Research on Sintering Burn-Through Point Prediction Based on Multi Optimization Algorithms and Deep Learning

Zhen Dong[1], Menghui Bai[1], Haifeng Song[2]
[1]*College of Artificial Intelligence, North China University of Science and Technology, Tangshan, Hebei, China*
[2]*College of Science, North China University of Science and Technology, Tangshan, Hebei, China*

Abstract: With the collaborative evolution of computing power and algorithms, machine learning has become the core supporting technology for solving complex problems in metallurgical engineering. Burn-Through Point (BTP) soft sensing faces typical big data challenges such as strong coupling of multiple variables, non-stationary across time scales, and high-frequency noise. This paper proposes a prediction model that integrates Osprey Optimization Algorithm (OOA), Time varying Filter Empirical Mode Decomposition (TVF-EMD), Dwarf Mongoose Optimiza-tion (DMO), and Gated Recurrent Unit (GRU). Firstly, perform missing value completion and Min Max normalization on industrial measured data sampled at 10 months and 30 minutes; Subsequently, with the minimization of sample entropy as the fitness, OOA is used to globally optimize the bandwidth threshold and B-spline order of TVFEMD, achieving efficient de-noising and multi-scale decomposition. All IMFs are reorganized into high, medium, and low frequency collaborative mode functions (Co-IMF) through K-Means clustering based on sample entropy; Further use VMD deep narrowband reconstruction to enhance key intermediate frequency features for Co-IMF0, which has the highest energy proportion. In the prediction stage, DMO adaptively searches for hyperparameters such as batch_size and epoch of GRU to form a DMO-GRU predictor, and outputs weighted superposition of the three sub signals. Compared with seven baseline models including TVFEMD-GRU and VMD-PSO-LSTM, this method reduces the RMSE from 2.86 to 1.89 and improves the R2 to 0.990 on the test set, demonstrating robustness to high-frequency noise, in-sensitivity to local optima of hyperparameters, and ac-curate capture of key modal information. the research provides a replicable new paradigm for green, low-carbon, and intelligent manufacturing in the sinter-ing process of the steel industry.

Keywords: Burn-Through Point; Multiscale Decom-Posetion; OOA-TVFEMD; DMO-GRU

1. Introduction

Global crude steel production has surged to approximately 1960 Mt in 2024, with the sintering process accounting for over 70% of energy consumption and CO2 emissions in iron production [1-2]. To achieve "dual carbon" goals, digital trans-formation is imperative. the Burn-Through Point (BTP), defined as the location where the flame front burns through the material layer, is the critical indicator for sinter quality and fuel efficiency. BTP fluctuations lead to unstable permeability, increased return ore rates, and fuel waste. Therefore, stabilizing BTP within a standard window via precise rolling prediction is essential for intelligent manufacturing [3-9].

Existing BTP prediction methods have evolved through three stages: mechanism modeling, machine learning, and deep learning. Early mechanism models [10] and BP neural networks [11] established the foundation but struggled with complex nonlinearities. Machine learning approaches, such as Random Forest and GBDT [12-17], improved accuracy via feature engineering but often lacked temporal depth. Recently, deep learning models like LSTM and DNN [18-22] have shown superior performance. However, two major challenges remain limiting their implementation [23-27]:

To systematically address these issues, this paper proposes an integrated OOA-TVFEMD-DMO-GRU framework. First, an adaptive decomposition strategy is established where the Osprey Optimization Algorithm (OOA) globally

optimizes the bandwidth threshold and B-spline order of TVFEMD. By utilizing sample entropy minimization as the fitness criterion, this step achieves robust signal-to-noise separation compared to empirical parameter selection. Second, a feature reorganization mechanism is employed using K-Means clustering to reconstruct Intrinsic Mode Functions (IMFs) into three physically interpretable collaborative modes (Co-IMFs)—high-frequency noise, mid-frequency mechanism fluctuations, and low-frequency trends—thereby reducing input redundancy. Finally, the model implements dual optimization by applying Dwarf Mongoose Optimization (DMO) to adaptively tune critical GRU hyperparameters, such as batch size and iteration epochs. This approach effectively resolves the limitations of manual parameter selection and ensures high-precision rolling prediction.

