Prediction of Sintering Flue Gas Based on TCN Convolutional Neural Network

Weimin Yin[1], Jiayi Zhang[1], Jiaming Xu[2], Yanbo Hu[1], Tonghui Zhao[1], Junhu Wang[1], Jie Li[3,*]
[1]CHINA MCC22 GROUP CORPORATION LTD, Tangshan, China
[2]College of Science, North China University of Science and Technology, Tangshan, China
[3]College of Metallurgy and Energy, North China University of Science and Technology, Tangshan, China
**Corresponding Author*

Abstract: The CO concentration in the sintering flue gas during the sintering process is an important indicator reflecting the combustion state and permeability. Accurate short-term prediction of CO concentration helps in process stability and anomaly detection. However, in industrial environments, numerous sensor variables, significant data noise, and operational drift complicate high-dimensional modeling, making rapid deployment challenging. This paper proposes a lightweight short-term prediction framework based on causal dilated convolutions (TCN) for sintering flue gas data with a two-minute sampling interval to predict the target variable, cleaned_co. First, based on process mechanisms, relevant variables related to flue gas and exhaust are aggregated and selected, forming an input set with only 11 features. Next, multivariate sequences are constructed into supervised samples using a sliding window approach, employing causal convolutions to prevent future information leakage, and dilated convolutions to expand the temporal receptive field, thereby capturing the temporal dependencies between combustion and airflow. Experiments on real-world datasets validate the effectiveness of the proposed method. the results demonstrate that the model tracks the overall trend of CO concentration well within a 90-minute historical window, achieving stable prediction performance in the normal operating range. This method provides a concise and feasible technical approach for online modeling and rapid deployment of the sintering process, with fewer features.

Keywords: CO Concentration; Short-Term Forecasting; Industrial Sensor Variables; Causal Dilated Convolution (TCN)

1. Introduction

Sintering is a critical front-end process in steel metallurgy, and its combustion intensity, layer permeability, and heat and mass transfer processes directly affect the quality of sintered ore, fuel consumption, and pollutant emissions levels[1] . In the long process and highly coupled conditions of sintering production, achieving online prediction and early warning of key process indicators without incurring additional hard measurement costs has been an important research direction in process control and intelligent manufacturing[2] . In recent years, a relatively systematic research framework has been formed around mechanism-based modeling and data-driven modeling of the sintering process, with a particular emphasis on modeling robustness and deployability under varying operating conditions[3] .

Among the monitoring indicators for sintering flue gas, carbon monoxide (CO) concentration is a critical signal that reflects the combustion state and the completeness of gas-solid reactions: when there is local oxygen deficiency, incomplete combustion, or uneven airflow distribution in the material layer, CO concentration often increases and is closely related to the movement of the combustion zone, permeability, and exhaust system. Atmospheric CO primarily originates from industrial flue gas emissions, and its generation mechanism is mainly attributed to the incomplete combustion of fossil fuels[4] . As the core link in the steel production chain, iron ore sintering emits approximately 50 to 60 million tons of CO annually, leading to about 70% of the steel industry's annual pollution tax being associated with CO emissions. With the gradual implementation of CO pollutant regulation policies and the advancement of the "dual carbon" strategy, carbon reduction has become a primary policy goal across industries. Pei Yuandong et al. [5] , by analyzing the generation and distribution patterns of CO in the sintering process, emphasized that achieving CO reduction under current sintering conditions requires addressing both source control and process optimization. Li Qiankun et al. [6] conducted industrial trials on a sintering machine and, through enhanced process control, achieved a significant reduction in CO emission concentration in a 550 m² sintering machine, with a decrease of 782 mg/Nm³. Li Jie et al. [7] summarized the generation mechanisms of CO in sintering flue gas, including incomplete combustion of fuel and reduction reactions, and proposed source reduction strategies by analyzing emission patterns.

Regarding CO emissions and generation mechanisms, influencing factors, and control strategies in sintering, there have been specialized reviews and experimental simulations in the metallurgical field, pointing out significant correlations between CO and oxygen supply levels, fuel combustion, material layer structure, and airflow organization[8-10] . CO can also serve as an important basis for process state diagnosis and emission control. Meanwhile, industrial sites commonly use online monitoring and CEMS systems to continuously collect sintering flue gas data, providing a data foundation for data-driven prediction[11-14] .

