Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.081754

ARTICLE

A Computational Multi-Output Soft Sensing Framework for Sinter Quality Prediction Using Feature Selection and Hierarchical SVR Optimization

Zhenhua Yang[1,2], **Yifan Li**[1,2], **Aimin Yang**[1,2,*], **Jie Li**[2,3] **and Tao Xue**[1,2]

[1]College of Science, North China University of Science and Technology, Tangshan, China
[2]Tangshan Key Laboratory of Engineering Computation, Tangshan, China
[3]College of Metallurgy and Energy, North China University of Science and Technology, Tangshan, China

*Corresponding Author: Aimin Yang. Email: aimin@ncst.edu.cn

Received: 08 March 2026; Accepted: 27 April 2026; Published: 27 May 2026

ABSTRACT: Sinter quality prediction in iron ore sintering is a challenging computational modeling problem because of highly nonlinear process behavior, strong cross-variable interactions, and disturbances caused by changing operating conditions. This study develops a data-driven multi-index soft-sensing framework for sinter quality prediction by combining feature selection and hierarchical model optimization. An improved binary Greylag Goose Optimization algorithm is first employed to identify a compact subset of informative variables, reducing redundancy and multicollinearity in the original process data. A hierarchical two-stage Greylag Goose Optimization strategy is then designed to optimize the hyperparameters of a support vector regression model through coarse-to-fine search, balancing global exploration and local refinement in the parameter space. The proposed framework is validated on three key sinter quality indices under consistent data partitioning and equal optimization budgets. Experimental results show that the method achieves coefficients of determination of 0.975, 0.985, and 0.986 for yield, drum index, and $RDI_{+3.15}$, respectively, indicating strong predictive capability and robust generalization. Comparative experiments demonstrate that the proposed framework outperforms several representative baseline methods in terms of prediction accuracy and fitting performance. In addition, ablation analysis confirms the contribution of the hierarchical optimization mechanism to the overall model performance. The proposed framework offers an effective computational approach for multi-index quality modeling, online prediction, and intelligent decision support in complex industrial systems.

KEYWORDS: Data-driven modeling; multi-index quality prediction; support vector regression; feature selection; Greylag Goose Optimization; hierarchical optimization

1 Introduction

Sinter quality is a fundamental prerequisite for the stable operation of blast furnaces and for low-carbon and high-efficiency ironmaking. Indicators such as sinter yield, the tumble strength index, and the reduction degradation index directly reflect the output level, mechanical strength, and high-temperature metallurgical performance of the sintering process [1]. Because the sintering process involves the coupling of multiple factors, including raw material particle size distribution, water addition and mixing state, bed structure, ignition and suction regimes, and chemical composition [2], the process mechanism shows strong nonlinearity, strong disturbances, and multi-source uncertainty [3]. As a result, quality indices are prone to fluctuations in industrial production and are difficult to predict with high accuracy. Therefore, developing a modeling method oriented to real production data and capable of achieving high-accuracy prediction for

multiple quality indices has important engineering significance for optimized control of the sintering process and for quality stabilization [4].

Research on online prediction of key variables in the sintering process can generally be divided into two categories: soft sensing of product quality and composition, and prediction or stabilization of key process states. Existing studies have explored a variety of data-driven methods. Traditional machine learning and ensemble learning methods have been applied to the prediction of chemical composition and quality-related variables [5–8]. Deep learning and temporal modeling approaches, including recurrent networks, convolutional fusion models, and Transformer-based architectures, have further improved the prediction of FeO, basicity, TFe, and burn-through-point-related variables under complex dynamic conditions [9–13]. These methods have been extended and validated across different sinter quality indicators and modeling frameworks, demonstrating strong generalization for industrial time-series prediction tasks [14,15]. Graph-based learning and multi-source fusion strategies have also been introduced to capture variable coupling, delayed dependencies, and complementary information from heterogeneous data sources [16–19]. In addition, condition identification, condition-specific modeling, and prediction–control integration have been investigated to address operating-condition variability and process stabilization requirements [20–24]. More recently, joint prediction, uncertainty-aware learning, and optimization-based decision models have been explored for multi-index prediction and production-oriented decision support [25–27]. For physical and strength-related indices, feature-enhanced intelligent models and microstructure-assisted analysis have also been reported [28,29]. These studies demonstrate the potential of intelligent modeling for sintering systems; however, most existing methods still focus on specific variables, particular operating conditions, or individual modeling strategies, leaving room for further improvement in unified multi-index quality prediction.

Despite recent advances, multi-index sinter quality prediction remains challenging because existing studies often focus on a single variable or a specific operating scenario. In addition, sintering data are typically high-dimensional, strongly collinear, and affected by frequent operating-condition disturbances, while model performance is highly sensitive to hyperparameter settings. To address these challenges, this study proposes a bGGO-TSGGO-SVR framework for multi-index sinter quality prediction. The bGGO (Improved Binary Greylag Goose Optimization) is used for feature selection, and TSGGO (Two-Stage Greylag Goose Optimization) is adopted for coarse-to-fine hyperparameter tuning of support vector regression. These two components form a unified prediction framework for multiple sinter quality indices. The proposed method provides an effective data-driven solution for multi-index quality prediction and process optimization in complex industrial processes.

2 Data Processing

2.1 Data Description

The dataset used in this study was obtained from real sintering production trials, containing a total of 2000 samples. Each sample corresponds to a complete sintering process, and each record consists of 17 process and composition-related input variables and three quality output indices. The input variables include information on the size distribution of the blended mixture, thermal operation conditions and negative pressure, bed structure parameters, and major chemical compositions. These variables provide a relatively comprehensive representation of key factors that influence quality formation during sintering. The output variables are three representative quality indices, namely Product Yield, Drum Index, and $RDI_{+3.15}$ (Reduction Degradation $Index_{+3.15}$), which reflect the productivity, mechanical strength, and resistance to reduction degradation of the sinter, respectively. Table 1 lists all variables in the dataset. The first 17 parameters are inputs, and variables 18 to 20 are outputs.

Table 1: Data description.

Index	Parameter Description
1	Water addition ratio (percent)
2–6	Particle size of mixed materials, including less than 3 mm, 3 to 5 mm, 5 to 8 mm, greater than 8 mm, and the average particle size (percent)
7	Ignition temperature (°C)
8	Ignition negative pressure (kPa)
9	Sintering negative pressure (kPa)
10	Layer thickness (mm)
11	Sintering time (min)
12–16	TFe, CaO, MgO, SiO_2, and Al_2O_3 in sinter (percent)
17	Moisture (percent)
18	Product yield (percent)
19	Drum index (percent)
20	$RDI_{+3.15}$ (percent)

2.2 Data Cleaning

Because the data were collected from real sintering production trials, no missing values were observed. To address potential extreme disturbances and abnormal sampling points in industrial data, this study employed the box plot method to identify and handle outliers [30]. The box plot characterizes the data distribution using quantile statistics and detects outliers based on the interquartile range as the scale measure. Let Q_1 and Q_3 denote the first and third quartiles of a given variable, respectively. The interquartile range is defined as follows:

$$IQR = Q_3 - Q_1, \tag{1}$$

Based on the interquartile range, the lower and upper thresholds are defined as follows:

$$L = Q_1 - 1.5IQR, \tag{2}$$

$$U = Q_3 + 1.5IQR, \tag{3}$$

A sample value is considered an outlier if it satisfies $x < L$ or $x > U$. Box plots were generated for all variables to identify outliers, and the results are shown in Fig. 1. This method enables effective screening of outliers without relying on distributional assumptions, thereby reducing the influence of a small number of abnormal observations on model training and evaluation results. The identified outliers were replaced with the mean values of the respective variables to minimize their impact on the analysis and maintain the integrity of the dataset.