2. Data Preprocessing

The data used in this article comes from actual collected data during the sintering process of a certain steel enterprise, including multiple variables such as temperature, water flow rate, and machine speed, covering the entire process information from the sinter-ing machine head to the sintering machine tail. the original data was collected over a period of 10 months, with a sampling frequency of every 30 minutes, and a total of 11149 pieces of data. the target variable is the sintering endpoint.

Considering the significant dimensional di-fferences among different variables, the subsequent TVFEMD de-composition and GRU model training may be affected. Therefore, all features were nor-malized using the Min Max normalization method to scale them to the [0, 1] interval. the normalization formula is as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

To improve the efficiency of model training and reduce the risk of overfitting caused by redundant input variables, this study conducted a correlation analysis between the original features and the target variable, using Pearson Correlation Coefficient (PCC) as the criterion for discrimination. the calculation formula is as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

Among them, r_{xy} represents the degree of linear correlation between variables x and y, with a range of [−1, 1]. the larger the absolute value, the stronger the correlation.

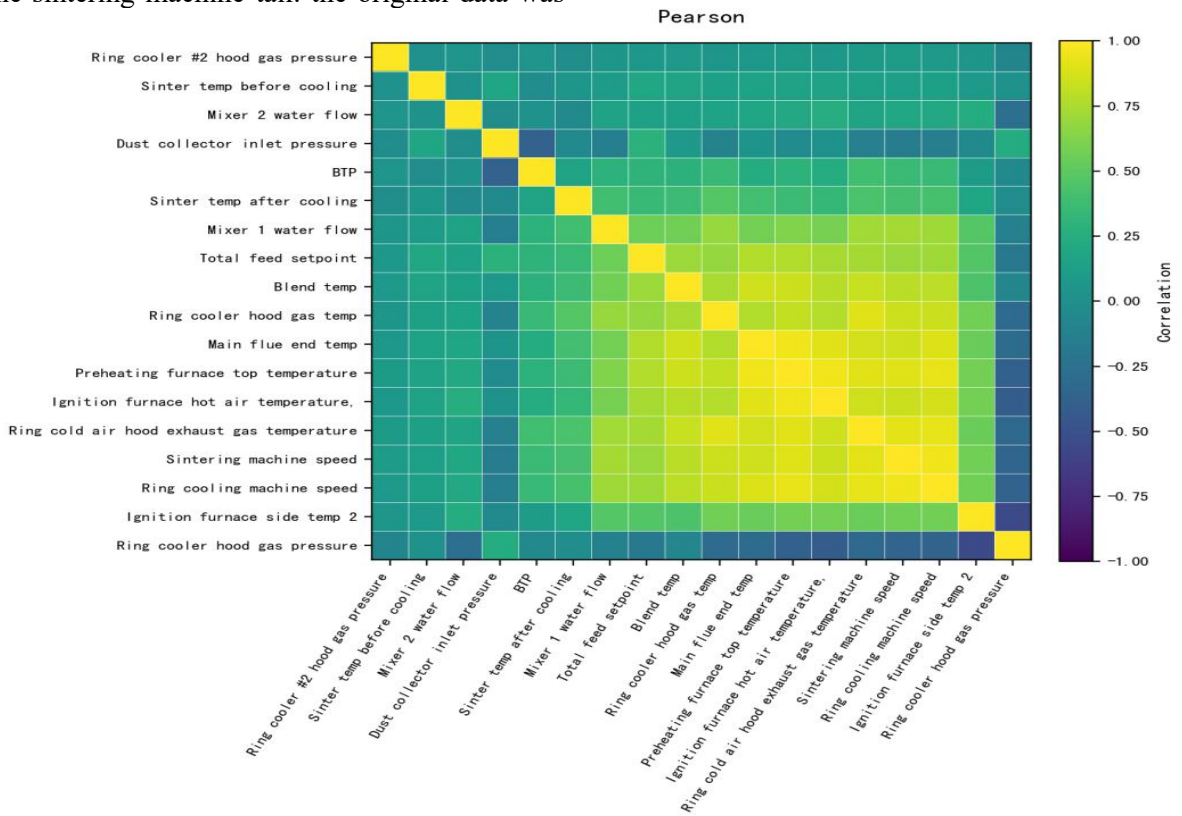


Figure 1. Heat Map

Calculate Pearson coefficients between all variables and the target variable, and draw a heatmap (as shown in Figure 1) to visually display the correlation between variables. By analyzing the heat map and specific coefficient values, several variables highly correlated with the sintering endpoint indicators were selected: sintering machine speed, ring cooling ma-chine speed, preheating furnace top temperature, ignition furnace side temperature, ring cold air hood exhaust gas temperature, ignition furnace hot air temperature, as input features for subsequent TVFEMD decomposition and GRU models. In addition, there is obvious collinearity among some variables. To further improve the modeling effect, denoising and modal reconstruction were carried out on these variables in the TVFEMD decomposition stage.

3. Model Establishment

3.1 OOA-TVFEMD Model

TVFEMD is an advanced signal processing technique that introduces time-varying filters to overcome mode mixing in traditional EMD. However, its performance depends heavily on two hyperparameters: the bandwidth threshold ξ and the B-spline order n.

Osprey Optimization Algorithm (OOA):To avoid empirical parameter selection, we employ OOA [28], a metaheuristic simulating the hunting behavior of ospreys. OOA is used to globally search for the optimal (ξ, n) pair. the objective function is to minimize the Sample Entropy (SE) of the decomposition result, as lower entropy indicates higher regularity and better noise separation:

$$MinimizeF = SE\big(Rec(\xi, n)\big) \tag{3}$$

where ξ represents the bandwidth threshold determining the filtering strength, and n denotes the B-spline order controlling the smoothness. the Sample Entropy is calculated as:

$$SE(m, r, N) = -\ln\left(\frac{A^m(r)}{B^m(r)}\right) \tag{4}$$

where m is the embedding dimension, r is the similarity tolerance, and N is the sequence length. $B^m(r)$ denotes the probability that two vectors match for m points, while $A^m(r)$ represents the probability that they match for $m + 1$ points.

Co-IMF Reconstruction via K-Means: To extract robust features, we calculate the SE of each Intrinsic Mode Function (IMF) and use K-Means clustering (K=3) to reorganize them into Collaborative Mode Functions (Co-IMFs):

1. Co-IMF0 (High-frequency): High SE, representing noise.
2. Co-IMF1 (Mid-frequency): Moderate SE, representing mechanism fluctuations.
3. Co-IMF2 (Low-frequency): Low SE, representing the baseline trend.

3.2 DMO-GRU Model

Gated Recurrent Unit (GRU): the network comprises a three-layer stacked GRU (128/64/32 units), followed by Dropout (0.2). GRU uses update and reset gates to capture temporal dependencies efficiently.

Dwarf Mongoose Optimization (DMO): DMO is to optimize the integer hyperparameters of GRU (e. g., batch size and epochs). DMO simulates the foraging behavior of the Alpha group. the position update is modeled as:

$$X_{i+1} = X_i + \phi \times (X_{best} - X_i) \tag{5}$$

where X_{i+1} and X_i are the updated and current positions of the $i-th$ individual, respectively; X_{best} represents the global optimal position (Alpha female); and ϕ is the distributed movement vector controlled by the "collective will" parameter. the fitness function is the minimization of Mean Squared Error (MSE) on the validation set.

3.3 OOA-TVFEMD-DMO-GRU Model

The proposed framework operates through a systematic process of "decomposition-reconstruction-integration. " Initially, the raw sintering endpoint data undergoes processing via the OOA-optimized TVFEMD algorithm, where the bandwidth threshold ξ and B-spline order n are globally searched to minimize sample entropy. the resulting components are then restructured into three Collaborative Mode Functions (Co-IMFs) via K-Means clustering, effectively separating high-frequency noise, mid-frequency mechanism fluctuations, and low-frequency trends. Subsequently, these three distinct Co-IMFs serve as parallel inputs for three independent GRU predictors. To accommodate the different temporal characteristics of each frequency band, DMO is employed to adaptively search for the optimal hyperparameters (batch size and epochs) for each specific GRU network, ensuring that each sub-model converges at an appropriate pace. Finally, the outputs of the three sub-models are integrated using a weighted superposition

strategy to generate the final BTP prediction $\hat{Y}_{final}$:

$$\hat{Y}_{final} = \sum_{k=0}^{2} \omega_k \cdot GRU_k(Co - IMF_k) \quad (6)$$

where w_k denotes the adaptive weight for the k-th collaborative mode. These weights are fitted via non-negative least squares on the validation set to dynamically balance the contribution of trend information and fluctuation details. For practical deployment, a rolling window strategy is adopted, updating the GRU hidden states and Co-IMF inputs with every new batch of data to achieve low-latency online prediction.

