performance prediction; less attention has been paid to the complexity of real operating environments and the instability of data samples [6]. These factors affect the accuracy of real-time representation of a DT system. The closed-loop control of the system is based on steady-state statistical data, and the optimal regulation of the control strategy cannot reach the effect of dynamic feedback and real-time adjustment [7]. The evolution prediction of key equipment state depends more on expert experience or prior degradation characteristics. Thus, the inclusive and dynamic evaluation ability of the evaluation system has limitations [8]. To overcome the existing research problems, this paper makes the following contributions:

- A dynamic regulation mechanism is proposed in the new DT cluster system, integrating IIoT with algorithm optimization. The dynamic perception and intelligent feedback of complex equipment states are realized.
- A strong screening dual-model optimization prediction method based on Stacking integration and fusion is proposed to improve the accuracy of dynamic regulation in the DT system.

- The real-time state representation and life-cycle performance prediction of large key equipment are significantly optimized.

The remainder of this research article is organized as follows. Section 2 reviews and discusses related work. Section 3 establishes the dynamic real-time feedback model based on a DT system, and analyzes the optimization mechanism of the SSDO algorithm. Section 4 verifies the effectiveness of the proposed model and method through experimental evaluations. Section 5 covers the conclusions of this study and discusses future work.

2 Related Work

2.1 DT Dynamic Perception Combined with IIoT

Digital twinning is an effective means of establishing the all-platform digital state representation. In real IIoT scenarios, dynamic representation based on a DT is gradually applied. The combination of the IIoT and DT has gone through several signature phases [9]. In the single simulation, physical entity simulation [10], industrial environment DT capture, DT industrial environment application, DT and IIOT effective communication stages [11], some scholars have carried out continuous technical exploration. It shows obvious advantages in real-time information capture and dynamic data display [12]. For example, in an intelligent transportation system, the DT-supported IIoT architecture is built to capture information from the equipment base station to achieve intelligent perception and intelligent control of the infrastructure [13]. Platenius-Mohr et al. [14] built an interoperable digital twin under the IIoT system to realize flexible transformation of information model. Cheng et al. [15] used DT technology to build an enhanced framework for the IIoT, and propose improvement strategies for effective data transmission from DT systems. Cecil et al. [16] proposed a network physical framework based on the Internet of Things, which includes five collaborative entities: management system, cloud services, network components, physical information interaction and operating equipment. The combination of DT with IIoT, PLC control system, cloud storage and transmission, and web visualization has potential applications in solving the coupling problem of different system architectures. Tao et al. [17] conducted in-depth studies on standard model and model standardization construction. However, there remain some difficulties in terms of explaining the real-time information interaction of complex data-driven by big data. The industrial application of DT technology and efficient information communication remains a major challenge for the combination of DT technology and IIoT.

To solve these problems, the authors focus their attention on dynamic representations of industrial information with feedback regulation. Combined with IIoT, a new DT architecture is constructed. The real-time interactive perception between the DT system, control system, multi-sensor network, and the physical entity is preliminarily realized. The real-time dynamic representation of complex equipment operating state can be completed.

2.2 DT Dynamic Model Based on Machine Learning Algorithm Optimization

Digital twins can present and regulate some complex industrial operating states that are difficult to identify intuitively. Combined with big data machine learning processing methods and deep learning algorithm optimization, the analysis platform can independently learn, reason, train, and model.

Many scholars have studied improving the self-learning ability and self-optimization ability of the DT system. The earliest combination of machine learning and DT models is predictive analysis for fault diagnosis [18]. It has been gradually improved in the aspects of data cleaning, deep learning performance prediction [19], data fusion feature extraction and complex equipment optimization [20].

Xia et al. [21] proposed a reinforcement learning framework to adjust the risk of real-time abnormal values to the system. Xiao et al. [22] established a bionic information physical system for smart power plants and suggests that a smart system should possess the abilities of self-learning, self-drive optimization, self-coordination, and global optimization. Rasheed et al. [23] modified the hybrid prediction model driven by DT technology and uses the hybrid improved algorithm to predict the component life, thus effectively improving the prediction accuracy of the system.

If data modeling analysis is carried out on complex DT models with insufficient data samples, obtaining an accurate evolutionary model will be challenging. Kucera et al. [24] completed a large number of theoretical and experimental studies on the virtual-real interaction and logic control optimization of production systems. However, the control evaluation model relies on the experience and knowledge of experts, and the system is prone to problems such as inaccurate, untimely, and unintelligent representation results. Liu et al. [9] proposed a state representation and predictive analysis method for machine learning algorithms at the application level. However, the lack of a comprehensive evaluation mechanism needs to be considered when realizing the state representation of the whole life cycle using this method.

To solve the above problems, the authors focus on establishing a dynamic regulation mechanism in DT system. The existing DT-IIoT platform combined with machine learning algorithm is used for optimization. The collected data is continuously added in real-time and the valid samples are supplemented to engage in the global control evaluation. The proposed dynamic regulation mechanism improves the representation and prediction ability of the DT model effectively.

2.3 Multi-Model Integration Fusion for DT Evolution Model

As the operating environment of real industrial sensors is complex, full sample data needs to be quickly and effectively screened. Malakuti et al. [25] solved the unstable data-dependent interference in machine learning systems by integrating DT models. Li et al. [26] proposed that multi-sensor information fusion and error analysis are key in ensuring the accuracy of information extraction from the source. Multi-model integration and fusion can effectively improve the prediction performance of the model. Currently, model fusion based on weighted combination and bagging algorithm is comparatively more widely used than fusion based on Stacking model. Some of the commonly-used Stacking-based models include: Kernel Ridge Regression Model (KRR), Decision Tree Model (dt_m), Gradient Boosting Regression (GBDT), Random Forest (RF), and Neural Network Model (NN). Li et al. [27] used the regression single model to optimize the parameters of a heating boiler and realize the multi-operating condition performance optimization of the boiler. An XGBoost model is adopted for load prediction; the load prediction of similar days is analyzed via a clustering method. Chen et al. [28] compared this model with deep learning Long-Short Term Memory (LSTM) networks, and conclude that the XGBoost method has more advantages in terms of higher accuracy and superior generalization ability. Combining different types of base Stacking models leads to comparative advantages and significantly improved prediction accuracy of the model.

The authors pay more attention to the inclusive optimization effect of the integration model. Through the comprehensive comparison of various models, we try to obtain a more inclusive performance prediction and evaluation system.

In summary, through the combination of DT-IIoT platform and machine learning algorithm optimization, the DT model has the capability of multi-sensor real-time perception and equipment performance prediction. In this paper, a dynamic, high-accuracy, and lightweight DT cluster system is studied by combining DT-IIoT with integration algorithm optimization. The DT-IIoT optimization model realizes real-time dynamic feedback regulation. It also provides accurate state representation and performance evaluation for industrial key equipment.

3 DT-IIoT Architecture and Optimization Methods

A DT-IIoT optimization model combined with integration algorithm optimization is used to construct a dynamic regulation mechanism. The system-level real-time bidirectional regulation is created based on algorithms and a control system. A strong-screening dual-model optimization prediction method based on stacking integration and fusion is proposed to improve the accuracy of the prediction algorithm. The framework of this DT-IIoT optimization model is presented in this section.

3.1 Basic Architecture of DT-IIoT Optimization Model

The DT-IIoT optimization model tries to build a DT cluster system with real-time feedback and dynamic regulation. It provides real-time perception, state representation, and performance prediction evaluation for industrial key equipment. The new DT system consists of an equipment acquisition layer, data perception layer, data transfer layer, intelligent analysis layer, and display layer, as shown in Fig. 1.