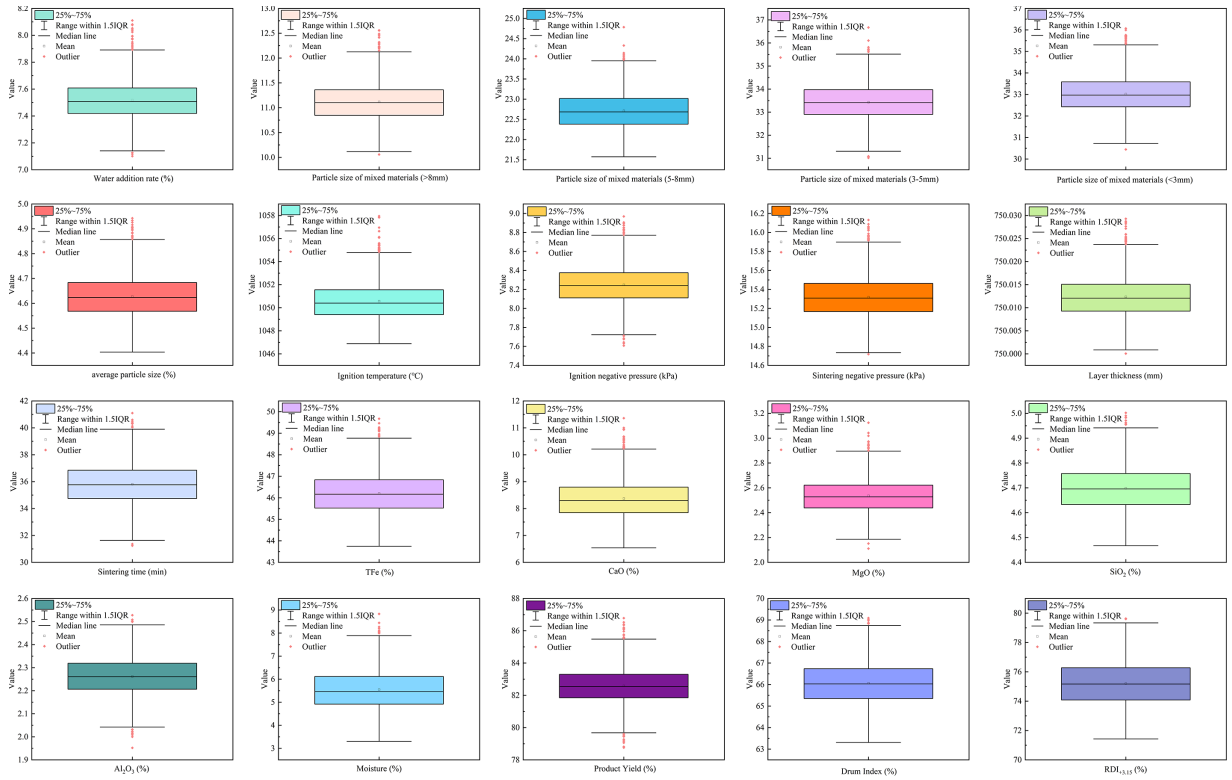

Figure 1: Box plots of all variables.

After removing outliers, the input features were standardized to eliminate the adverse effects of differences in variable scales on model training. Let the mean and standard deviation of the *j*th feature in the training set be μ_j and σ_j, respectively. The standardization is performed as follows:

$$x'_{ij} = \left(x_{ij} - \mu_j\right)/\sigma_j, \tag{4}$$

3 Model Development

To predict three quality indices in the sintering process, namely product yield, drum index, and $RDI_{+3.15}$, this study develops a bGGO-TSGGO-SVR model. The model uses support vector regression as the base predictor. Feature subset selection is performed using the bGGO optimization algorithm, and the SVR hyperparameters are tuned via gray goose optimization in a hierarchical manner with refined search, thereby improving prediction accuracy and generalization performance. The overall procedure consists of five stages, including data preprocessing, feature selection, stage one global coarse search, stage two local fine search, and final model training and prediction. To clearly illustrate the input-output relationships and execution order of each module, the model architecture is presented in Fig. 2. The model development procedure is described in detail below, organized by module.

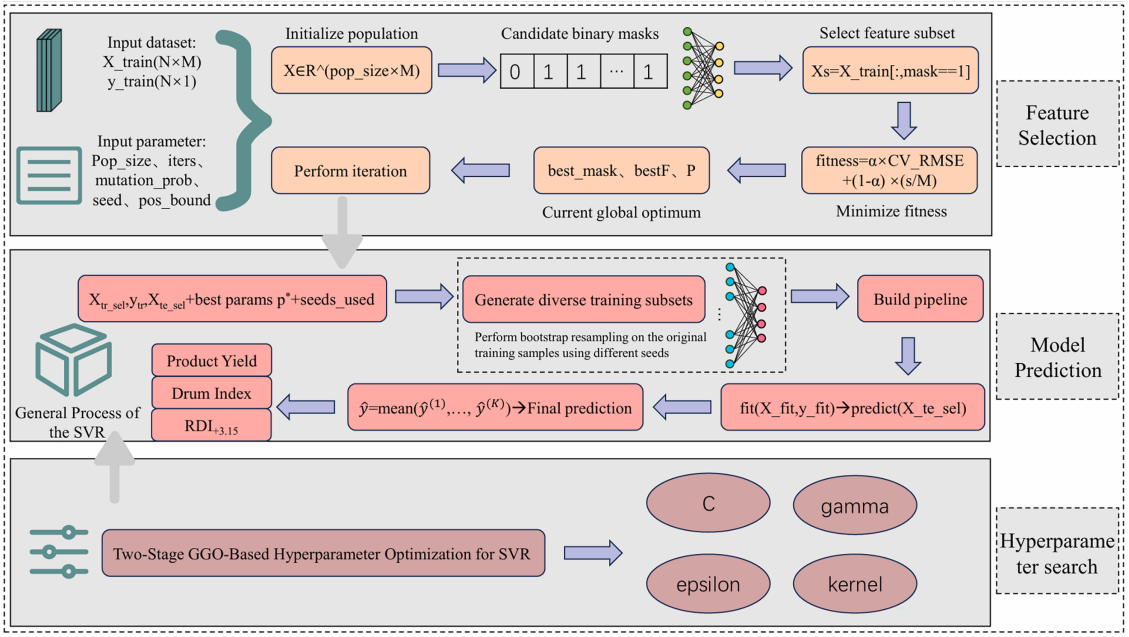

Figure 2: Overall execution flow of the model.

3.1 Feature Selection

The variables in the sintering process exhibit strong coupling and redundancy. Directly using all features as inputs can introduce noise and degrade the model's generalization capability. Therefore, based on Greylag Goose Optimization, this study develops a binary gray goose optimization algorithm, bGGO, for feature subset search [31]. By searching for an optimal feature combination in a binary space, bGGO enables adaptive screening of input variables.

The training dataset is denoted as $X \in \mathbb{R}^{N\times d}$, where N is the number of samples and d is the number of candidate features. The selected feature subset is represented by a binary vector $M = [m_1, m_2, \ldots, m_d]$, which satisfies constraint $m_j \in \{0,1\}$. If $m_j = 1$, the jth feature is selected; otherwise, it is removed. The selected feature subset can be expressed as follows:

$$S = \{j | m_j = 1\}, \quad |S| = \sum_{j=1}^{d} m_j, \tag{5}$$

Since swarm intelligence algorithms update individual positions in a continuous space, a Sigmoid function is used to map the continuous position vector $Z \in \mathbb{R}^d$ to a feature selection probability vector $p \in (0,1)^d$, as shown in:

$$p_j = \sigma(z_j) = 1/(1 + \exp(-z_j)), \tag{6}$$

A binary mask is then generated via random sampling, as shown in:

$$m_j = \begin{cases} 1, u_j < p_j \\ 0, u_j > p_j \end{cases}, \quad u_j \sim U(0,1), \tag{7}$$

To ensure feasibility, if $\sum_{j=1}^{d} m_j = 0$, the feature with the highest selection probability is forced to be selected. Furthermore, to satisfy the minimum feature number constraint $d_{\min}$, if $|S| < d_{\min}$, unselected features are added in descending order of p_j until the constraint is met.

The objective of feature selection is to reduce the feature dimensionality as much as possible while maintaining predictive accuracy. The fitness function is defined as a weighted sum of an error term and a feature size penalty term, as given in:

$$F(M) = \alpha \cdot RMSE_{cv}(X_S, y) + (1-\alpha) \cdot (|S|)/d, \tag{8}$$

Here, $\alpha \in [0,1]$ is a trade-off coefficient, y is the ground truth vector for the target quality index, and X_s denotes the input matrix that retains only the columns corresponding to subset S. The error term is evaluated using the k fold cross validation root mean squared error, as defined in:

$$RMSE_{cv}(X_S, y) = 1/k \times \sum_{k=1}^{K} \sqrt{\frac{1}{|v_k|} \sum_{n \in v_k} \left(y_n - \hat{y}_n^{(k)}\right)^2}, \tag{9}$$

v_k denotes the validation set of the k fold, $\hat{y}_n^{(k)}$ is the prediction for sample n produced by the model trained in the k fold. By minimizing the fitness value, the optimal feature mask M' and the corresponding feature subset is obtained. The selected subset is then used as the input for the subsequent two-stage hyperparameter optimization and prediction modeling.