From an engineering application perspective, achieving minute-level short-term CO prediction could help operators identify early signs of combustion anomalies, deterioration of permeability, and fluctuations in the exhaust system. This would provide forward-looking information for adjusting parameters such as air volume, ignition, and material moisture, thereby reducing the risk of amplified fluctuations caused by delayed adjustments[15-16] . In current intelligent sintering research, typical tasks include "end combustion zone position (e. g., BTP) prediction" and "quality soft measurement (e. g., composition, FeO, etc.)": for example, machine learning-based BTP prediction systems and decision rule construction have been validated in sintering process control; some studies have also proposed intelligent control frameworks aimed at BTP. In quality soft measurement, deep models like DNN and LSTM have been applied to sintering composition prediction, combined with feature selection and anomaly handling; and semi-supervised time series feature extraction models focused on sintering quality have also been explored[17-19] . These studies show that using multivariate process time series to construct "soft measurement/soft prediction" models is a feasible and effective technical approach. Although numerous sensors are typically deployed in sintering sites, actual modeling faces the following challenges:

(1) High-dimensional and complex correlation structure: the raw variables include valve positions, flow rates, pressures, temperatures, material layer status, and multi-stage airbox information. Directly using high-dimensional inputs leads to model complexity, difficult parameter tuning, and high deployment costs.

(2) Widespread noise and outliers: Sensor drift, missing data, and transient anomalies can introduce heavy-tail errors, causing the model to overfit to a small number of outliers during training, which affects generalization.

(3) Non-stationarity of operating conditions: Changes in raw materials, moisture, and operating strategies can lead to data distribution shifts over time, requiring the model to have certain robustness and transferability.

(4) Temporal dependencies and lag effects: CO response to exhaust, air supply, material layer, and moisture exhibits delays and thermal inertia, necessitating the model to capture multi-scale dependencies in the time dimension.

To balance prediction performance with engineering deployability, a minimal input+lightweight deep model approach is chosen: On one hand, key variables are selected and aggregated based on process correlations to minimize the number of features; on the other hand, a Temporal Convolutional Network (TCN) is used for short-term prediction. TCN offers advantages such as parallel computation, stable training, controllable receptive fields, and ease of deployment, making it suitable for industrial time-series tasks.

The main contributions of this paper include:

Proposal of a minimal feature construction strategy for CO prediction in sintering flue gas, aggregating numerous sensor variables into 11 key operational features to reduce data dimensionality and deployment complexity.

Development of a sliding window-based supervised learning framework with causal dilated convolutions to prevent future information leakage, enabling minute-level short-term prediction.

Visualization and error distribution analysis revealing that the model fits well in the normal range but exhibits amplitude compression in the peak segments, providing a basis for future improvements (such as robust loss, peak segment indicators, and event threshold evaluation).

2. Research Methodology and Model Development

2.1 Task Definition and Data Organization

Let the multivariate sensor sequence be denoted as $X_t \in \mathbb{R}^F$,and the target variable cleaned_co be denoted as y_t ,This paper adopts a sliding window approach to construct supervised samples:the input consists of the past L records(each representing 2 minutes),corresponding to 6 minutes of historical data, and the output is the cleaned_co value H steps ahead:

$$Input:[X_{t-L+1},\dots,X_t]\rightarrow Target:y_t+H \qquad (1)$$

To ensure a realistic industrial prediction scenario, the data is split into training, validation, and test sets in chronological order. Standardization of input features is performed using only the training set statistics.

2.2 TCN Convolutional Neural Network Model

To model industrial time-series dependencies, this paper adopts the Temporal Convolutional Network (TCN) architecture for prediction. the model flowchart is shown in Figure 1. TCN is a convolutional neural network architecture designed specifically for sequence modeling, where its core is composed of multiple stacked 1D convolution blocks. Causal convolutions are employed to ensure that the output at any given time depends only on the current and past inputs, thus avoiding future information leakage. To capture longer-range dependencies, dilated convolutions are used, with the dilation rate (e. g., 1, 2, 4) progressively increasing at each layer, thus enhancing the temporal receptive field without significantly increasing computational cost.

Causal convolution is the foundation of TCN, ensuring that the model's output at time step t only depends on the information from the input sequence at time steps t and earlier. In a standard 1D convolution, without padding, the output may depend on future inputs. To ensure causality, TCN adds zero-padding to the left side of the convolution kernel, making the output dependent only on past and current inputs. Each convolution block consists of two layers of causal dilated convolutions with residual connections.