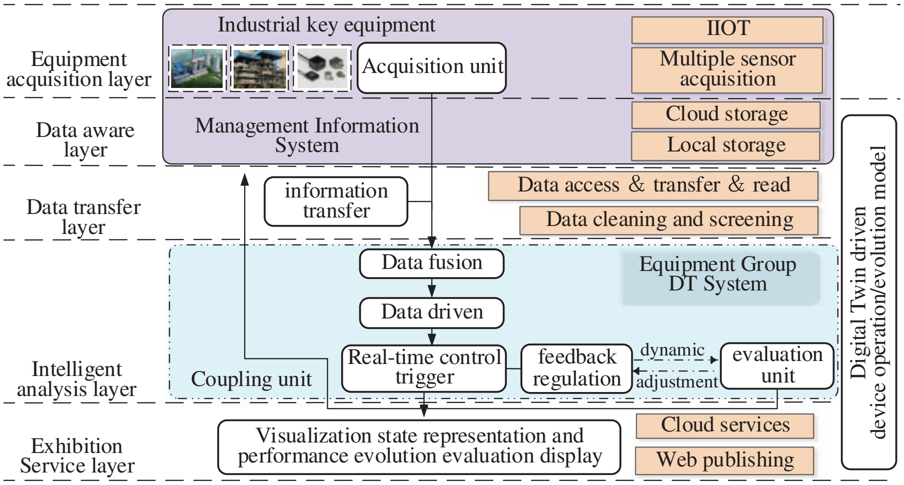

Equipment acquisition layer
Industrial key equipment
Acquisition unit
IIOT
Multiple sensor acquisition
Data aware layer
Management Information System
Cloud storage
Local storage
Data transfer layer
information transfer
Data access & transfer & read
Data cleaning and screening
Data fusion
Equipment Group DT System
Data driven
Intelligent analysis layer
Coupling unit
Real-time control trigger
feedback regulation
dynamic
adjustment
evaluation unit
Digital Twin driven device operation/evolution model
Exhibition Service layer
Visualization state representation and performance evolution evaluation display
Cloud services
Web publishing

Figure 1: Architecture of DT-IIoT optimization model for industrial key equipment

By taking advantage of real IIoT networking, multi-sensor information for industrial equipment is captured in real-time. Time series data from the equipment layer is collected into local storage or cloud storage to construct a real-time information management system. A DT virtual equipment cluster model is established by 3D modeling. Data preprocessing, data state representation and performance prediction analysis of time series data are completed in virtual space. The DT-IIoT optimization model optimizes the above processes in subsequent order.

Data preprocessing optimization strategy: To solve the influence of discrete data instability on data representation, the DT-IIoT optimization model uses DT virtual space to implement multiple rounds of strong screening and optimal feature dimension reduction selection strategies to obtain more stable data samples. By accessing, transmitting, reading, and screening the collected data in the Data transfer layer, a data-driven sample of the DT system is obtained, and real-time data-driven trigger control is realized.

Data state representation optimization strategy: The dynamic regulation mechanism of DT system with global feedback response is established. In the Intelligent analysis layer, the DT-IIoT platform combined with machine learning algorithm is used for optimization. Continuous adding of the data collected in real-time and supplementation of the valid samples allow participation in the global control evaluation. The DT system continuously adds data from the management system to expand the valid samples of the analysis model. Real-time data is used to participate in the training and correction of machine learning models. The calculation results are fed back to the management control system to realize the global real-time control evaluation.

Performance prediction optimization strategy: A SSDO prediction method based on Stacking integration and fusion is proposed in the dynamic systematic regulation. A lightweight evolution prediction system is constructed based on the inclusiveness of the integrated model, and the equipment performance in the whole life cycle is represented digitally.

3.2 Dynamic Regulation Mechanism Implementation

The dynamic regulation mechanism, which combines the advantages of the IIoT platform and machine learning algorithm optimization, is the global feedback regulation. It is helpful to link the feedback adjustment of DT system to the real system and the guiding adjustment of the originally collected data to the DT system. The steps of dynamic regulation mechanism are shown in Fig. 2. (1) Transmitting multi-sensor data samples from the acquisition layer to DT-IIoT optimization model. (2) Checking feedback real-time operating data between DT model and the control system. (3) Sensing multi-sensor features to realize control triggers based on equipment performance data. (4) The lightweight prediction model based on the Stacking-integrated optimization algorithm is constructed to realize the evolution prediction of existing data. (5) Global real-time control evaluation: Combined with the real-time supplementary data, the evolutionary data value is compared with the expected value. If it meets the expectation, the iteration is continued; if not, the adjustment strategy is fed to the real control system.

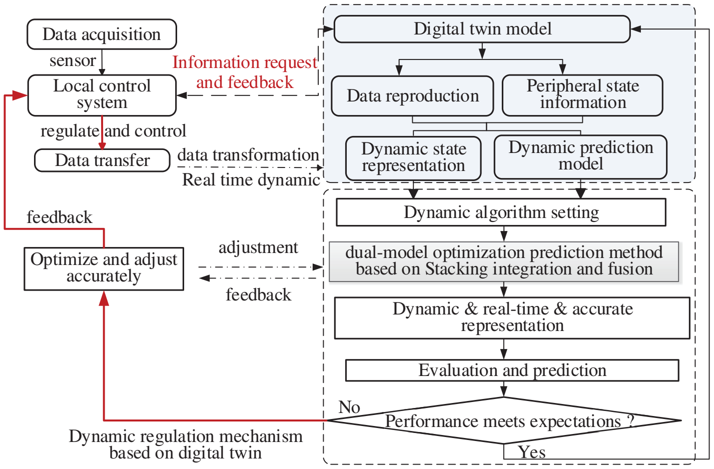

Data acquisition
sensor
Local control system
Information request and feedback
regulate and control
Data transfer
data transformation
Real time dynamic
feedback
Optimize and adjust accurately
adjustment
feedback
Digital twin model
Data reproduction
Peripheral state information
Dynamic state representation
Dynamic prediction model
Dynamic algorithm setting
dual-model optimization prediction method based on Stacking integration and fusion
Dynamic & real-time & accurate representation
Evaluation and prediction
No
Performance meets expectations ?
Yes
Dynamic regulation mechanism based on digital twin

Figure 2: Implementation steps of dynamic regulation mechanism

3.3 Stacking Integrated Optimization Algorithm Implementation

Because of the complexity, diversity and instability of industrial data, the dynamic representation and prediction of equipment real-time performance are greatly affected. In the self-adjustment and reconstruction stage of the DT model, the inclusive lightweight prediction model is more conducive to real-time representation. A SSDO prediction method based on stacking integration and fusion is proposed. Machine learning algorithms are used to optimize the multi-sensor information fusion process.

The stacking integration algorithm is introduced into the global fusion optimization of the DT-IIoT optimization model. A complete dual-model prediction optimization is constructed by multi-round optimization of the base model and meta-model. The dynamic regulation mechanism proposed in Section 2.2 is used to achieve real-time performance prediction and evaluation. The construction method is shown in Fig. 3. Implementation details are shown in the following section.