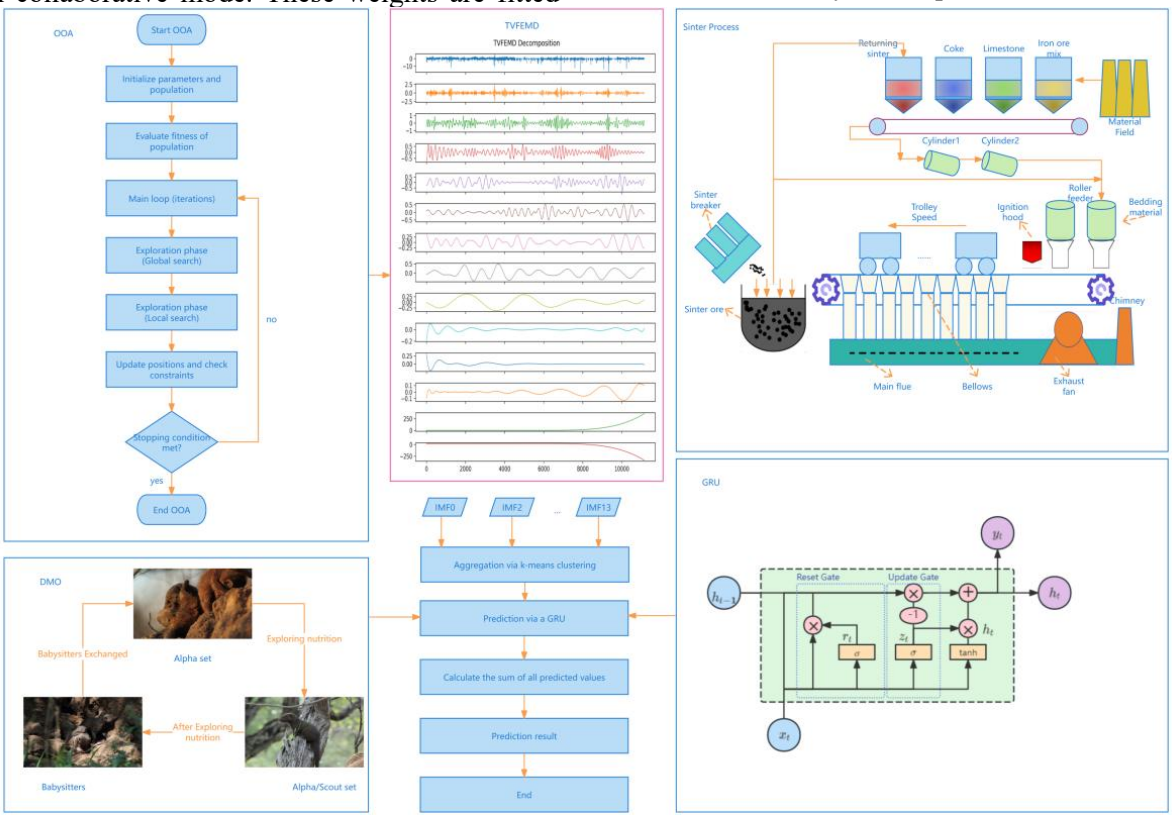


Figure 4. Flow Chart

4. Results and Discussion

4.1 Visualization of Signal Decomposition and Feature Extraction

Fig. 5 shows the signal stripped into 13 IMFs. High-frequency components show sharp random fluctuations, while low-frequency components converge to a gentle trend. the Sample Entropy analysis (Fig. 6) reveals a clear "decreasing complexity" law: SE peaks at IMF1 and drops to a low plateau after IMF6. This confirms that high-order IMFs are deterministic trends, while low-order IMFs contain high uncertainty.