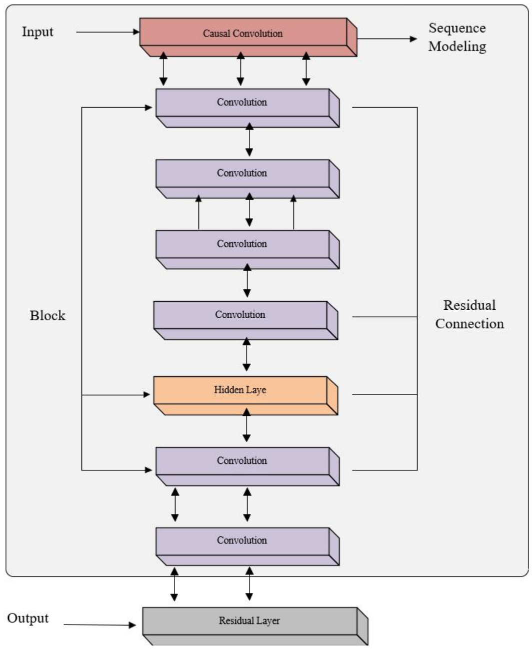

Figure 1. Overall architecture of the TCN

Assuming the input sequence is $x=(x_1,x_2,...,x_T)$,the convolution kernel size is k,and the weights are $w=(w_1,w_2,...,w_k)$,the output of the causal convolution is given by:

$$y_t=\sum_{i=1}^{k} w_i \cdot x_{t-(k-i)} \qquad (2)$$

To capture long-range dependencies without increasing network depth or the number of parameters, TCN introduces dilated convolutions. By skipping certain input elements, dilated convolution enlarges the receptive field (the range of input values the model can consider). the dilation factor d controls the sampling interval of the convolution kernel. Typically, in TCN, the dilation factor grows exponentially across network layers (e. g., 1, 2, 4, 8,...), allowing the shallow layers to capture local patterns and deeper layers to capture global patterns. the receptive field size increases exponentially with the number of layers, enabling the model to handle very long sequences.

TCN consists of multiple residual blocks stacked together, with each residual block containing two dilated causal convolution layers, an activation function (e. g., ReLU), weight normalization, and dropout (for regularization). the residual connections help train deep networks and mitigate the vanishing gradient problem. the entire TCN model can be represented as a stack of residual blocks, with the final regression output produced using the hidden state at the last time step:

$$\hat{y}_{t+H} = f_{\theta}([X_{t-L+1}, \dots, X_t]) \quad (4)$$

2.3 Feature Construction

Considering that sintering CO concentration is highly correlated with the flue gas flow field, oxygen supply, exhaust strength, material layer, and moisture, this paper aggregates and selects features from the original sensor data set to construct the feature set, as shown in **Table 1.**

Let the input be the windowed time-series block $H \in \mathbb{R}^{B \times C \times T}$ (batch size B, number of variables C, feature length T).

Table 1. Characteristic Variable

Number	Feature
1	Flue o2 mean
2	Flue temp mean
3	Flue negp mean
4	Mainfan flow
5	Mainfan press mean
6	gas_flow_col
7	Sintering circulating fan inlet flow detection
8	Sintering machine speed L1 setting
9	Bed thick mean
10	Moist col
11	Water total

3. Experimental Results and Analysis

3.1 Experimental Setup

The model training uses Mean Squared Error (MSE) as the loss function, and the optimal model parameters are selected based on the validation set. Considering that industrial data often exhibits heavy-tail distributions and spikes, the formulas are as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} | X_{\mathrm{obs}} - X_{\mathrm{model}} |}{n} \quad (5)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(X_{obs} - X_{model} \right)^2}{n}} \quad (6)$$

In addition to MAE and RMSE, it is recommended to use R^2(coefficient of determination) to reflect the explanatory power and comparability, and WAPE as a commonly used relative error metric in industry to avoid the instability of MAPE in low-value ranges.

3.2 Performance Comparison

The experimental dataset used in this study is sourced from the actual operation process of a 260 m² sintering machine in a large steel enterprise. the production line is equipped with an advanced Distributed Control System (DCS) and real-time online monitoring equipment, enabling precise continuous monitoring and data recording of key sintering process indicators, such as temperature, pressure, flow, and composition. the data is exported from the historical database, covering the time period from July 1, 2025, to July 30, 2025, with a total of 30 days of complete production records.

To ensure the richness and comprehensiveness of the model training samples, the sampling interval is set to 2 minutes after removing the time delay effects, and data during equipment maintenance and downtime is excluded, resulting in 20, 563 raw sample records. These samples cover hundreds of feature columns, including raw material properties, sintering process state parameters, and operation control parameters. To enhance the accuracy and robustness of the AI prediction model, a detailed data preprocessing step is carried out to effectively remove outliers and missing values, thus minimizing the negative interference of these noises on the prediction performance of CO concentration in sintering flue gas. This ensures the reliability of data quality and provides a solid foundation for subsequent machine learning model training.