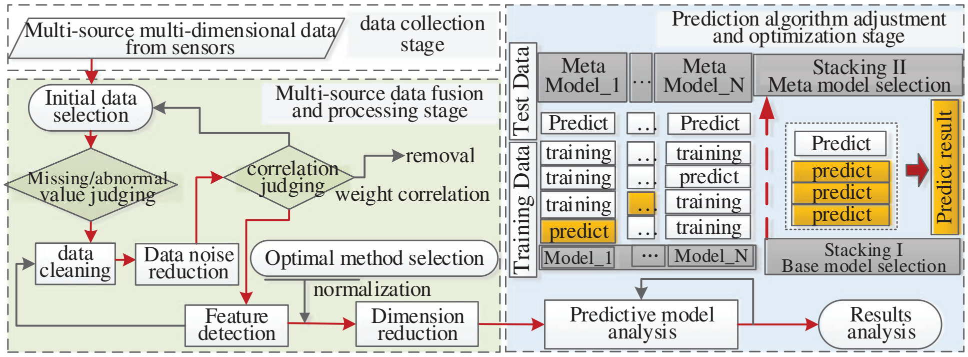


Figure 3: The overall strategy of strong screening dual-model optimization (SSDO) prediction algorithm

3.3.1 Data Preprocessing Optimization Implementation

(1) Data Lightweight Strong Screening

To eliminate the influence of discrete data instability on data representation, Spearman correlation analysis is used to clean the multi-sensor data in DT system. Feature dimension reduction of target performance data is implemented by a multicollinearity check. Multivariate collinearity test results of variance inflation factor (VIF) are used as the weight value to participate in the feature heterogeneity evaluation of Spearman correlation coefficients. Lightweight screening and accurate feature retention of massive data are realized. The basic weight value optimization method is shown in Eqs. (1)–(3):

$$W_E = \frac{1}{1 - R^2} \tag{1}$$

$$r = 1 - \frac{6}{n\left(n^2 - 1\right)} \sum_{k=1}^{n} \left(R_k - Q_k\right)^2 \tag{2}$$

where R^2 is the coefficient of determination. Suppose $(\mathrm{X}_k, \mathrm{Y}_k)$ is the sample taken from the population sample (X, Y), $\mathrm{k} = 1, 2, \dots, \mathrm{n}$, and n is the total number of samples. Rank the elements in X, Y in ascending order. Using R_k to represent the rank of X_k in X. Using Q_k to represent the rank of Y_k in r. The feature values samples with a certain correlation between x and y are selected as follows:

$$Fea_s = \left\{ \left(W_1 \times Fea_1, \dots, W_i \times Fea_i, \dots, W_n \times Fea_n\right)_{opt}, W_i = \prod_{i=1}^{n} r_i \times \mathrm{W}_{\mathrm{E}_i} \right\} \tag{3}$$

Fea is the category of feature values, S is the optimal selection value; r is the correlation coefficient; i is the number of feature samples;

(2) Lightweight Dimension Reduction for Multi-feature Data

Lightweight dimension reduction is considered in this part to preserve data integrity. The typical dimension reduction methods include linear dimension reduction methods, such as PCA, ICA (Independent Component Analysis), and MDS (Multidimensional Scaling); Kernel-based nonlinear dimension reduction methods, such as KPCA; Feature values-based nonlinear dimension reduction methods (manifold learning), such as ISOMAP, LLE (Locally Linear Embedding), and Spectral Embedding methods [29]. To compare the adaptability of the above methods to industrial discrete data processing, the implementation steps are as follows: (1) Selecting the target prediction model Model_aim preliminarily. (2) Standardizing the preprocessed data, using 7 different classical methods to reduce the dimension of data samples. The obtained new data samples $\mathrm{Data_i}$ are divided into new training sets and new test sets. (3) New data sets for each category are brought back into Model_aim to obtain prediction results. (4) Evaluating the effects of dimensionality reduction algorithms by evaluation index (such as R^2). The optimal preprocessing algorithm Fea $_{\mathrm{OPT}}$ is obtained.

$$\mathrm{Data_{new}} = \mathrm{train\_test\_split(Data_i, train,\ target)}\ \mathrm{i} = 0....6, \tag{4}$$

$$\mathrm{Fea\ alg_{OPT}} = \max\left\{\mathrm{R^2\_score(test\_target,\ Pred_i)}\right\}\mathrm{i} = 0....6 \tag{5}$$

Stable data samples are obtained through multiple rounds of strong screening and comparison of the feature dimension reduction method.

3.3.2 Dual-Model Stacking Integrated Optimization Implementation

Conventional Stacking integration models divide data sets into multiple subsets. Independent performance prediction is performed for each multilevel training subset. The predicted results are used as the new reference group to obtain the final prediction results. In model selection, multiple base models and one meta-model are used to directly obtain prediction results, as shown in Fig. 3. Because the Stacking integration structure absorbs the processing advantages of every single model, the accuracy of the prediction model is effectively improved.

In this paper, dual-model stacking integration optimization is proposed based on the Stacking integration structure. Dual-model and multiple rounds hyperparameter optimization are performed for both the base model and meta-model. The implementation steps are as follows: (1) Dividing k-fold subsets, the grid search method is used to optimize the hyperparameter of different target base models. (2) Changing the combination number of base models, and implementing global optimization of model accuracy, the lightweight combination structure with the highest adaptability is obtained. (3) Optimizing the single base model with the highest accuracy again. The GWO algorithm, GA algorithm, PSO algorithm, CV-Grid Search, and other optimization strategies are used in obtaining the optimal meta-model. (4) The new meta-model is brought back into the Stacking integration model to participate in the global optimization. In each round of global optimization of the integrated model, the CV-Grid Search method is used for the first optimization. The second optimization adopts an automatic hyperparameter tuning method, and the global optimal lightweight prediction model is obtained. The optimization strategy flow is shown in Fig. 4.

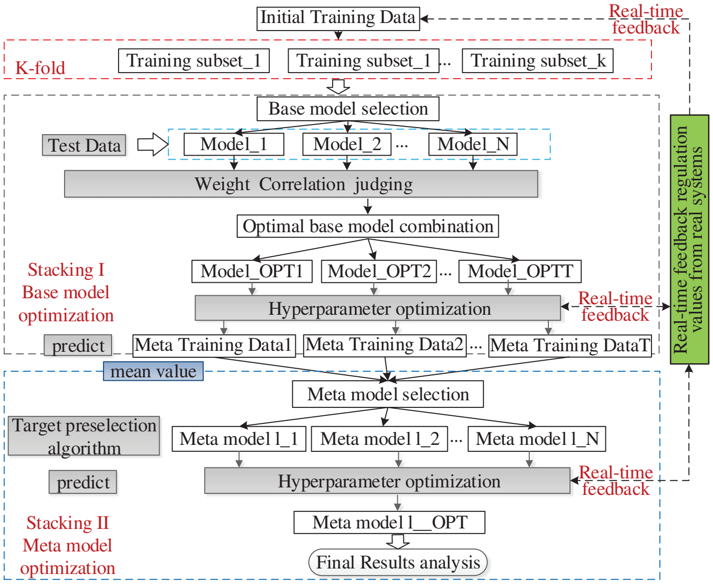


Figure 4: Dual-model stacking integrated optimization algorithm

In the stackingI layer, eight typical algorithms are selected as comparison targets to participate in performance comparison and combinatorial optimization. These algorithms include: Random Forest (RF), Gradient Enhancement (GB), Adaptive Boosting (Ada Boost), K-nearest Neighbors (KNN), Support Vector Machine (SVR), XGBoost, Extreme Tree (ET), and Decision Tree (dt). By correlation analysis and combinatorial optimization, a stackingI minimalist model with a high fitting degree and the minimum number of basic models is obtained. In the stackingII layer, a prediction model with a higher fitting degree and stronger generalization ability is selected based on the comparison results from stackingI. The optimal real-time lightweight model is obtained through the use of local hyperparameter tuning with feedback. Through secondary optimization, the generalization ability and prediction accuracy of the algorithm are improved comprehensively.

3.4 *Evaluation Index*

To verify the global optimization performance of the system, Mean Squared Error (MSE) and determination coefficient: R^2 (R-Square) are adopted as evaluation indexes, expressed as:

(1) Mean Squared Error: MSE

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{6}$$

where $y_i - \hat{y}_i$ is the real value - the predicted value on the test set. m is the number of samples.

(2) Determination Coefficient: R^2 (R-square).