3.2 Two Stage Hyperparameter Optimization

The predictive performance of support vector regression is highly sensitive to hyperparameter settings. To avoid reliance on empirical choices and to improve optimization efficiency, after the feature subset is determined, this study employs TSGGO to conduct a hierarchical search for SVR hyperparameters. This design enables the optimization process to achieve both global exploration capability and fine-grained local refinement.

In stage one, a global coarse search is performed. The continuous hyperparameters to be optimized are represented as a vector $v = [v_1, v_2, \ldots, v_n]$, and a logarithmic mapping is used to decode them into the actual SVR hyperparameters. The logarithmic scale improves comparability across parameters of different magnitudes and enhances search stability, allowing the swarm-based search to update in a more balanced space. Stage one minimizes the cross-validation error to obtain a candidate solution with strong generalization performance, which serves as the initialization and center for the subsequent fine search.

In stage two, local fine-tuning is conducted within the neighborhood of the stage one optimum. Let v' denote the optimal vector obtained in stage one. Stage two constructs a contracted search region centered at v' and continues the gray goose optimization iterations within this local space, thereby refining the hyperparameter combination and reducing uncertainty caused by random fluctuations. Compared with stage one, stage two operates in a more focused search region, which enables more precise optimization with fewer evaluations. Both stages use the same objective function, with the k-fold cross-validation root mean squared error as the evaluation metric. Let a hyperparameter configuration be denoted as θ, and let the corresponding cross-validation error be $RMSE_{cv}(\theta)$. The final output is the optimal hyperparameter configuration obtained in stage two, which is then used to build the SVR prediction model for the target quality index.

3.3 Construction of the SVR Predictor

After feature subset selection and two-stage hyperparameter optimization, support vector regression is used to model the nonlinear mapping between sintering process variables and quality indices. For a given quality index, the input vector is $x \in \mathbb{R}^{|s'|}$, and the SVR regression function is expressed as follows:

$$\hat{y} = f(x) = \sum_{i=1}^{n_s} \left(\alpha_i - \alpha_i^{'}\right) K(x_i, x) + b, \tag{10}$$

x_i denotes a support vector, α_i and $\alpha_i^{'}$ are Lagrange multipliers, b is the bias term, n_s is the number of support vectors, and $K(,)$ is the kernel function. To capture the nonlinear relationship between process variables and quality indices, both the radial basis function kernel and the linear kernel are considered. The kernel type is selected adaptively through the two-stage optimization procedure.

When the radial basis function kernel is used, the kernel is defined as follows:

$$K(x_i, x) = \exp\left(-\gamma \, ||x_i - x||^2\right) \tag{11}$$

When the linear kernel is used, the kernel is defined as follows:

$$K(x_i, x) = x_i^T x \tag{12}$$

Here, γ is the kernel parameter that controls the local response range of the radial basis function kernel. Through the kernel mapping, SVR can learn complex nonlinear relationships in a high-dimensional feature space, making it suitable for quality prediction tasks in the sintering process where multivariable coupling is present.

For the multi output prediction problem, a target wise modeling strategy is adopted, in which independent SVR predictors are constructed for product yield, drum index, and $RDI_{+3.15}$, respectively. For the jth quality index, the prediction function $f_j()$ is trained using the optimal feature subset $S_j^{'}$ obtained from feature selection and the optimal hyperparameter configuration $\theta_j^{'}$ obtained from two stage optimization. For any input sample, the predicted values of the three quality indices are given by:

$$\hat{y}_j = f_j\left(xs_j^{'}\right), \quad j \in \{1, 2, 3\} \tag{13}$$

4 Experiments

4.1 Evaluation Metrics

To comprehensively evaluate the prediction performance of the proposed model on sinter quality indices, four metrics are adopted, including root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). RMSE is more sensitive to large errors, MAE reflects the average absolute deviation, MAPE describes the relative error level, and R^2 measures the proportion of variance in the data that is explained by the model. The test set contains n samples, For the ith sample, the ground truth and predicted values are denoted as y_i and $\hat{y}_i$. The metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{14}$$

$$MAE = 1/n \times \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{15}$$

$$MAPE = 100\%/n \times \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \tag{16}$$

$$R^2 = 1 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n} (y_i - \overline{y})^2 \tag{17}$$

Here, $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ is the mean of the ground truth values. For RMSE, MAE, and MAPE, smaller values indicate lower prediction errors. For R^2, a value closer to 1 indicates better fitting and generalization performance.

4.2 Model Comparison

To validate the effectiveness of the proposed bGGO-TSGGO-SVR model for predicting sinter quality indices, five representative state-of-the-art methods were selected as benchmarks, including AST-MLP [32], DBO-XGBoost [33], GSA-CatBoost [34], PSO-LightGBM [35], and POA-SVR [36]. All models were evaluated on the same dataset using an identical data split, with the same training and test samples. The data was randomly divided into training, validation, and test sets to ensure independence between these subsets and to prevent any potential data leakage. For the benchmark models that incorporate swarm intelligence optimization, as well as the proposed method, the same optimization parameters were applied. Specifically, the population size was set to 12, and the maximum number of iterations was set to 30 for all models. These consistent settings ensured that the search budget and iteration settings were unified, allowing for a fair comparison of model performance. Performance evaluation was conducted under the same metric system to ensure a fair comparison. The prediction results of all models on the three quality indices, namely Product Yield, Drum Index, and $RDI_{+3.15}$, are summarized in Table 2.

Table 2: Comparison of experimental results.

Prediction Metric	Model	RMSE	MAE	MAPE	R^2
Product Yield	AST-MLP	0.459	0.361	0.443	0.826
	DBO-XGBoost	0.374	0.304	0.372	0.885
	GSA-CatBoost	0.324	0.256	0.315	0.913
	PSO-LightGBM	0.347	0.275	0.331	0.901
	POA-SVR	0.258	0.179	0.202	0.956
	bGGO–TSGGO–SVR	0.175	0.140	0.178	0.975
Drum Index	AST-MLP	0.223	0.172	0.265	0.949
	DBO-XGBoost	0.229	0.175	0.274	0.946
	GSA-CatBoost	0.176	0.138	0.216	0.968
	PSO-LightGBM	0.192	0.146	0.226	0.962
	POA-SVR	0.165	0.129	0.182	0.969
	bGGO–TSGGO–SVR	0.122	0.097	0.157	0.985

(Continued)

Table 2 (continued)

Prediction Metric	Model	RMSE	MAE	MAPE	R^2
$Rdi_{+3.15}$	AST-MLP	0.419	0.327	0.437	0.927
	DBO-XGBoost	0.384	0.297	0.396	0.930
	GSA-CatBoost	0.241	0.185	0.253	0.973
	PSO-LightGBM	0.290	0.223	0.303	0.960
	POA-SVR	0.195	0.154	0.192	0.974
	bGGO–TSGGO–SVR	0.179	0.139	0.181	0.986

Table 2 and Fig. 3 compare different models on three quality indices: Product Yield, Drum Index, and $RDI_{+3.15}$. bGGO-TSGGO-SVR achieves the lowest RMSE, MAE, and MAPE and the highest R^2 across all tasks, indicating superior error control and fitting consistency. Relative to the best baseline, POA-SVR, it reduces Product Yield RMSE from 0.258 to 0.175, a 32.17 percent decrease. MAE drops from 0.179 to 0.140, a 21.79 percent decrease, and MAPE drops from 0.202 to 0.178, an 11.88 percent decrease. R^2 rises from 0.956 to 0.975, an absolute gain of 0.019. For Drum Index, RMSE falls from 0.165 to 0.122, a 26.06 percent decrease. MAE falls from 0.129 to 0.097, a 24.81 percent decrease, and MAPE falls from 0.182 to 0.157, a 13.74 percent decrease. R^2 increases from 0.969 to 0.985, an absolute gain of 0.016. For $RDI_{+3.15}$, RMSE decreases from 0.195 to 0.179, a 8.21 percent decrease. MAE decreases from 0.154 to 0.139, a 9.74 percent decrease, and MAPE decreases from 0.192 to 0.181, a 5.73 percent decrease. R^2 increases from 0.974 to 0.986, an absolute gain of 0.012. Overall, AST-MLP shows larger errors, while ensemble tree methods reduce errors but remain inferior to SVR based approaches. Combining bGGO feature selection with TSGGO hyperparameter optimization yields lower errors and higher coefficients of determination under identical data splits and iteration budgets, confirming robustness for multi index sinter quality prediction.