Based on the sample entropy feature vectors, the K-Means algorithm automatically stratifies all IMFs into three distinct Collaborative Mode Functions (Co-IMFs), as visualized in Fig. 7. the first component, Co-IMF0, is dominated by high-frequency signals characterized by large amplitude peaks and high entropy values. This cluster primarily aggregates random noise and instant sensor interference, which are effectively isolated in this stage to prevent error amplification in downstream modeling.

In contrast, the remaining components capture the meaningful physical characteristics of the sintering process. Co-IMF1 gathers intermediate-frequency signals manifested as dense quasi-periodic fluctuations. These features are phy-siologically significant as they reflect the core dynamics of the combustion zone and the coupling effects of air pressure. Meanwhile, Co-IMF2 is composed of low-frequency components with the minimum sample entropy and the highest energy proportion, representing the stable, slow-changing baseline trends of the operating conditions. This hierarchical restructuring successfully validates the proposed signal processing strategy of "high-frequency suppression, mid-frequency highlighting, and low-frequency retention, " laying a robust feature foundation for the subsequent prediction steps.

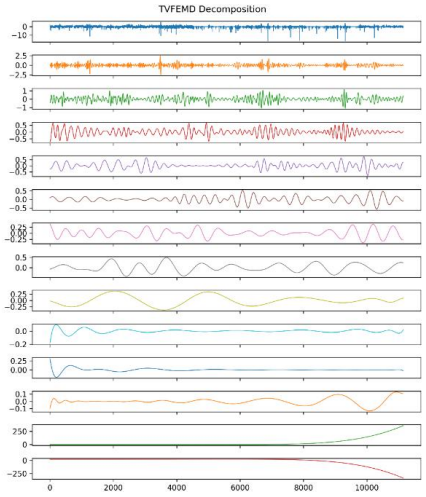


Figure 5. TVFEMD Decomposition Results

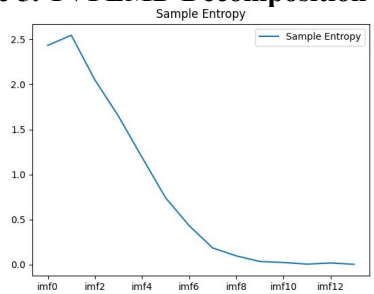


Figure 6. Comparison of Sample Entropy of Various IMF

4.2 Analysis and Comparison of Prediction Results

To verify the effectiveness of the method proposed in this article, eight representative models were selected for comparison under the same operating conditions. 250 sets of training samples and 100 sets of test samples were uniformly used, and R2, RMSE, MAE, and MAPE were used as evaluation indicators. the point by point prediction curves of 100 test samples are shown in Figure 8.

The comprehensive comparison reveals the critical necessity of the proposed dual-optimization strategy. the OOA-TVFEMD-DMO-GRU model achieved the lowest RMSE of 0.137, significantly outperforming the OOA-TVFEMD-GRU variant (RMSE 0.225). This discrepancy highlights that optimizing decomposition alone is insufficient; without the adaptive hyperparameter tuning provided by DMO, the standard GRU fails to fully exploit the purified features, potentially leading to overfitting on over-smoothed data. Conversely, relying solely on prediction-end optimization (TVFEMD-DMO-GRU) yielded only marginal improvements (RMSE 0.151) over the baseline. This suggests that even with optimal model parameters, the prediction accuracy remains fundamentally limited by the noise aliasing and modal instability present in the unoptimized decomposition stage.