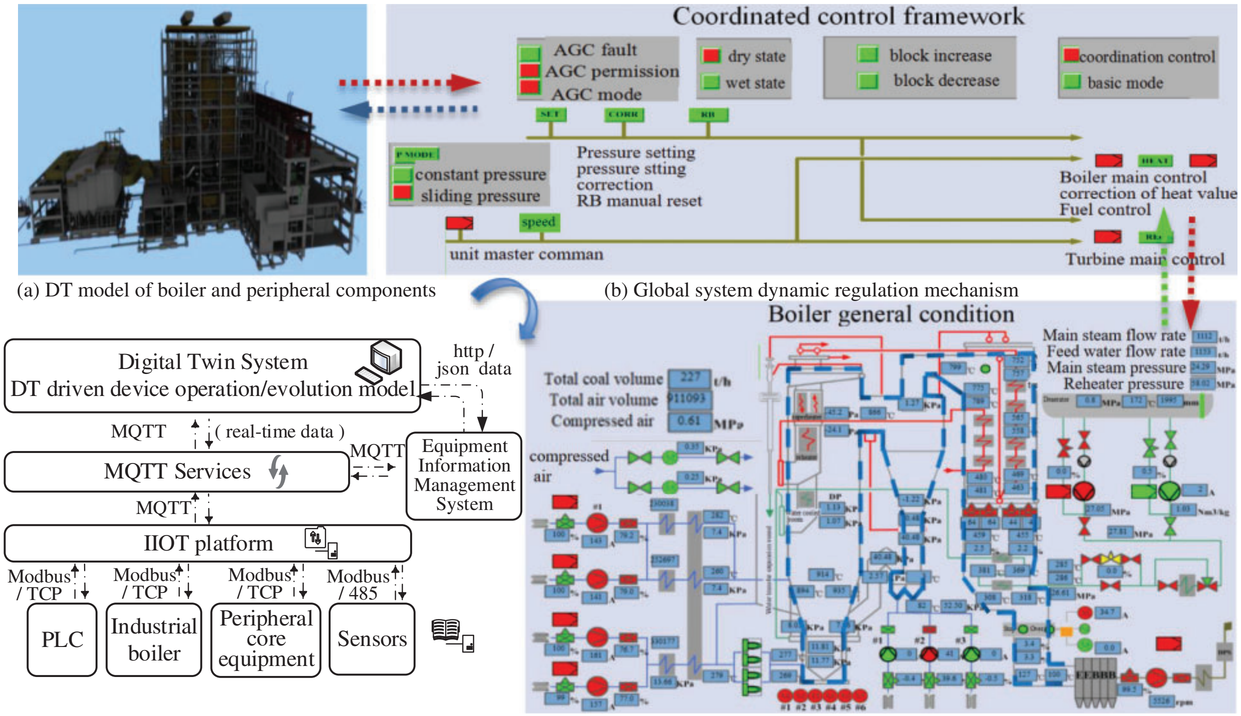
$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \tag{7}$$

where y_i: real value $\hat{y}_i$: predicted value.

4 Results and Discussion

4.1 Construction of Digital Twin System

To verify the optimization effect of the DT-IIoT optimization model, the key equipment boiler in a large power plant is applied as the research object in this paper. Taking the actual IIoT operation data of a power plant located in Shanxi Province as a case study, all used data is self-collected data. The steam production performance of a 2 × 350 MW circulating fluidized bed boiler is analyzed. The DT-IIoT model of the boiler and peripheral equipment is established, as shown in Fig. 5.



(a) DT model of boiler and peripheral components

(b) Global system dynamic regulation mechanism

(c) DT-IIoT system information communication

(d) Boiler and sensor feature real-time representation

Figure 5: Digital twin model construction and control strategy framework

The IIoT and multi-sensor information collection are used to sense and read all kinds of real information about boiler steam quantity and boiler peripheral equipment. The boiler is a supercritical pressure operation boiler. It adopts a single furnace, M layout, primary intermediate reheating, and circulating fluidized bed combustion mode. The main parameters collected by the sensor are: mainstream flow, main steam pressure, main steam temperature, reheating steam pressure, reheating steam temperature, feed water flow, feed water temperature, furnace negative pressure, smoke exhaust temperature, economizer outlet flue gas oxygen content, total air supply pressure, primary air pressure, pressure difference between bellows and furnaces, preheater outlet air temperature, coal feed to coal feeder, primary air volume, primary air temperature, bed temperature, bed pressure, re-feed fluidized air pressure, etc. The information collected by the sensor is mainly from the equipment information management system of the local control system. The DT-IIoT optimization model conducts data-based strong screening and data dimension reduction. The proposed strong screening dual model optimization prediction method based on Stacking integration and fusion is used to dynamically adjust the evolution prediction of steam quantity. Local 3D modeling representation based on digital twinning is shown in Fig. 5a. The IIoT system uses Modbus and TCP communication protocols to obtain PLC control data, boiler operation parameters, peripheral key equipment parameters, and other auxiliary performance parameters. The actual operation data is transferred to the DT system, subscribed, and received/sent via the MQTT protocol. The DT virtual engine is connected by MQTT to realize the control trigger. The dynamic regulation mechanism proposed in this paper is used to realize the global systematic feedback regulation and state representation, as shown in Fig. 5c.

The global system dynamic regulation mechanism is shown in Fig. 5b. A machine learning-based DT-IIoT platform works with control system to establish bidirectional response regulation. The DT model constantly introduces data from the management system, expands the analysis model, and uses real-time data to participate in the training and correction of machine learning model. The training results are fed back to the management system to realize the global control evaluation. The implementation architecture of the dynamic regulation mechanism, in this case, is shown in Fig. 6.

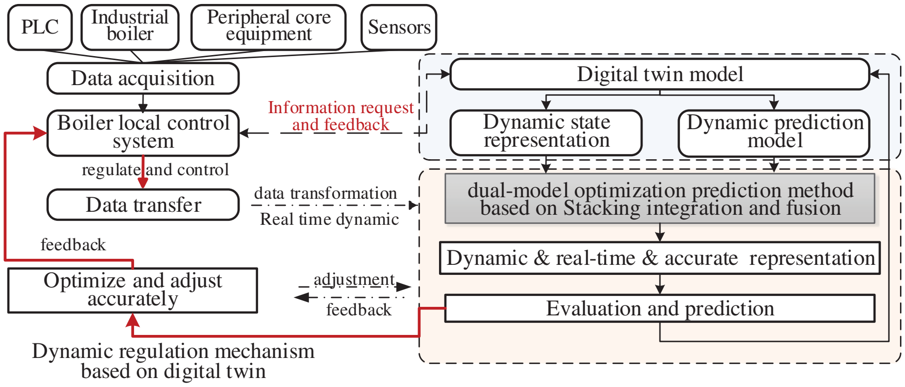

PLC
Industrial boiler
Peripheral core equipment
Sensors
Data acquisition
Boiler local control system
Information request and feedback
regulate and control
Data transfer
data transformation
Real time dynamic
feedback
Optimize and adjust accurately
adjustment
feedback
Dynamic regulation mechanism based on digital twin
Digital twin model
Dynamic state representation
Dynamic prediction model
dual-model optimization prediction method based on Stacking integration and fusion
Dynamic & real-time & accurate representation
Evaluation and prediction

Figure 6: Dynamic regulation of the whole system

DT-IIoT optimization model uses the proposed dual-model integrated optimization algorithm to establish an evolution prediction and evaluation system. Real-time samples and prediction deviation are introduced into DT dynamic representation by dynamic regulation mechanism. Based on multi-sensor historical samples, the state characteristics of the target equipment are obtained. The historical data used in this paper are the actual operation data of the boiler and peripheral components from 2021 to 2022. The local IIoT systems acquire the long-term operational data that is collected by sensors in a 10 min acquisition cycle. Multiple rounds of strong screening and effective dimension reduction are carried out for discrete data. The optimal applicability model is obtained through the dual-model integration optimization to realize the prediction and evolution evaluation of steam quantity. The real-time status display of each data collection point is shown in Fig. 5d. The data set covers the target predicted value, which is steam quantity, and other 37 operating parameters that affect steam production. This paper uses a set of collected data with a total of 2886 × 38 data features. There are many unstable factors in the data, such as missing values, abnormal values, and noise interference data, which challenge the accuracy of steam quantity prediction.