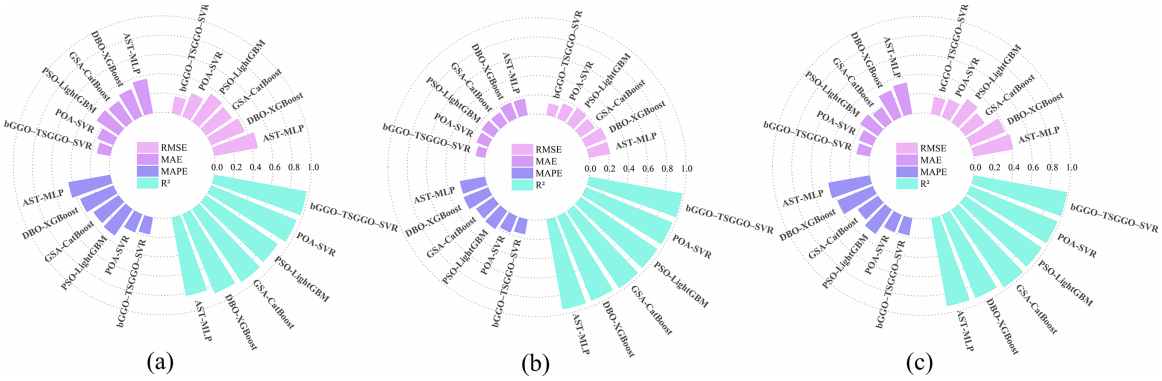

Figure 3: (**a**) Model performance comparison across Product Yield indicators; (**b**) model performance comparison across Drum Index indicators; (**c**) model performance comparison across $RDI_{+3.15}$ indicators.

The error comparison for Product Yield is shown in Fig. 4. The error curve of bGGO-TSGGO-SVR lies closer to the zero line with smaller fluctuations, indicating stronger tracking capability for Product Yield. For the first 80 samples, the mean absolute error of bGGO-TSGGO-SVR is approximately 0.141. The error distribution is more concentrated and shows a lower tail risk. In comparison, although POA-SVR performs well overall, it produces a larger number of samples with relatively large errors under the same threshold. The errors of GSA-CatBoost and PSO-LightGBM are more dispersed, with high error points occurring more

frequently. These results indicate that the proposed model not only achieves lower average errors but also provides stronger suppression of extreme deviations.

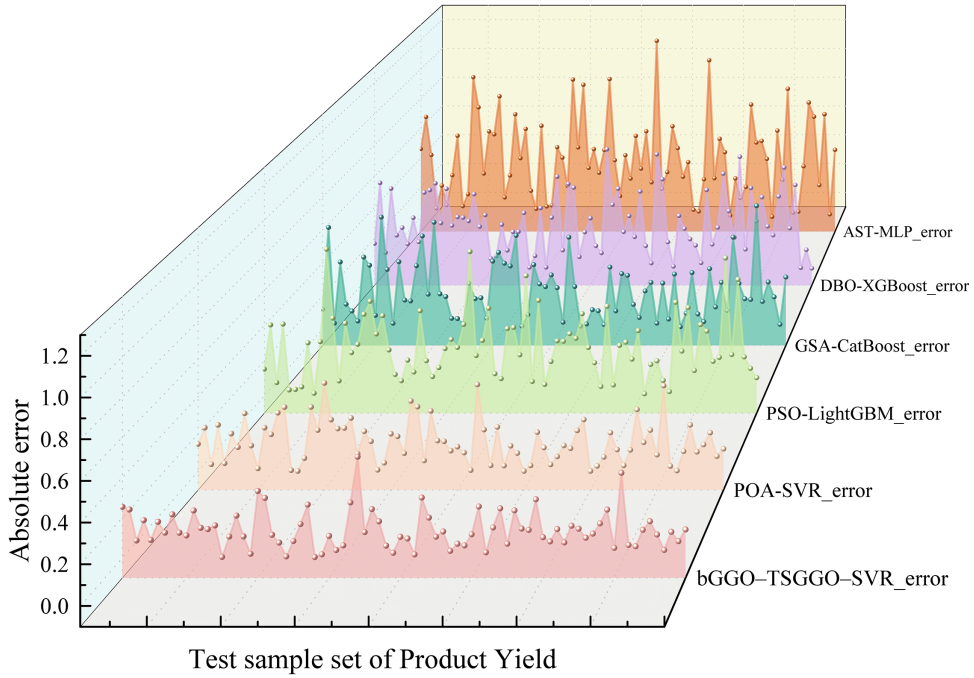

Figure 4: Error comparison of Product Yield for different models.

The error comparison for Drum Index is shown in Fig. 5. Overall, the SVR-based models outperform the neural network model and some ensemble tree models. Among them, bGGO-TSGGO-SVR and POA-SVR exhibit the most similar error levels, and both maintain small magnitude error fluctuations. From the perspective of the error distribution, bGGO-TSGGO-SVR provides better control of tail errors. Approximately 90 percent of the samples have an absolute error no greater than about 0.194, whereas the corresponding value for POA-SVR is about 0.210. This indicates that the proposed method is more stable for samples associated with larger errors. In summary, for the Drum Index task, the proposed method maintains low average errors while further narrowing the tail region of the error distribution, resulting in more stable predictions.

The error comparison for $RDI_{+3.15}$ is shown in Fig. 6, where the advantage of bGGO-TSGGO-SVR is more evident. Its error curve is smoother overall and contains fewer abrupt deviation points. For the first 80 samples, the mean absolute error of bGGO-TSGGO-SVR is approximately 0.130, and about 90 percent of the samples have an absolute error no greater than 0.267. The overall error distribution is more concentrated than that of the other models. The mean absolute error of POA-SVR is approximately 0.142, indicating that the proposed method not only reduces the average error for this index but also lowers the frequency of high error samples. In contrast, some models show pronounced error dispersion and extreme values, with high deviation samples occurring more frequently. This reflects insufficient generalization stability under fluctuating RDI conditions. Therefore, the proposed method can more effectively suppress extreme errors and improve prediction consistency for the RDI index.

Figure 5: Error comparison of Drum Index for different models.

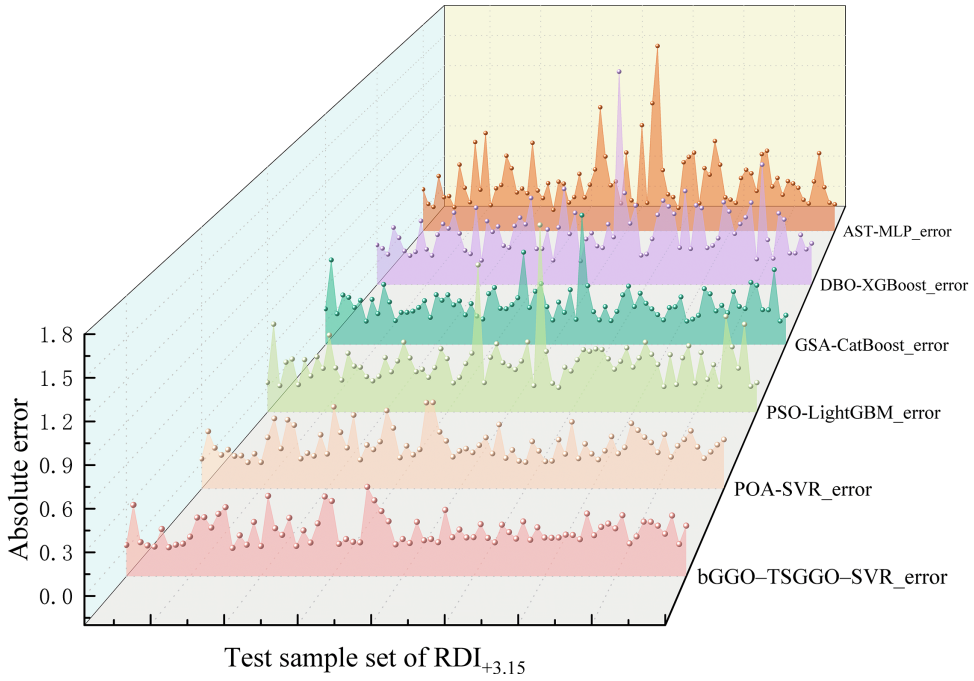

Figure 6: Error comparison of $RDI_{+3.15}$ for different models.