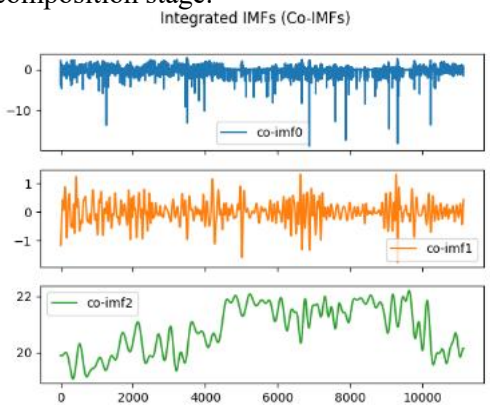
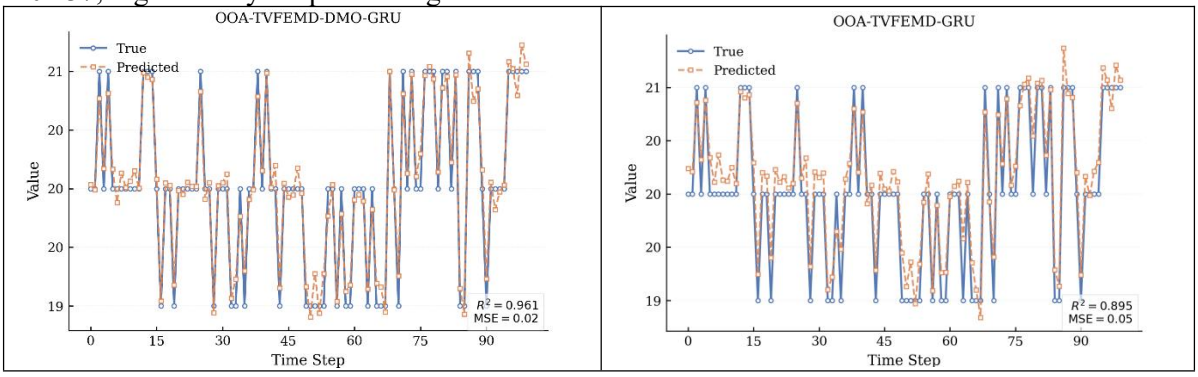


Figure 7. Co-IMF Restructuring Results

From the perspective of error morphology, as shown in Fig. 8, the proposed combination significantly compresses amplitude and phase errors during peak fluctuation periods and maintains low bias during operating condition switches. This demonstrates that the synergistic mechanism—where OOA stabilizes feature extraction and DMO ensures convergence in a clean feature space—fundamentally resolves the twin bottlenecks of "noise amplification" and "local optima. " Consequently, the model exhibits the robustness required for online engineering deployment, surpassing the capabilities of models that rely on single-point optimization.