4.2 Discrete and Unstable Data Screening

The DT-IIoT optimization model makes strong screening, reduction, and fusion processing for discrete and unstable data. The main research results are as follows:

(1) Screening of Original Oscillation Data

A box plot is used to plan the distribution of the original data obtained in the data set, as shown in Fig. 7. The comprehensive distribution of all original data is relatively uniform, but it still contains data values with large deviations, such as V9, V25 and V16. Therefore, it is necessary to delete the deviation information of the multi-sensor data that has whole sequence features, and to reduce the dimension of complex features.

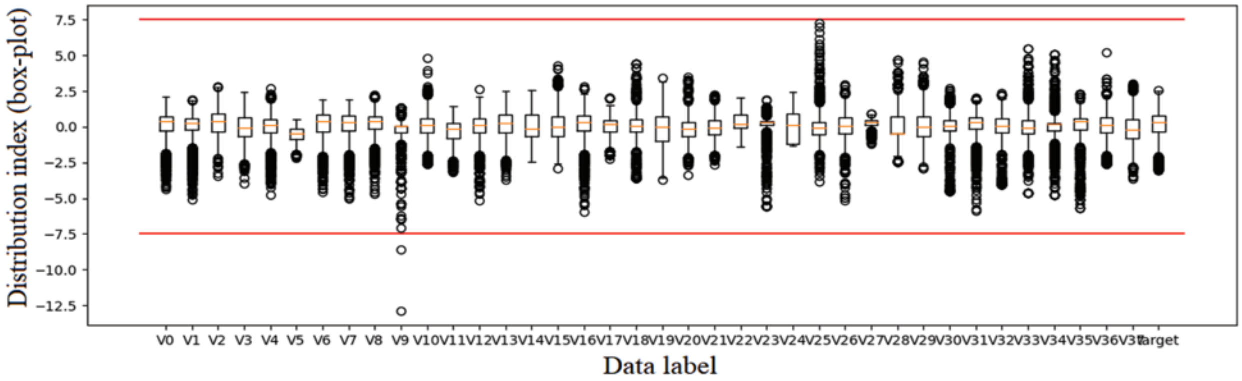
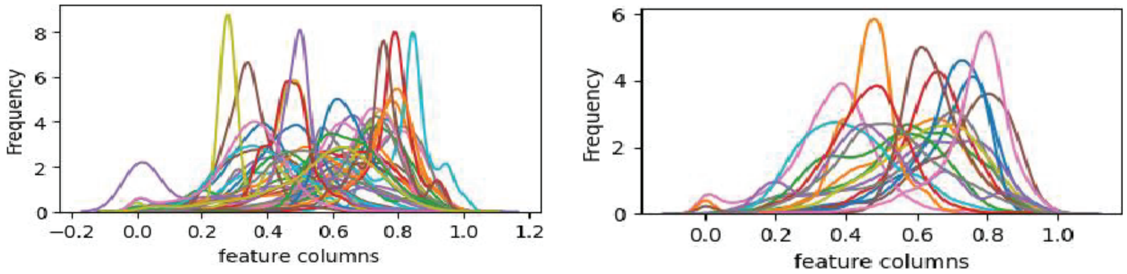


Figure 7: Data dispersion and distribution characteristics

The 37 data features are linked with the steam quantity target value. More stable data types are obtained by fusion screening of Spearman correlation coefficient and multicollinearity test. Kernel Density Estimation (KDE) for each feature is finally obtained, as shown in Fig. 8.



(a) Feature distribution (after preliminary screening) (b) Featuredistribution (after multicollinearity test)

Figure 8: Data screening and feature distribution

The original screening characteristics are shown in Fig. 8a. Each feature deviates from the core region in an oscillatory manner between (0.3–0.9). The overall stability of the distribution is poor. After strong screening, a total of 18 features are obtained. The amplitude of the highest occurrence frequency decreases by 20%, and the core region is between (0.4~0.8). The overall characteristic data is relatively more stable. After two rounds of data screening, the data evolution results obtained by the DT system are shown in Table 1.

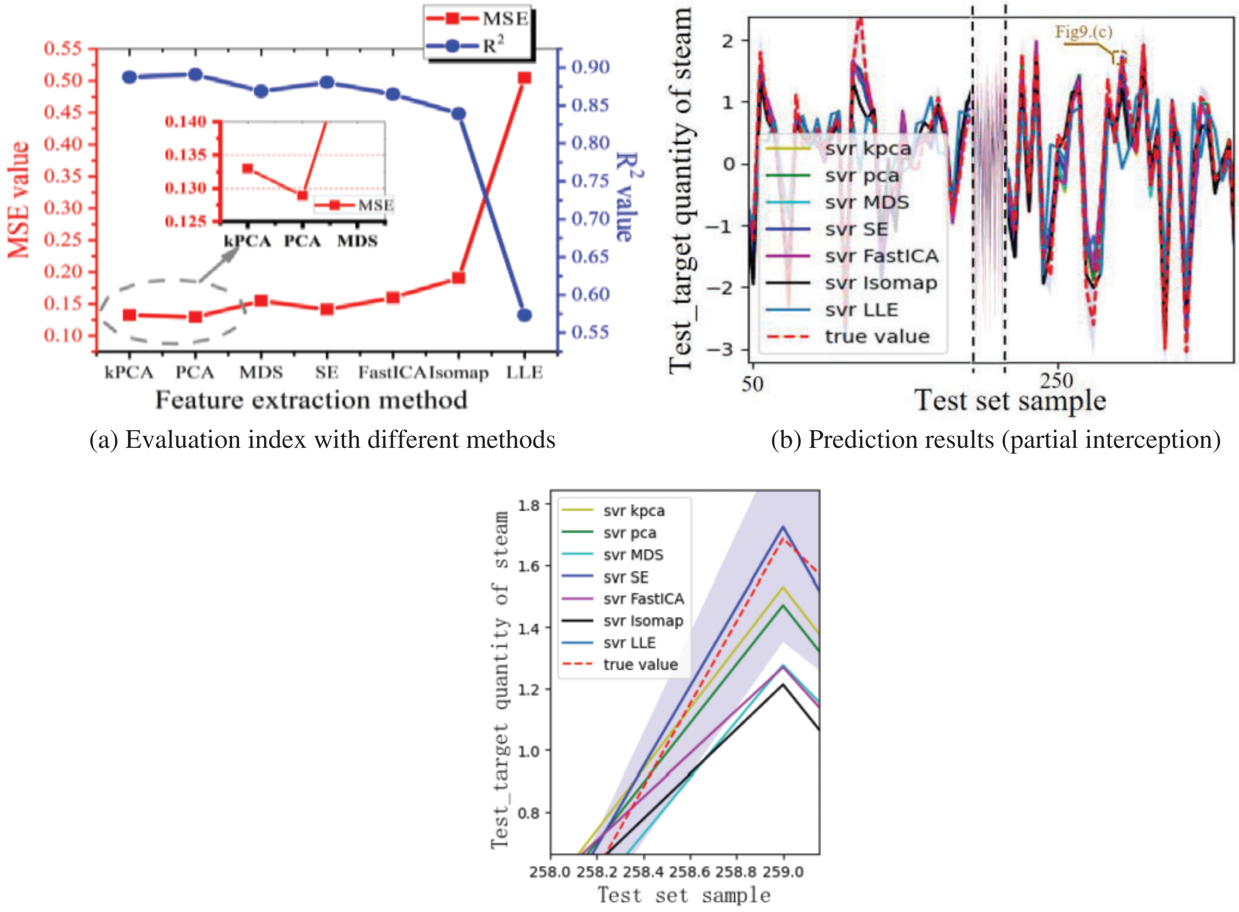
After the original oscillation data screening and the complex feature data dimension reduction screening, the accuracy of the DT data representation results is effectively improved. The effective self-improvement ability reaches 6.61% and 15.13%, respectively, and the accuracy rate increases from 85.8% to 89.11%, which has a certain improvement effect.