4.3 Ablation Study

To quantitatively evaluate the contribution of each module to performance improvement, four ablation variants were compared under the same dataset and an identical train-test split. The variants include the

baseline SVR, GGO-SVR with only single-stage hyperparameter optimization, bGGO-SVR with only feature selection, and the complete framework bGGO-TSGGO-SVR. The RMSE, MAE, MAPE, and R^2 results of these variants on the three quality indices are summarized in Table 3 and the ablation comparison is further visualized in Fig. 7.

Table 3: Ablation experiment results.

Prediction Metric	Model	RMSE	MAE	MAPE	R^2
Product Yield	SVR	0.418	0.325	0.396	0.855
	GGO-SVR	0.225	0.147	0.173	0.965
	bGGO-SVR	0.229	0.175	0.217	0.952
	bGGO-TSGGO-SVR	0.175	0.140	0.178	0.975
Drum Index	SVR	0.244	0.197	0.304	0.939
	GGO-SVR	0.152	0.119	0.168	0.969
	bGGO-SVR	0.168	0.127	0.181	0.961
	bGGO-TSGGO-SVR	0.122	0.097	0.157	0.985
$RDI_{+3.15}$	SVR	0.348	0.282	0.383	0.943
	GGO-SVR	0.188	0.154	0.191	0.971
	bGGO-SVR	0.295	0.235	0.311	0.959
	bGGO-TSGGO-SVR	0.179	0.139	0.181	0.986

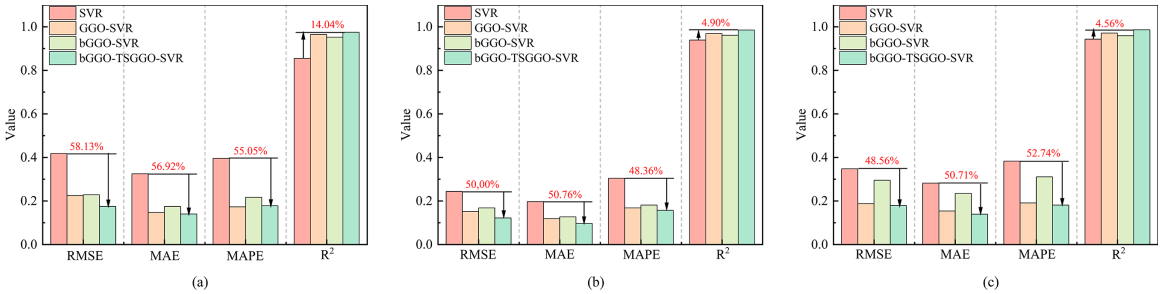

Figure 7: (**a**) Comparison of ablation-study metrics for Product Yield; (**b**) comparison of ablation-study metrics for Drum Index; (**c**) comparison of ablation-study metrics for $RDI_{+3.15}$.

The results in Table 3 and Fig. 7 show that baseline SVR yields large errors across all three indices, indicating that default settings fail to capture the strong nonlinear coupling between process variables and quality. With GGO hyperparameter optimization, performance improves significantly: RMSE drops from 0.418 to 0.225 for Product Yield, from 0.244 to 0.152 for Drum Index, and from 0.348 to 0.188 for $RDI_{+3.15}$, while R^2 rises above 0.965 for all tasks, confirming that hyperparameter tuning is the main driver of SVR generalization.

In the bGGO-SVR variant, which uses feature selection alone, we observe a reduction in errors, but the gains are less stable and exhibit clear variation between the indices. This suggests that removing features may lead to some information loss, which impacts the model's ability to generalize effectively across all quality indices. Feature selection alone, therefore, provides some improvement but is not sufficient to fully optimize the model's performance.

When bGGO is combined with two-stage optimization in the bGGO-TSGGO-SVR model, the best results are achieved. This combination yields the lowest errors and the highest R^2 values across all indices, showing that feature selection and hierarchical hyperparameter tuning complement each other. The combination improves both stability and consistency while preserving high accuracy, further supporting the claim that the two methods work synergistically to enhance model performance.

To ensure that the high R^2 values reported are not due to overfitting, we employed 5-fold cross-validation during the hyperparameter optimization process. The dataset was divided into five subsets, with each subset serving as the test set once, while the remaining four subsets were used for training. The summarization of each fold's cross-validation result of R^2 is shown in Table 4. These results show the consistency of the model performance across different splits of the data, demonstrating that the reported high R^2 values are stable and not due to overfitting.

Table 4: Cross-validation results.

K	Product Yield	Drum Index	$RDI_{+3.15}$
1	0.9741	0.9850	0.9845
2	0.9753	0.9852	0.9843
3	0.9747	0.9847	0.9842
4	0.9739	0.9853	0.9847
5	0.9752	0.9848	0.9846
Average	0.9746	0.9850	0.9845

To further explain the mechanism of the feature selection module, the final feature subsets identified by bGGO for the three quality indices were statistically summarized and analyzed. As shown in Table 5, the optimal input combinations differ noticeably across indices. This indicates that the dominant factors for the three indices are not fully consistent, and it also suggests that sintering process variables contribute differently to different quality characterizations. For the Product Yield task, 14 features were retained, and 3 were removed. For the Drum Index and RDI tasks, 10 features were retained in each case, indicating that these two indices rely on a more concentrated set of key variables. In contrast, Product Yield shows a stronger dependence on a broader combination of process and composition information.

Table 5: Feature selection results.

Prediction Metric	Number	Selected Features
Product Yield	14	Water addition rate, Particle size of mixed materials(<3 mm, 5–8 mm, >8 mm), average particle size, Ignition negative pressure, Layer thickness, Sintering time, TFe, CaO, MgO, SiO_2, Al_2O_3, Moisture
Drum Index	10	Water addition rate, Particle size of mixed materials (5–8 mm, 3–5 mm, <3 mm), Ignition temperature, Ignition negative pressure, Sintering negative pressure, SiO_2, Al_2O_3, Moisture
$RDI_{+3.15}$	10	Water addition rate, Particle size of mixed materials (3–5 mm), average particle size, Ignition negative pressure, Sintering time, TFe, CaO, MgO, Al_2O_3, Moisture

The bGGO-based feature selection results indicate that the optimal input combinations for the three quality indices share commonalities while also exhibiting clear differences. First, all three indices retain the water addition rate, the ignition negative pressure, Al_2O_3, and Moisture. This suggests that the moisture condition of the mixture and the airflow organization during ignition act as common constraints on sinter quality formation. It also indicates that the gangue component Al_2O_3 has a general influence on bonding phase formation and structural evolution. Second, the pairwise shared features reveal differences in index relatedness. Product Yield and Drum Index both retain Particle size of mixed materials (less than 3 mm and 5 to 8 mm) and SiO_2, indicating that particle size gradation and gangue content jointly drive both product formation proportion and strength performance. Product Yield and $RDI_{+3.15}$ share average particle size, sintering time, TFe, CaO, and MgO. This implies that yield and reduction degradation behavior depend more strongly on the mineralogical basis determined by composition and on the sufficiency of the thermal history. The only shared feature between Drum Index and $RDI_{+3.15}$ is Particle size of mixed materials (3 to 5 mm), suggesting partial consistency at the level of particle size structure, while the dominant factors are not fully the same. Finally, index-specific features further reveal mechanistic differences. Product Yield additionally retains Particle size of mixed materials (greater than 8 mm) and Layer thickness, highlighting the key role of bed structure and permeability in product formation. Drum Index uniquely retains Ignition temperature and Sintering negative pressure, emphasizing the importance of thermal intensity and suction conditions for densification of the bonding structure and impact resistance. No unique feature is identified for $RDI_{+3.15}$, indicating that its key influencing factors are mainly associated with shared variables and with the composition and thermal history variables that it shares with Product Yield. Overall, these feature selection results provide data-level evidence that the three indices share common dependencies on moisture condition and airflow organization. They also suggest that Product Yield is more strongly constrained by bed structure and overall process conditions, Drum Index is more sensitive to thermal and suction conditions, and $RDI_{+3.15}$ depends more on the compositional and mineralogical basis together with thermal history control.

4.4 Analysis of Model Prediction Results

The model conducted hyperparameter optimization based on GGO, and the search range for the hyperparameters in the first stage is listed in Table 6.

Table 6: Hyperparameters.