Data preprocessing includes: Time alignment and sorting: Parsing the time column as datetime and sorting it in ascending order. Numerical conversion and missing value imputation: Converting input features and targets to numerical values and using forward filling (ffill) to handle short-term missing values. Valid sample exclusion: Removing samples with remaining missing values. Training set standardization: Standardizing input features using the training set mean and variance through z-score normalization to avoid data leakage.

For feature construction, to balance prediction performance and engineering deployability, this study aggregates numerous raw sensors into 11 key operational features, including:

Main duct: Average O_2, average temperature, average negative pressure.

Main exhaust fan: Sum of north and south inlet flow, average inlet pressure.

Flue gas flow: Desulfurization inlet flue gas flow (using exhaust outlet flow as a substitute when missing).

Process state: Circulating fan inlet flow detection, machine speed setting, material layer thickness (average of 1-5), post-second mixing moisture, total water addition (first+second mixing).

3.3 Influence of Window Length L

The window length L affects the amount of historical information available to the model. **Table 2** presents the comparison of MAE and RMSE on the test set

Table 2. MAE/RMSE

L(records)	Historical length	MAE	RMAE
15	30min	850.606	1028.372
30	60min	580.16	803.456
45	90min	463.115	671.707

When L increases from 15 to 45, the error significantly decreases, indicating that the short-term evolution of cleaned_co is influenced by thermal inertia and transport inertia, requiring a longer historical context. However, when L continues to increase (to 60/90), the error increases, suggesting that excessively long history introduces operational drift and irrelevant fluctuations, making the learning target non-stationary, which reduces generalization.

3.4 Comparative Experiment Analysis

As shown in Figure 2, a comparison is made between the Temporal Convolutional Network (TCN) and Convolutional Neural Network (CNN) for the sintering flue gas cleaned_co prediction task. the experiment used industrial furnace operation data, selecting the first 1, 500 samples of the test set for analysis, with a 2-minute prediction step (H=1) and a sliding window length of 90 minutes (L=45).

From the comparison chart (Figure 1), it is evident that the TCN model provides smoother predictions, closely aligning with the true values, especially during dramatic changes in the sintering flue gas signal. This suggests that TCN is effective in capturing the temporal dependencies in the time series data. In contrast, the CNN model shows significant deviations, especially in capturing the peaks and valleys of the signal, indicating its limitations in processing time series data and its inability to accurately capture dynamic changes.

According to the evaluation metrics, the TCN model outperforms the CNN model in RMSE, MAE, and R^2, further verifying its superior performance in predicting future values of sintering flue gas based on historical data. While the CNN model can capture some patterns, it performs poorly in capturing more complex dynamic changes, resulting in lower prediction accuracy.

This analysis indicates that model selection is crucial for time-series prediction tasks. Models specifically designed to handle temporal dependencies, such as TCN, perform better than traditional CNN architectures in handling complex time-related signals.

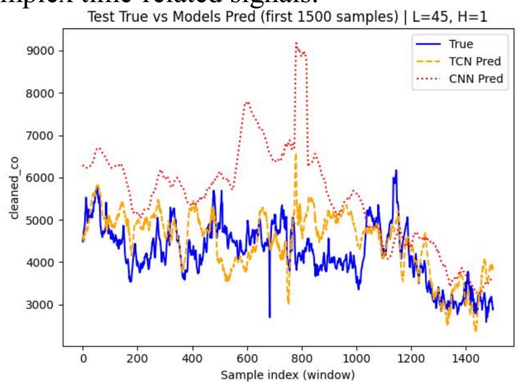

Figure 2. Comparison Chart of TCN and CNN Results

3.5 Visualization Result Analysis

Furthermore, Figure 3 shows a scatter plot of the true CO concentration values versus the predicted values. the x-axis represents the true values, and the y-axis represents the predicted values. the points are distributed around the reference line y=x showing a clear linear trend. This indicates that the TCN model successfully captured the main patterns in the data, especially within the 2000–5000 concentration range, where the predicted and true values are highly consistent.

As shown in Figure 4, the residual histogram further quantifies the distribution characteristics of the prediction error. the x-axis represents the residual (predicted value minus true value), and the y-axis represents the sample count. the distribution is approximately bell-shaped, with the peak near zero, and the count reaches about 500. This reflects the model's unbiased nature, with most prediction errors falling within a controllable range.