Table 1: Comparison of data representation accuracy before and after data instability processing

Sample size	Data representation index considering initial instability	
	MSE value	Accuracy of digital twin data representation
Sample of original discrete data	0.1582	0.858
Stable full data sample	0.1484	0.869
Sample after complexity dimension reduction	0.1289	0.8911

(2) Adaptability Comparison of Feature Extraction Methods

Seven typical algorithms (PCA, ICA, MDS, KPCA, ISOMAP, LLE and SE) are used to test the adaptability of different methods to discrete data representation. The results are shown in Fig. 9:



(a) Evaluation index with different methods

(b) Prediction results (partial interception)

(c) Vertex trend capture in Fig. 9(b)

Figure 9: Comparison of prediction results with different feature extraction algorithms

PCA method has the best comprehensive performance, and the fitting degree is 0.8911. KPCA, MDS, SE and FastICA have similar fitness levels, and the fit degree fluctuated around 0.87, as shown in Fig. 9a. The deviations between each method and PCA are shown in Table 2. Compared with PCA, the MSE errors of ISOMAP and LLE algorithms are 32.23% and 74.46%, respectively, which are not suitable for the screening of this sample. The Manifold Dimension Reduction method has poor adaptability to noisy data.

Table 2: Accuracy evaluation with different feature dimension reduction methods

Implement deviation evaluation indicators	kPCA	PCA	MDS	Spectral embedding	FastICA	Isomap	Locally linear embedding
MES	0.1330	0.1289	0.1549	0.1412	0.1597	0.1902	0.5047
R^2	0.8876	0.8911	0.8691	0.8807	0.8650	0.8392	0.5734
Relative deviation rate (based on MSE value)	3.06%	0%	16.79%	8.68%	19.28%	32.23%	74.46%

Part of the fitting region is randomly selected, and the range of bandwidth ±20% is used to fit the adaptive results (blue bandwidth region, as shown in Fig. 9c). The results show that KPCA, PCA-SVR, MDS and SE can capture the vertex features well, while FastICA, ISOMAP and LLE have poor anti-distortion ability and weak adaptive ability.

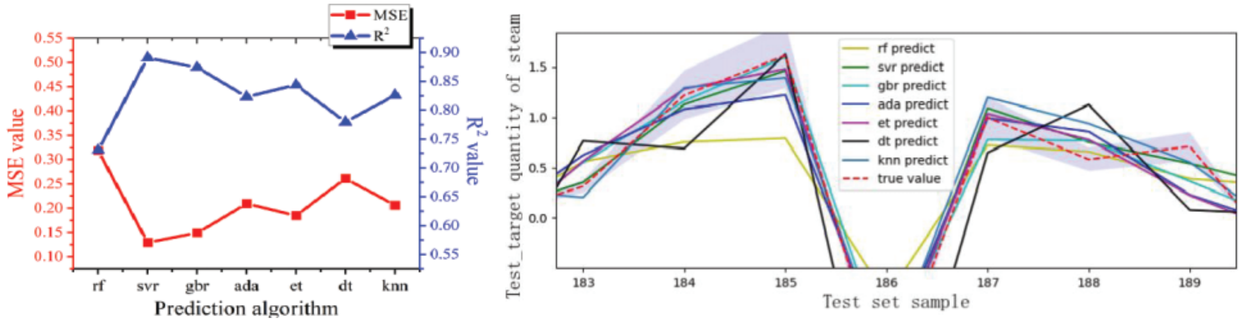
4.3 Dual-Model Stacking Integration Model Optimization

The multi-round selection and optimization of the dual model are key in improving the accuracy of the prediction model. The optimal lightweight model selection results are as follows:

4.3.1 Selection and Optimization of Base Model

Seven classical single algorithms are selected for comparison selection and hyperparameter optimization, and the base model with the highest adaptability is obtained. The prediction results of seven prediction algorithms (RF, GBR, SVR, Ada Boost, ET, dt and KNN) are fitted and compared with the XGBoost method with strong adaptability.

The results show that the fitting accuracy of SVR is higher than that of other models. The relative error loss values of other algorithms relative to SVR reach 146.97%, 15.57%, 8.68%, 62.21%, 43.33% and 59.54%, respectively. RF has poor predictive performance; XGBoost with strong adaptability has a higher fitting degree of prediction results, and its R^2 value reaches 0.8796, which is more significant than most single-fitting models. SVR performance (with $R^2 = 0.8911$) is more prominent, as shown in Fig. 10a and Table 3. Part of the fitting region is randomly selected, and the range of bandwidth ±20% is used to fit the adaptive results (blue bandwidth region, as shown in Fig. 10b). The results show that the prediction results of the initial base model cannot accurately capture the important features of the data, and have a large deviation from the actual situation, which needs to be further improved.



(a) Comparison of evaluation index

(b) Comparison of prediction results in sub-regions

Figure 10: Prediction results of the initial base model

Table 3: Comparison of base models pre-selection results

General models	RF	SVR	GBR	Ada	ET	dt	KNN	XGB
MSE	0.3183	0.1289	0.1490	0.2091	0.1848	0.2611	0.2057	0.1425
R^2	0.7309	0.8911	0.8741	0.8233	0.8438	0.7793	0.8262	0.8796
Relative deviation rate (based on MSE value)	146.97%	0%	15.57%	8.68%	62.21%	43.33%	59.54%	10.55%

Based on the above analysis results, the results of second-round hyperparameter tuning of each model are compared. Using 5 K Fold cross-validation, adaptive real-time deviation data adjustment is introduced to realize real-time optimization and improvement based on dynamic hyperparameter adjustment. The results are shown in Table 4.

Table 4: Comparison of optimization and self-promotion ability of base models

Base model	Second optimization mode		MSE	$\mathbf{R^2}$	Self-promotion (based on MSE value)
RF	Before optimization	Gen model	0.318	0.731	31.76%
	After optimization	Opt model	0.217	0.817	
SVR	Before optimization	Gen model	0.129	0.891	17.05%
	After optimization	Opt model	0.107	0.906	
GBR	Before optimization	Gen model	0.149	0.874	14.77%
	After optimization	Opt model	0.127	0.893	
Ada	Before optimization	Gen model	0.209	0.823	2.87%
	After optimization	Opt model	0.203	0.829	
KNN	Before optimization	Gen model	0.206	0.826	5.34%
	After optimization	Opt model	0.195	0.836	
XGB	Before optimization	Gen model	0.142	0.879	9.86%
	After optimization	Opt model	0.128	0.892	

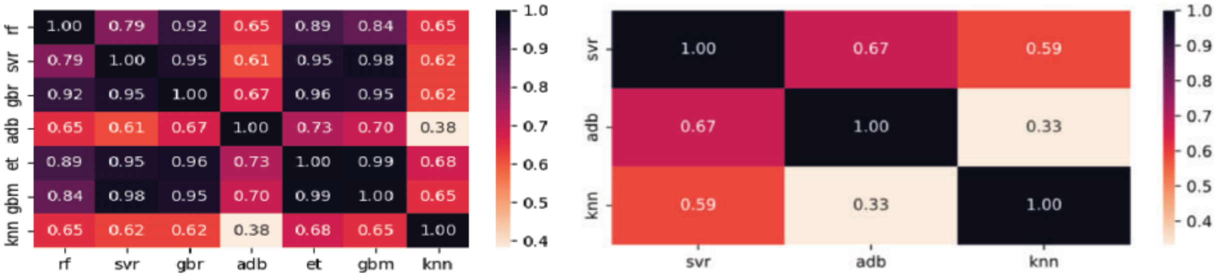
After adaptive adjustment and hyperparameter optimization, the self-improvement ability of each model is improved. The RF model has the strongest self-improvement ability, with an improvement rate of 31.76%. Ada Boost's adjustment ability is limited and is only increased by 2.87%. The average self-improvement of models is more than 10%. The XGB model combined with strong generalization ability also gets a good promotion effect. Overall, SVR is the most effective option in base model optimization.