Parameter	Description	Search Variable	Search Range	Decoded Actual Range
Kernel	Kernel type	kernel_flag	[0, 1]	RBF or linear
C	Regularization penalty coefficient	log10 (C)	[−1, 3]	$C \in [10\text{^}-1, 10\text{^}3]$
Gamma	RBF (Radial Basis Function) kernel width parameter	Log10 (gamma)	[−4, 1]	$\gamma \in [10\text{^}-4, 10\text{^}1]$
Epsilon	Epsilon insensitive loss width	Log10 (epsilon)	[−3, 0]	$\varepsilon \in [10\text{^}-3, 10\text{^}0]$

This stage aims to explore a broad range of values with larger step sizes (set to 0.5 in logarithmic space) to quickly identify a region with promising solutions. Stage one accounts for 80% of the total iterations,

enabling broad exploration of the parameter space. In stage two, local fine-tuning is performed within a smaller search region centered around the optimal solution from stage one. The search ranges in stage two are narrowed to:

$$C: [0.5 \times stage1(C), 2 \times stage1(C)], \tag{18}$$

$$gamma: [0.5 \times stage1(gamma), 2 \times stage1(gamma)], \tag{19}$$

$$epsilon: [0.5 \times stage1(epsilon), 2 \times stage1(epsilon)] \tag{20}$$

The step sizes for the fine search are set to 0.1 in logarithmic space. Stage two accounts for 20 percent of the total iterations and focuses on refining the optimal solution obtained from stage one, minimizing uncertainty caused by random fluctuations.

The kernel determines the mapping from the input space to a high-dimensional feature space. The RBF kernel is capable of representing strongly nonlinear relationships, whereas the linear kernel corresponds to weakly nonlinear or approximately linear scenarios. The parameter C is the regularization penalty coefficient that balances training error and model complexity. A larger C encourages lower training error but may introduce overfitting, while a smaller C imposes stronger regularization and yields a smoother model. The RBF kernel parameter gamma controls the effective influence range of the kernel. When gamma is large, the model becomes more sensitive to local samples, and the decision function becomes more complex. When gamma is small, the model is smoother and more robust to noise. The parameter epsilon defines the width of the epsilon-insensitive loss region, within which errors are not penalized. Increasing epsilon can improve robustness to noise but may reduce accuracy, whereas an overly small epsilon makes the model sensitive to minor perturbations. Because reasonable values of C, gamma, and epsilon often span multiple orders of magnitude, logarithmic encoding is adopted to improve optimization efficiency and stability. The optimized hyperparameters for each prediction target are summarized in Table 7.

Table 7: Hyperparameter optimization results.

Prediction Metric	Kernel	C	Gamma	Epsilon
Product Yield	RBF	704.889	0.004	0.153
Drum Index	RBF	42.929	0.065	0.008
$RDI_{+3.15}$	RBF	115.127	0.006	0.025

To visually assess the fitting consistency of the model on the test set, the ground truth and predicted curves for Product Yield, Drum Index, and $RDI_{+3.15}$ were plotted, together with the 95 percent confidence interval of the predictions. As shown in Fig. 8, for Product Yield, the predicted curve closely matches the ground truth over the first 80 test samples and can stably track peak and trough variations. Quantitatively, for the first 80 samples, the MAE is 0.141, and the MAPE is 0.171 percent. The 95th percentile of the absolute error is 0.298, and the maximum absolute error is 0.503. Moreover, 95.0 percent of samples have an absolute error no greater than 0.30. Considering the results over the full test set, the RMSE for Product Yield is 0.175, and the R^2 is 0.975, which further indicates that the model achieves high accuracy and stability for this index.

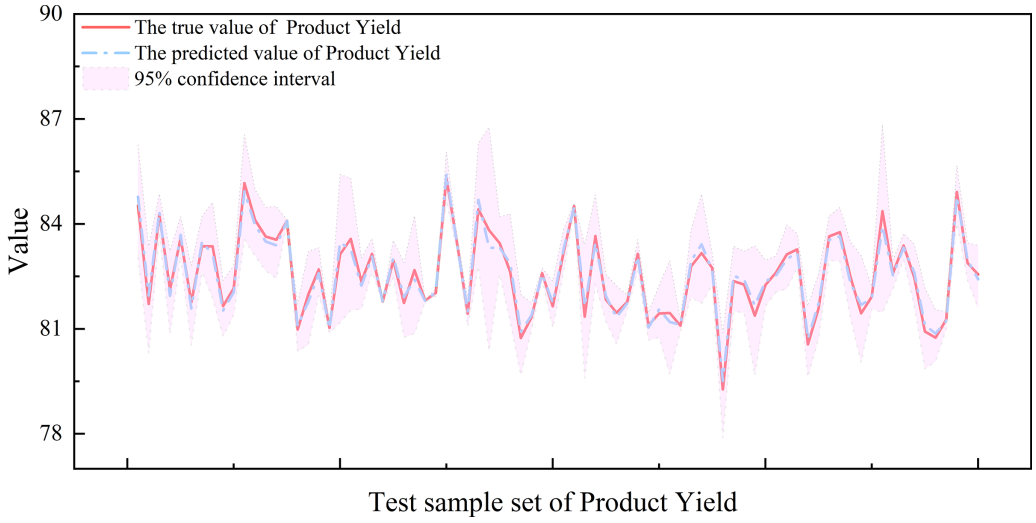

Figure 8: Comparison of ground truth and predicted values for Product Yield.

The results for Drum Index are shown in Fig. 9. The predicted values closely follow the ground truth, with particularly small deviations even during rapid fluctuations, indicating that the model has strong fitting capability for the strength-related index. For the first 80 samples, the MAE is 0.102, the MAPE is 0.155 percent, the 95th percentile of the absolute error is 0.228, the maximum absolute error is 0.367, and 91.25 percent of the samples have an absolute error no greater than 0.20. Over the entire test set, the RMSE for Drum Index is 0.122, and the R^2 reaches 0.985, confirming the model's high precision in predicting this index.

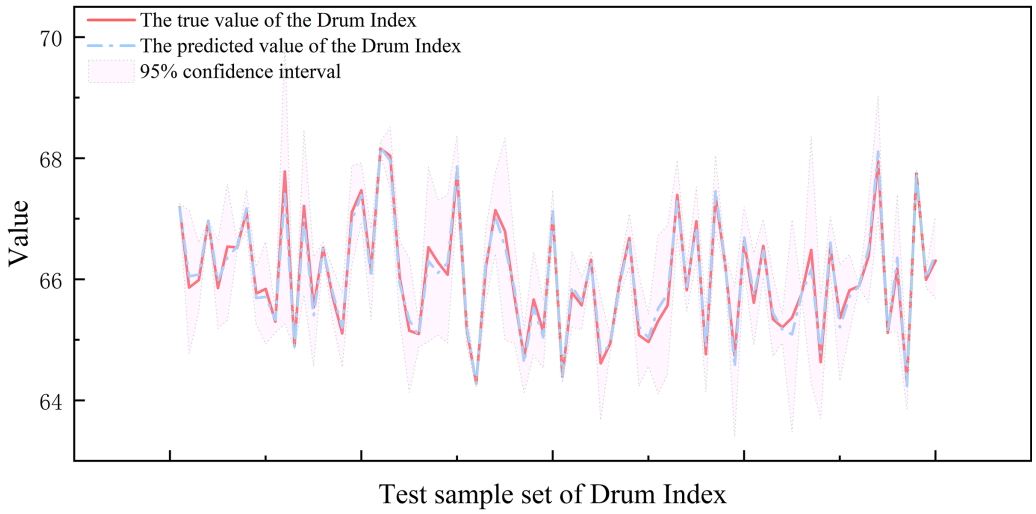

Figure 9: Comparison of ground truth and predicted values for Drum Index.

The results for $RDI_{+3.15}$ are shown in Fig. 10. The predicted curve closely matches the ground truth at most sample points, effectively capturing the overall trend and local fluctuations of $RDI_{+3.15}$. This indicates that the model also demonstrates strong learning and generalization capabilities for the reduction degradation-related indices. For the first 80 samples, the MAE is 0.130, the MAPE is 0.172 percent, the 95th percentile of the absolute error is 0.338, the maximum absolute error is 0.439, and 92.5 percent of the samples

have an absolute error no greater than 0.30. Over the entire test set, the RMSE for $RDI_{+3.15}$ is 0.179, and the R^2 is 0.986, indicating that the model maintains high prediction consistency and stability for this index.