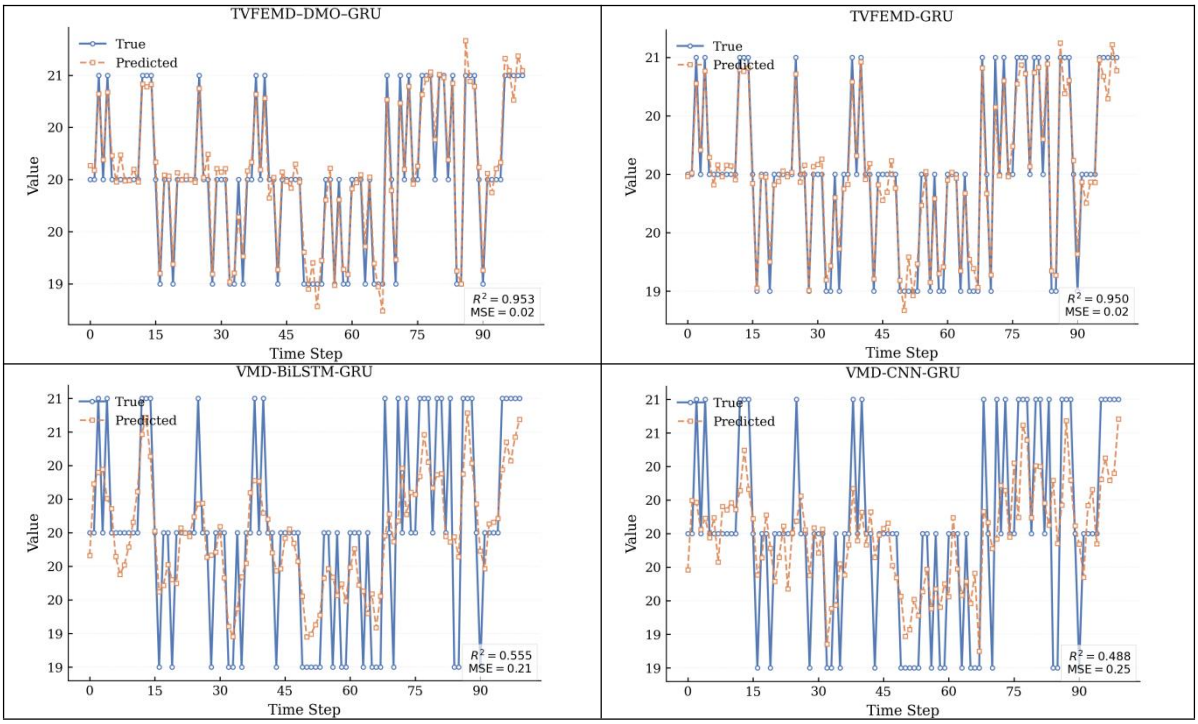


Figure 8. Point by Point Prediction Curve of Test Samples

Table 1. Comparison of Predictive Performance of Eight Models on the Test Set

Model	R^2	RMSE	MAE	MAPE
OOA–TVFEMD–DMO–GRU	0.961	0.137	0.110	0.55%
OOA–TVFEMD–GRU	0.895	0.225	0.190	0.96%
TVFEMD–DMO–GRU	0.953	0.151	0.121	0.61%
TVFEMD–GRU	0.950	0.155	0.126	0.63%
VMD–LSTM	0.337	0.564	0.464	2.32%
VMD–PSO–LSTM	0.815	0.300	0.250	1.24%

5. Conclusion

Facing multivariable coupling, multi-scale non-stationarity and high-frequency noise in BTP soft sensing, this paper proposes an integrated OOA–TVFEMD–DMO–GRU prediction model and validates it on industrial data. Under a sample entropy minimization criterion, OOA globally optimizes the TVFEMD bandwidth threshold and B-spline order so that the decomposed IMFs have clearer modal boundaries, higher signal-to-noise ratio and can be reorganized into high-, medium- and low-frequency Co-IMFs, with VMD further enhancing key intermediate frequency modes when needed. DMO then performs swarm intelligence search over GRU hyperparameters (such as batch size and maximum epochs), improving convergence and generalization compared with manual tuning. Only using decomposition-side optimization (OOA–TVFEMD–GRU) or only using prediction-side op-timization (TVFEMD–DMO–GRU) brings limited gains, whereas the joint scheme effectively mitigates mode aliasing, noise amplification and local optima. the final model achieves R^2 = 0.961, RMSE = 0.137, MAE = 0.110 and MAPE = 0.55% on the industrial test set, out-performing TVFEMD–GRU, TVFEMD–DMO–GRU and OOA–TVFEMD–GRU, and maintaining stable accuracy in peak and working-condition transition segments, which demonstrates its suitability for online rolling prediction and engineering deployment in intelligent sintering systems.

References

[1] Jianlin Lin, DeJin Kong. Study on “Industrial Internet” and “Industrie 4.0”[J]. Strategic Study of CAE, 2015, 17(07):141-144.

[2] Kesha Guo. the Trend of Industrial Structure Adjustment and Upgrading in China and the Thoughts of Policy Adjustment during the “14th Five-Year Plan”[J]. China Industrial Economics, 2019, (07):24-41.