4.3.2 Stacking Multi-Model Integration Optimization

The real-time analysis of massive industrial data requires the system to be lightweight. The lightweight of prediction model is the key to the dynamic regulation mechanism. In this section, the advantage model obtained in Section 4.3.1 will be integrated and optimized again to obtain a lightweight dual-model optimization system. Finally, a double improvement of the prediction algorithm is realized.

(1) The Establishment of Lightweight Integrated Algorithm Architecture

The optimized SVR OPT model is used as a meta-model, and 7 different base models are combined and optimized to obtain correlation and error analysis. The results are shown in Fig. 11a. The difference among SVR, Ada and KNN is the largest and the correlation of error distribution is the lowest, but the richness of the model will be lost, as shown in Fig. 11b. RF, GBR, ET and XGBoost are all tree-based models, and since similarity is too high, the combination is not conducive to the structural optimization of the prediction model. The data representation effects of different integration combinations are shown in Table 5. After deleting the base model with high similarity, the accuracy of data representation is improved. Compared with the accuracy of data representation, the accuracy of five models' combination reaches 0.9065, and showed stronger advantages. The precision of a three-model combination is slightly lower due to the decrease in richness.



(a) Predictioncorrelation with 7 base models (b) Prediction correlation with simplified model (3 base models)

Figure 11: Prediction error correlation results with the base model

To ensure the accuracy of the modeling, the data used for the first modeling is the operation parameters of the boiler and its auxiliary equipment within one year. A lightweight model greatly reduces the data processing time of the system. The establishment of lightweight model is of great significance for the first large-scale computation. When 7 basic models are reduced to 5 basic models, the running time is reduced by 132.81% from 243.12 to 104.43 s. After 5 base models are reduced to 3 base models, their running time will be reduced again by 67.74% from 104.43 to 62.39 s, as shown in Table 5.

Table 5: Systematic variance analysis of data representation

Base model (Opt models)	Stacking metamodel	Loss MSE	$\mathbf{R}^2$	Run time/s
RF+SVR+GBR+ADA+ET+GBM+KNN	SVR cv-opt	0.1116	0.9037	243.1245
SVR+GBR+ADA+ET+GBM+KNN	SVR cv-opt	0.1117	0.9053	195.4269
SVR+GBR+ADA+ET+KNN	SVR cv-opt	0.1106	0.9065	104.4317
SVR+ADA+ET+KNN	SVR cv-opt	0.1094	0.9059	94.6459
SVR+ADA+KNN	SVR cv-opt	0.1108	0.9063	62.3906

The optimal lightweight model obtained from the first large-scale calculation is brought into the single real-time data again, which will effectively improve the stability of the analysis model.

(2) Stacking Fusion Construction and Optimization

Based on the effectiveness of SVR for discrete data representation, automatic tuning of the SVR meta-model is continued in stackingII. PSO (particle swarm optimization), GA (genetic algorithm) and GWO (Gray Wolf Optimizer) are used for real-time automatic optimization of hyperparameters, which makes the system obtain a higher representation effect. Empirical hyperparameter tuning and XGB self-tuning optimization [30] are used to conduct a second-round evaluation of the integrated optimization model. The optimization results are shown in Table 6. In the automatic optimization of SVR series algorithms, the self-lifting ability of SVR-Gwo is the highest, reaching 18.16%. The self-lifting capacity of SVR-PSO is the lowest, with 11.68%. The mean value of the SVR series self-lifting is 15.85%. The two-round optimization has an obvious gain effect.

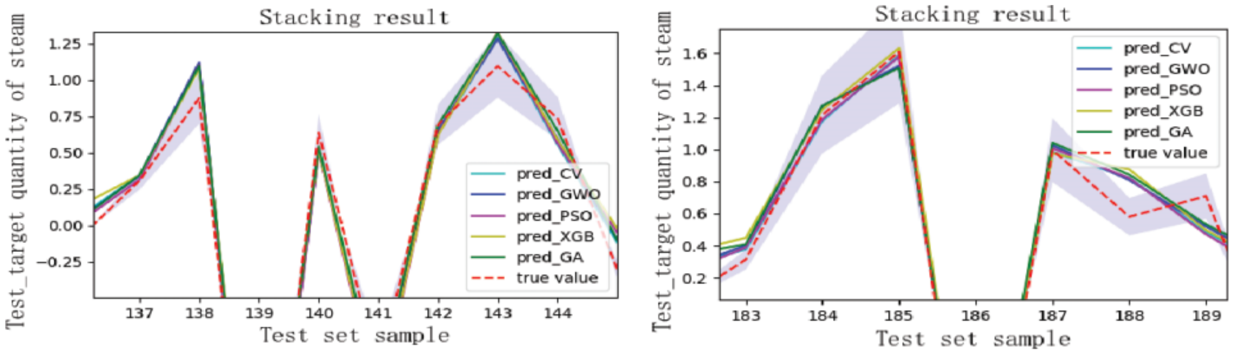
Table 6: Comparison of base-meta dual model optimization

Base model general	Stacking model	MSE	$\mathbf{R}^2$	Base model opt	Stacking model	MSE	$\mathbf{R}^2$	Relative self-lifting general to opt	Self-lifting mean value
SVR+GBR+ ADA+ET+ KNN	SVR-CV	0.1288	0.8912	svr+gbr+ ada+et+knn	SVR-CV	0.1106	0.9065	16.46%	(SVR series optimization) 15.85%
	SVR-PSO	0.1233	0.8958		SVR-PSO	0.1104	0.9067	11.68%	
	SVR-GA	0.1267	0.8929		SVR-GA	0.1082	0.9085	17.1%	
	SVR-Gwo	0.1269	0.8928		SVR-Gwo	0.1074	0.9092	18.16%	
	XGB	0.1412	0.8807		XGB	0.1144	0.9033	23.43%	23.43%

GWO has the best performance in similar optimization, with $R^2 = 0.9092$ after optimization, exceeding the predicted performance of all existing models. The optimization model reduces the loss rate (MSE) from the initial 0.318 to the optimal 0.1074, and increases R^2 from the initial 0.731 to the optimal 0.9092. The accuracy and stability of the model are improved. Stacking integrated optimization effectively improves the adaptability and representation accuracy of the prediction model. The second optimization of the meta-model is an effective improvement of the dual-model integrated optimization structure. The results show that the loss deviation of the secondary optimization is reduced by 20% on average after the dual-mode synchronous self-regulation optimization. The optimization effect of SVR-GWO is the best, the accuracy rate reaches 90.92%, and the self-improvement ability reaches 18.16%. XGBoost, which initially has strong adaptability, has a greater performance improvement after dual-mode secondary tuning, with the largest increase of 23.43%. The strong screening dual-model integrated optimization structure can promote the accurate representation of discrete data.

4.4 *Result Representation*

The representation effect of the system on part of discrete data is randomly selected, and the range of bandwidth ±20% is used to fit the adaptive results (blue bandwidth region, as shown in Fig. 12b).



(a) Partially random region

(b) Partially random region compared with Figure 10(b)

Figure 12: Base-meta dual model optimization predicted results

The results show that the system representation ability is more stable after multi-stage optimization. Compared with the original analysis results in Fig. 10b, all alternatives are within the error margin of 10%. The prediction value is consistent with the actual value, and the capturing ability of data discreteness is effectively enhanced.