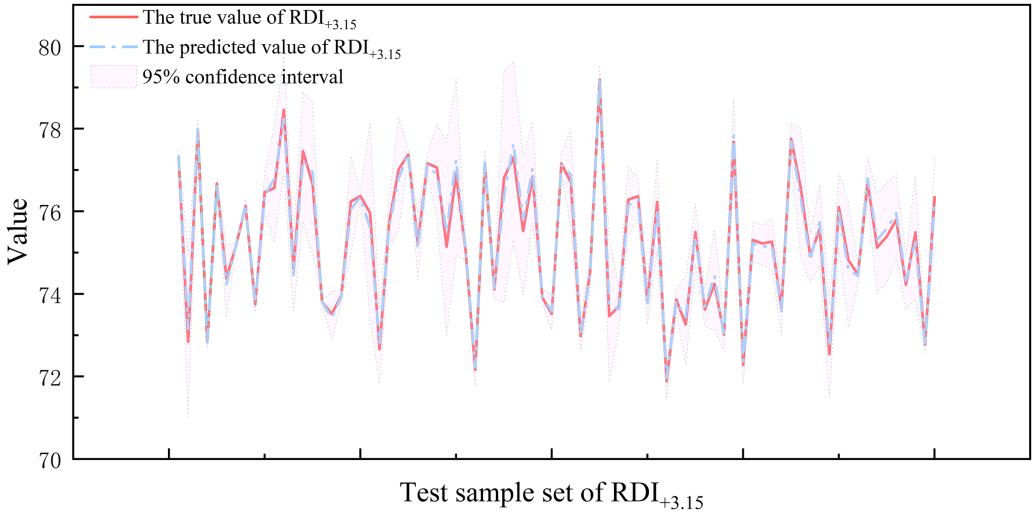

Figure 10: Comparison of ground truth and predicted values for $RDI_{+3.15}$.

4.5 Model Application

To improve the practical applicability of the proposed bGGO-GGO-SVR model, a prototype sinter quality prediction system was developed based on PyQt5 [37]. The system provides a graphical user interface for model loading, sample input, prediction, result display, and trend visualization, thereby linking offline modeling with engineering applications. As shown in Fig. 11, the system supports both single-sample and batch prediction, while historical results can be visualized as line charts to reflect the variation trends of multiple quality indices. This design improves the interpretability and usability of the model in practical quality monitoring and process control.

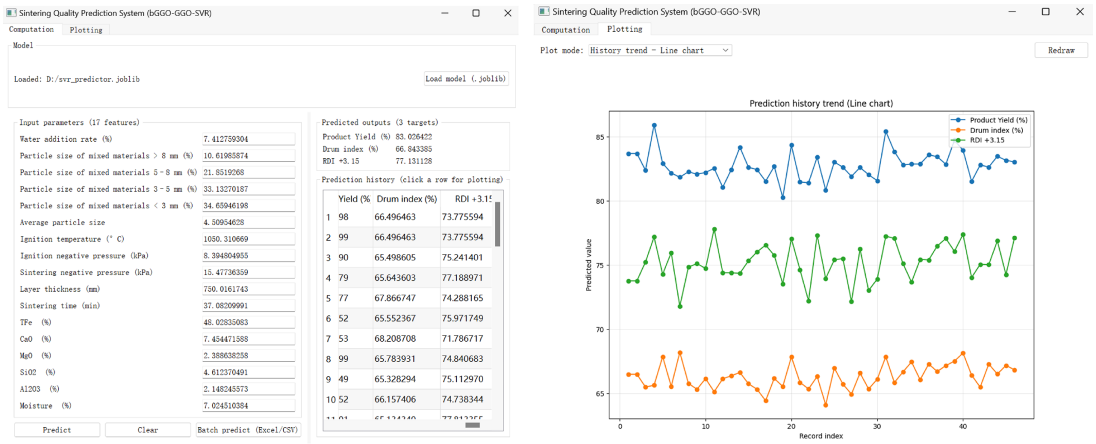

Figure 11: System interface. (**a**) Calculation interface; (**b**) plotting interface.

5 Conclusion

This study proposed a bGGO-TSGGO-SVR framework for multi-index sinter quality prediction using operational data from industrial trials. The framework employed support vector regression as the core predictor, combined with improved binary Greylag Goose Optimization for feature selection, and adopted two-stage Greylag Goose Optimization for coarse-to-fine hyperparameter tuning. In this way, redundant and collinear variables were reduced, while model fitting ability and generalization performance were improved under strongly nonlinear and multi-factor coupled conditions. Experimental results showed that the proposed bGGO-TSGGO-SVR method achieved significant improvements in prediction accuracy compared to five representative baseline models for Product Yield, Drum Index, and $RDI_{+3.15}$. Specifically, the model reduced RMSE by 32.2 percent for Product Yield, by 26.1 percent for Drum Index, and by 8.2 percent for $RDI_{+3.15}$. Additionally, the R^2 values increased by 0.019 for Product Yield, by 0.016 for Drum Index, and by 0.012 for $RDI_{+3.15}$, demonstrating the model's superior performance. Ablation studies further revealed that hierarchical hyperparameter optimization played the dominant role in performance improvement, while feature selection was most effective when combined with appropriate hyperparameter tuning. These results indicate that the proposed framework offers a highly effective data-driven solution for multi-index sinter quality prediction and process optimization.

To ensure the framework's real-time applicability in industrial environments, future work should focus on adapting the optimization process to handle online predictions efficiently. While the current two-stage optimization can be computationally intensive, leveraging parallel processing, distributed computing, and model compression techniques can reduce computational costs. Additionally, the framework's scalability can be improved by using incremental learning methods, allowing for continuous adaptation as new data and quality indices are introduced in industrial settings. These considerations will help make the framework suitable for large-scale and real-time industrial deployment. Since this study is conducted on static data under steady operating conditions, future research could also explore the model's robustness by evaluating its performance under various process disturbances and nonstationary conditions, which are commonly encountered in real-world industrial scenarios.

As a continuation and further development of this work, several lines for future research could be explored.

Incorporating economic variables: One potential direction involves integrating economic factors, such as operating costs, into the current framework. By not only predicting process quality but also assessing the associated economic impact, the model can evolve into a decision-support tool that aligns with industrial optimization objectives. This would provide a more holistic solution for process management and cost-efficiency in industrial settings.

Interpretability techniques: Another promising area is the incorporation of interpretability techniques, such as methods based on variable importance analysis. This would help in understanding the relationships between input variables and the model's output, thus enhancing the transparency of the model. Interpretability is particularly important in industrial applications, where decision-making processes often require clear justification for the recommendations made by predictive models.

Evaluation across different operational scenarios: Finally, to further validate the robustness and generalizability of the model, it would be beneficial to evaluate its performance across various operational scenarios or datasets. This would help assess the reliability of the framework under different industrial conditions, thus expanding its potential applications in diverse real-world environments.

These research directions, encompassing the integration of economic considerations, interpretability, and validation across diverse scenarios, will not only improve the framework's real-world applicability but also broaden its scope for industrial optimization and decision-making.

Acknowledgement: None.

Funding Statement: This work was supported by National Natural Science Foundation of China (52074126); the Project of Yanzhao Iron and Steel Laboratory (25364004D), and the Graduate Student Innovation Fund of North China University of Science and Technology (No. 2026S27).