[3] KoŜtial I, Dorĉák E, Terpak J. Optimal

control of the sintering process [J]. IFAC Proceedings Volumes, 2005, 38(1):280-285.
[4] Fumin LI, Jucai Hou, Xiaojie Liu, et al. Development status and prospect of smart manufacturing of sintering system [J]. Sintering and Pelletizing, 2024, 49(01):27-34.
[5] Xuefeng Zhang, Jingjing Tang, Liusong Huang, et al. Research progress of intelligent sintering prediction technology based on big data [J]. Metallurgical Industry Automation, 2024, 48(04):64-76.
[6] Shanwen Xu. End-point Control Method and System Research of Sintering Process [J]. Metallurgical Power, 2023, (03):99-102+107.
[7] Bing Xu, Shilai Huang, Yanyan Yang, et al. Development and Application of a Balanced Material Flow Control Model for Sintering [J]. Hebei Metallurgy, 2021, (10):54-56+84.
[8] Xu Jiang. Application of intelligent analysis by big data in blast furnace ironmaking [J]. Modern Transportation and Metallurgical Materials, 2023, 3(01):83-88.
[9] Lingkun Chen. Analysis of intelligent control system used for high efficiency smelting in blast furnace [J]. Metallurgical Industry Automation, 2021, 45(03):2-10.
[10] Patisson F, Bellot J P, Ablitzer D. Study of moisture transfer during the strand sintering process [J]. Metallurgical Transactions B, 1990, 21(1):37-47.
[11] Ren Y, Huang C, Jiang Y, et al. Neural network prediction model for sinter mixture water content based on KPCA-GA optimization [J]. Metals, 2022, 12(8):1287.
[12] Rahman R, Otridge J, Pal R. IntegratedMRF: random forest-based framework for integrating prediction from different data types [J]. Bioinformatics, 2017, 33(9):1407-1410.
[13] Jie Hu, Sheng Du, Min Wu, et al. Advanced Control and Intelligent Optimization of Raw-materials Preparation Process in Ironmaking [J]. Information and Control, 2018, 47(04):411-420.
[14] Fubin Wang, Rui Wang, Chen Wu. Short-Term Prediction of Sintering State Based on Improved Random Forest [J]. Laser & Optoelectronics Progress, 2022, 59(18):382-388.
[15] Wang S H, Li H F, Zhang Y J, et al. A hybrid ensemble model based on ELM and improved AdaBoost. RT algorithm for predicting the iron ore sintering characters [J]. Computational intelligence and neuroscience, 2019, 2019(1):4164296.
[16] Wei Chen, Pengfei Wu, Baoxiang Wang, et al. Intelligent optimization system of burden structure in the whole process of sintering-blast furnace iron-making [J]. Sintering and Pelletizing, 2020, 45(05):8-13+24.
[17] Tao Liu. Improving the quality of sintered ore under variable raw material conditions [J]. Hebei Metallurgy, 2022(007):000. DOI:10.13630/j. cnki. 13-1172.2022.0706.
[18] Zengchao Tian, Wenlong Zhan, Jiangya Kang, et al. Development of DNN-based sintering endpoint prediction and intelligent adjustment system [J]. Sintering and Pelletizing, 2023, 48(06):101-108.
[19] Liu S, Liu X, Lyu Q, et al. Comprehensive system based on a DNN and LSTM for predicting sinter composition [J]. Applied Soft Computing, 2020, 95:106574.
[20] Jinyang Wang, Zhaoxia Wu, Zhongzheng Li, et al. Prediction Model of Burning Through Point Based on JITL-XGBoost [J]. Journal of Northeastern University(Natural Science), 2025, 46(02):28-34+41.
[21] Yueyang Luo, Bocun He, Xinmin Zhang, et al. Research on multi-step prediction, method for sintering burn-through point prediction based on spatio-temporal encoding task decoding [J]. Metallurgical Industry Automation, 2025, 49(02):43-52.
[22] Zhen Zhang, Xiaojie Liu, Song Liu, et al. Research on combined forecasting model of return fine amount based on deep learning [J]. Sintering and Pelletizing, 2023, 48(01):57-66. DOI:10.13403/j. sjqt. 2023.01.009.
[23] Yuan Ding, Bin Wang, Jinchong Yan, et al. Prediction of burning through point based on two-dimensional interval autoregressive model [J]. Sintering and Pelletizing, 2017, 42(03):1-6+15.
[24] Cao W, Zhang Y, She J, et al. A dynamic subspace model for predicting burn-through point in iron sintering process [J]. Information Sciences, 2018, 466:1-12.
[25] Yan F, Zhang X, Yang C, et al. Data‐driven modelling methods in sintering process: Current research status and perspectives [J]. the Canadian Journal of Chemical Engineering, 2023, 101(8):4506-4522.
[26] Xinpeng Qu, Chunjie Yang, Jinsong Meng,

et al. Digital twin driven sintering endpoint prediction method [J]. Metallurgical Industry Automation, 2024, 48(S1):12-17.

[27] Chen X, Lan T, Shi X, et al. A semi-supervised linear–nonlinear least-square learning network for prediction of carbon efficiency in iron ore sintering process [J]. Control Engineering Practice, 2020, 100:104454.

[28] Dehghani M, Trojovský P. Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems [J]. Frontiers in Mechanical Engineering, 2023, 8:1126450.

[29] Liu, B., Cai, G., Qian, J., Song, T., & Zhang, Q. (2023). Machine Learning Model Training and Practice: A Study on Constructing a Novel Drug Detection System. International Journal of Computer Science and Information Technology, 1(1), 139-146.