5 Conclusions

Digital representation and accurate prediction of key performance for large industrial equipment is an important way to realize the whole life-cycle management for the key equipment. To solve the difficulties in real-time information interaction and dynamic prediction of the DT-IIoT platform, a dynamic regulation mechanism is proposed in the DT system, integrating the IIoT with algorithm optimization. A digital dynamic representation and evolution regulation mechanism driven by industrial big data is established to realize dynamic intelligent regulation. An SSDO prediction method, based on Stacking integration and fusion, is proposed in the dynamic systematic regulation. Taking the performance of key equipment boilers in large power plants as an example, the evolution prediction analysis of boiler steam quantity is realized. The results show that the multi-stage tuning lightweight structure composed of the five-element base model + the improved SVR meta-model can promote the accurate representation of discrete data. The running time is reduced by 132.81%. The loss deviation of the secondary optimization is reduced by 20% on average after the dual-model synchronous self-regulation optimization. Stacking integrated optimization effectively improves the adaptability and representation accuracy of the prediction model. Compared with the original analysis results, all alternatives are within the error margin of 10%. The prediction and representation effect of SVR-GWO optimization is the best, as the accuracy reaches 90.92%, and the self-improvement ability reaches 18.16%. The adaptability and reliability of the model are comprehensively improved, and better prediction and analysis results are achieved.

Funding Statement: This research was funded by: Major Science and Technology Project of Anhui Province (Grant Number: 201903a05020011), Talents Research Fund Project of Hefei University (Grant Number: 20RC14), and the Natural Science Research Project of Anhui Universities (Grant Number: KJ2021A0995). Graduate Student Quality Engineering Project of Hefei University (Grant Number: 2021Yjyxm09). Enterprise Research Project: Research on Robot Intelligent Magnetic Force Recognition and Diagnosis Technology Based on DT and Deep Learning Optimization.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] B. A. Miguel, O. M. Carlos and D. R. Javier, "Data-driven energy prediction modeling for both energy efficiency and maintenance in smart manufacturing systems," *Energy*, vol. 238, no. 8, pp. 121691, 2022.

[2] S. Jamil, M. Rahman and Fawad, "A comprehensive survey of digital twins and federated learning for Industrial Internet of Things (IIoT), Internet of Vehicles (IoV) and Internet of Drones (IoD)," *Applied System Innovation*, vol. 5, pp. 56–71, 2022.

[3] S. Malakuti and S. Gruner, "Architectural aspects of digital twins in IIoT systems," in *Proc. of European Conf. on Software Architecture*, Madrid, Spain, pp. 1–2, 2018.

[4] Y. Lu, C. Liu, K. Wang, H. Huang and X. Xu, "Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues," *Robotics and Computer Integrated Manufacturing*, vol. 61, pp. 101837, 2019.

[5] Y. He, J. Guo and X. Zheng, "From surveillance to digital twin: Challenges and recent advances of signal processing for industrial internet of things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 120–129, 2018.

[6] C. Zhang, G. H. Zhou, J. He, Z. Li and W. Cheng, "A data-and knowledge-driven framework for digital twin manufacturing cell," in *Proc. of CIRP Conf. on Industrial Product-Service Systems*, China, pp. 345–350, 2019.

[7] K. Zhang, T. Qu, D. J. Zhou and H. F. Jiang, "Digital twin-based opti-state control method for a synchronized production operation system," *Robotics and Computer Integrated Manufacturing*, vol. 63, no. 3, pp. 101892, 2020.

[8] W. C. Luo, T. L. Hu and Y. X. Ye, "A hybrid predictive maintenance approach for CNC machine tool driven by digital twin," *Robotics and Computer-Integrated Manufacturing*, vol. 65, no. 1, pp. 101974, 2020.

[9] M. N. Liu, S. L. Fang, H. Y. Dong and C. Z. Xu, "Review of digital twin about concepts, technologies, and industrial applications," *Journal of Manufacturing Systems*, vol. 58, pp. 346–361, 2021.

[10] S. Rikard, K. Wärmefjord, J. S. Carlson and L. Lindkvist, "Toward a digital twin for real-time geometry assurance in individualized production," *CIRP Annals*, vol. 66, no. 1, pp. 137–140, 2017.

[11] C. Gehrmann and M. Gunnarsson, "A digital twin based industrial automation and control system security architecture," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 669–680, 2019.

[12] Y. M. Zhao, L. Li, Y. Liu, Y. X. Fan and K. Y. Lin, "Communication-efficient federated learning for digital twin systems of industrial Internet of Things," *IFAC-PapersOnLine*, vol. 55, no. 2, pp. 433–438, 2022.

[13] A. Niaz, S. Khan, F. Niaz, M. U. Shoukat, I. Niaz *et al.,* "Smart city IoT application for road infrastructure safety and monitoring by using digital twin," in *Proc. of Int. Conf. on IT and Industrial Technologies*, Chiniot, Pakistan, pp. 1–6, 2022.

[14] M. Platenius-Mohr, S. Malakuti and S. Grüner, "Interoperable digital twins in IIoT systems by transformation of information models: A case study with asset administration shell," in *Proc. Int. Conf. on the Internet of Things*, Bilbao, Spain, pp. 1–8, 2019.

[15] J. Cheng, H. Zhang and F. Tao, "DT-II: Digital twin enhanced industrial internet reference framework towards smart manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 62, no. 4, pp. 101881, 2020.

[16] J. Cecil, S. Albuhamood and A. Cecil-Xavier, "An advanced cyber physical framework for micro devices assembly," *IEEE Transactions on Systems Man Cybernetics-systems*, vol. 49, no. 1, pp. 92–106, 2019.

[17] F. Tao and X. Ma, "Research on digital twin sttandard system," *Computer Integrated Manufacturing System*, vol. 25, no. 3, pp. 2405–2418, 2019.

[18] W. Luo, T. Hu, C. Zhang and Y. Wei, "Digital twin for CNC machine tool: Modeling and using strategy," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 3, pp. 1129–1140, 2019.

[19] J. S. Yoon, J. Jordon and M. van der Schaar, "GAIN: Missing data imputation using generative adversarial nets," in *Proc. of Int. Conf. on Machine Learning*, Stockholm, Sweden, pp. 2640–3498, 2018.

[20] Z. Ren, J. Wan and P. Deng, "Machine-learning-driven digital twin for lifecycle management of complex equipment," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 1, pp. 9–22, 2022.

[21] K. Xia, C. Sacco, M. Kirkpatrick, C. Saidy, L. Nguyen *et al.,* "A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment interfaces and intelligence," *Journal of Manufacturing Systems*, vol. 58, no. 3, pp. 210–230, 2021.

[22] X. W. Xiao and F. Wang, "Bionic structure and cyber-physical system for intelligent power plant oriented to the industrial internet," *Transactions of China Electrotechnical Society*, vol. 35, no. 23, pp. 4898–4911, 2020.

[23] A. Rasheed, O. San and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 99, pp. 21980–22012, 2020.

[24] L. Kucera and J. Vachalek, "The digital twin of a measuring process within the Industry 4.0 concept," in *Proc. of Int. Conf. of Machine Design Departments*, Demanovska Dolina, Slovakia, pp. 333–341, 2019.

[25] S. Malakuti, R. Borrison, A. Kotriwala and B. Kloeppe, "An integrated platform for multi-model digital twins," in *Proc. of Int. Conf. on the Internet of Things*, St. Gallen, Switzerland, pp. 9–16, 2022.

[26] G. Li and S. Zhao, "YOLO-RFF: An industrial defect detection method based on expanded field of feeling and feature fusion," *Electronics*, vol. 11, no. 24, pp. 4211, 2022.

[27] H. Li and H. Q. Wang, "Concept, system structure and operating mode of industrial digital twin system," *Computer Integrated Manufacturing System*, vol. 27, no. 12, pp. 3373–3390, 2021.

[28] M. H. Chen, Q. Y. Liu and J. S. Zhang, "XGBoost-based algorithm for post-fault transient stability status prediction," *Power System Technology*, vol. 44, pp. 1026–1033, 2020.

[29] B. Hu and Z. Q. Zhan, "Prediction research on short-term photovoltaic output based on PCA-GA-ELMAN," *Acta Energiae Solaris Sinica*, vol. 41, pp. 256–263, 2020.

[30] J. Q. Yu, W. Q. Jing and A. J. Zhao, "Cold load prediction model based on improved PSO-BP algorithm," *Journal of System Simulation*, vol. 33, pp. 54–61, 2021.