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Yifan Li; methodology, Zhenhua Yang and Yifan Li; software, Zhenhua Yang and Tao Xue; validation, Zhenhua Yang; formal analysis, Zhenhua Yang, Jie Li and Tao Xue; resources, Yifan Li, Aimin Yang and Jie Li; data curation, Zhenhua Yang, Yifan Li and Tao Xue; writing—original draft preparation, Zhenhua Yang; writing—review and editing, Aimin Yang and Jie Li; visualization, Zhenhua Yang and Tao Xue; supervision, Aimin Yang and Jie Li; project administration, Aimin Yang and Jie Li; funding acquisition, Aimin Yang and Jie Li. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Aimin Yang, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li Q, Wang JS, She XF, Xue QG, Guo ZC, Wang G, et al. Principles and support technologies of low-carbon ironmaking for blast furnace. Renew Sustain Energy Rev. 2025;211(9):115363. doi:10.1016/j.rser.2025.115363.
2. Du S, Ma X, Fan H, Hu J, Cao W, Wu M, et al. Intelligent prediction and soft-sensing of comprehensive production indicators for iron ore sintering: a review. Comput Ind. 2025;165(6):104215. doi:10.1016/j.compind.2024.104215.
3. Hu J, Li H, Liu J, Du S. Review of intelligent modeling for sintering process under variable operating conditions. Processes. 2025;13(1):180. doi:10.3390/pr13010180.
4. Yan F, Zhang X, Yang C, Hu B, Qian W, Song Z. Data-driven modelling methods in sintering process: current research status and perspectives. Can J Chem Eng. 2023;101(8):4506–22. doi:10.1002/cjce.24790.
5. Niu L, Liu Z, Zhang J, Sun Q, Schenk J, Wang J, et al. Prediction of sinter chemical composition based on ensemble learning algorithms. J Sustain Metall. 2023;9(3):1168–79. doi:10.1007/s40831-023-00717-x.
6. Li Y, Duan Y, Zhou Y, Yang J, Li F, Yang A. Research on prediction model of iron ore powder sintering foundation characteristics based on FOA-Catboost algorithm. Alex Eng J. 2024;86(8):603–15. doi:10.1016/j.aej.2023.12.015.
7. Li X, Liu X, Li H, Liu R, Zhang Z, Li H, et al. Research on sinter quality prediction system based on granger causality analysis and stacking integration algorithm. Metals. 2023;13(2):419. doi:10.3390/met13020419.
8. Wang S. Applying genetic algorithm to extreme learning machine in prediction of tumbler index with principal component analysis for iron ore sintering. Sci Rep. 2025;15(1):4777. doi:10.1038/s41598-025-88755-1.
9. Li Y, Yang C, Sun Y. Sintering quality prediction model based on semi-supervised dynamic time feature extraction framework. Sensors. 2022;22(15):5861. doi:10.3390/s22155861.
10. Zhao L, Meng S, Zhao X, Sun K, Cheng Z. An improved informer model for predicting sinter alkalinity based on multi-scale feature fusion. JOM. 2025;77(5):3837–48. doi:10.1007/s11837-025-07278-2.
11. Yang C, Yang C, Li J, Li Y, Yan F. Forecasting of iron ore sintering quality index: a latent variable method with deep inner structure. Comput Ind. 2022;141(2):103713. doi:10.1016/j.compind.2022.103713.

12. Li H, Wu M, Du S, Hu J, Zhang W, Chen L, et al. Prediction model of burn-through point with data correction based on feature matching of cross-section frame at discharge end. J Process Control. 2024;140(3):103265. doi:10.1016/j.jprocont.2024.103265.
13. Li S, Cao Y, Zhou Z, Li X, Zhu Y. Research on prediction method of ferrous oxide content in sinter based on optimized neural network. Minerals. 2025;15(6):553. doi:10.3390/min15060553.
14. Yang C, Yang C. Deep fusion of time series and visual data through temporal features: a soft-sensor model for FeO content in sintering process. Expert Syst Appl. 2025;268(2):126243. doi:10.1016/j.eswa.2024.126243.
15. Zhao Z, Feng W, Liu S, Xiong Z, Zhao Y, Zhang H, et al. Sinter quality prediction based on multi-features CNN + LSTM. Arab J Sci Eng. 2024;49(3):4271–86. doi:10.1007/s13369-023-08343-1.
16. Li Y, Wang B, Zhou Z, Yang A, Bai Y. Research progress of intelligent ore blending model. Metals. 2023;13(2):379. doi:10.3390/met13020379.
17. Chen X, Hu Y, Liu C, Chen A, Chi Z. Dynamic spatio-temporal graph network based on multi-level feature interaction for sinter TFe prediction. J Process Control. 2025;148(8):103401. doi:10.1016/j.jprocont.2025.103401.
18. Yan F, Yang C, He W, Mu J, Guo H. Knowledge and data dual-driven graph neural network for tumbler strength prediction in sintering process. IEEE Trans Instrum Meas. 2024;73(3):1–14. doi:10.1109/tim.2024.3428604.
19. Shi Z, Zhou K, Li T, Wu Z. Sinter physical indices prediction: a model integrating spatiotemporal feature fusion and graph attention network. JOM. 2025;77(12):10222–34. doi:10.1007/s11837-025-07824-y.
20. Chi Z, Chen X, Xia H, Liu C, Wang Z. An adaptive control system based on spatial–temporal graph convolutional and disentangled baseline-volatility prediction of bellows temperature for iron ore sintering process. J Process Control. 2024;140(12):103254. doi:10.1016/j.jprocont.2024.103254.
21. Hu J, Wu M, Chen X, Cao W, Pedrycz W. Multi-model ensemble prediction model for carbon efficiency with application to iron ore sintering process. Control Eng Pract. 2019;88(1):141–51. doi:10.1016/j.conengprac.2019.05.009.
22. Hu J, Wu M, Chen L, Cao W, Pedrycz W. Real-time dynamic prediction model of carbon efficiency with working condition identification in sintering process. J Process Control. 2022;111(1):97–105. doi:10.1016/j.jprocont.2022.02.002.
23. Du S, Wu M, Chen L, Cao W, Pedrycz W. An intelligent decision-making strategy based on the forecast of abnormal operating mode for iron ore sintering process. J Process Control. 2020;96(7):57–66. doi:10.1016/j.jprocont.2020.11.001.
24. Chen A, Chen X, Liu C, Shi X, Yu B, Guo Q. Spatio-temporal feature extraction model based on dynamic identification of operating conditions for sintering tumbler strength prediction. Control Eng Pract. 2025;164(170):106484. doi:10.1016/j.conengprac.2025.106484.
25. Li Y, Cao Y, Yang J, Wu M, Yang A, Li J. Optuna-DFNN: an Optuna framework driven deep fuzzy neural network for predicting sintering performance in big data. Alex Eng J. 2024;97(9):100–13. doi:10.1016/j.aej.2024.04.026.
26. Feng S, Wang B, Zhou Z, Xue T, Yang A, Li Y. Research on multi-decision sinter composition optimization based on OLS algorithm. Metals. 2023;13(3):548. doi:10.3390/met13030548.
27. Fang Y, Jiang Z, Gui W, Shen L. Towards reliable control: uncertainty-aware domain preserving stacked auto-encoder for data-driven modeling in large-scale industrial systems. Control Eng Pract. 2025;164(3):106383. doi:10.1016/j.conengprac.2025.106383.
28. Liu W, Bai Y, Zhang C, Wang Z, Yang A, Wu M. PSO-DFNN: a particle swarm optimization enabled deep fuzzy neural network for predicting the pellet strength. Alex Eng J. 2024;106(14):505–16. doi:10.1016/j.aej.2024.08.069.
29. Jursova S, Pustejovska P, Brozova S. Study on reducibility and porosity of metallurgical sinter. Alex Eng J. 2018;57(3):1657–64. doi:10.1016/j.aej.2017.03.007.
30. Williamson DF, Parker RA, Kendrick JS. The box plot: a simple visual method to interpret data. Ann Intern Med. 1989;110(11):916–21. doi:10.7326/0003-4819-110-11-916.
31. El-kenawy EM, Khodadadi N, Mirjalili S, Abdelhamid AA, Eid MM, Ibrahim A. Greylag goose optimization: nature-inspired optimization algorithm. Expert Syst Appl. 2024;238(22):122147. doi:10.1016/j.eswa.2023.122147.

32. Chen Z, Ran Q, Chen N, Zhang Y, Tang J, Long J. Spatio-temporal MLP model incorporating physical-informational neural networks for remaining useful life prediction. Mech Syst Signal Process. 2026;244(23):113815. doi:10.1016/j.ymssp.2025.113815.
33. Wang L, Wang Y, Wang Y. Enhancing tetracycline degradation prediction with hybrid DBO-XGBoost on g-C_3N_4 photocatalysts. Environ Technol Innov. 2026;41(3):104711. doi:10.1016/j.eti.2025.104711.
34. Ruan B, Hu Y, Liu C, Zhou Z. Prediction of compression modulus of Nantong fine-grained soil based on an interpretability-driven GSA-optimized CatBoost model. Adv Eng Inform. 2026;69(20):104034. doi:10.1016/j.aei.2025.104034.
35. Li Z, Huang Q, Liu Z, Tian X. A method for predicting lithium-ion battery capacity using an enhanced LightGBM model. Energy Rep. 2025;14:4212–27. doi:10.1016/j.egyr.2025.11.032.
36. Zhao J, Du X, Yu Y, Wu M, Nian H, Yang L. Remaining useful life prediction of lithium-ion batteries based on POA-SVR method in experimental platforms. Energy Rep. 2025;14(1):5360–70. doi:10.1016/j.egyr.2025.11.029.
37. Li Z, Wang M, Xu R, Jiang J, Li J, Zhang Z, et al. Prediction of viscosity of blast furnace slag based on NRBO-DNN model. Alex Eng J. 2025;119(11):124–37. doi:10.1016/j.aej.2025.01.126.