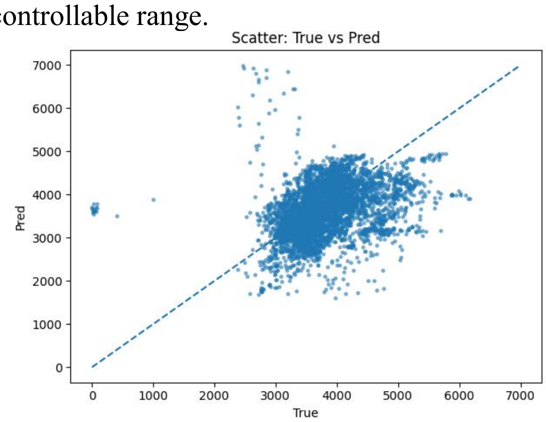

Figure 3. Scatter Plot of True vs. Predicted Values

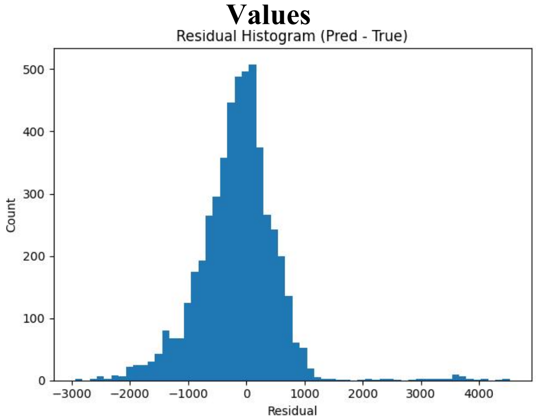

Figure 4. Histogram of Residual Distribution

4. Conclusion

This study develops a "minimal feature—lightweight model" approach for minute-level short-term prediction of CO in sintering flue gas based on two-minute sampled data. By aggregating numerous raw sensor data points according to process mechanisms and using only 11 key features (such as duct O_2/temperature/negative pressure, exhaust and flue gas transport, machine speed, material layer thickness, moisture, and water addition), stable predictions were achieved, demonstrating the strong representational power of this feature set for CO evolution.

The TCN model, designed to avoid future information leakage through causal dilated convolutions, successfully captured the dynamic dependencies of combustion and transport processes with a multi-scale temporal receptive field. the experimental results show that the window length significantly impacts performance, following a pattern of "improvement followed by degradation" from short to long historical windows. the optimal performance was achieved with L=45 (about 90 minutes), with a test MAE of 463.115 and RMSE of 671.707. This indicates that the short-term changes in cleaned_co depend not only on recent fluctuations but also on longer time scales, such as thermal and flow inertia. However, excessively long windows introduce operational drift and irrelevant information, reducing generalization. Further visualization analysis shows that the model follows the overall trend of CO well, with the error concentrated near zero, reflecting a good fit to normal operating conditions. the prediction curve was smoother than the true values, with systematic underestimation in the peak value sections and long-tailed residuals, suggesting that peak condition samples are sparse, the MSE objective is biased toward the mean section, and that potential time delays and sensor misalignment issues remain key factors affecting accuracy.

Overall, this method achieves usable online prediction capabilities with minimal feature cost, providing a simple and effective modeling pathway for monitoring, early warning, and control in the sintering process. It also lays the foundation for further improvements in peak prediction performance through robust loss, high-value segment weighting, or threshold event evaluation. Although this method achieves usable short-term prediction of CO with minimal features, limitations remain, such as the amplitude compression and underestimation of peak values in the prediction curve. This highlights the issues of sparse peak condition samples and the MSE objective's preference for the mean section. the long-tailed residual distribution suggests that large error events caused by sudden changes in operating conditions or sensor anomalies still exist.

Acknowledgments

Fund project: The funding projects are as follows: Enterprise-commissioned horizontal project: "Collaborative Control Technology and Equipment for Multi-pollutant Treatment in Metallurgical Flue Gas (20250213). " Supported by Project of Yanzhao Iron and Steel Laboratory: "Research on Collaborative Matching Mechanism and Quality Control of Sintering Flue Gas CO Treatment Technology (25364004D). "

References

[1] He, K., Li, H., Kang, J., et al. (2025). "Pilot-scale practice of catalytic oxidation for CO in sintering flue gas. " Sintering and Pelletizing, 50(05), 102-108. DOI: 10.13403/j. sjqt. 2025.05.084.

[2]. Sun, Q., Hao, X., Li, J., et al. (2025). "Mechanism of CO and NO synergistic control in the sintering process. " China Metallurgy, 35(10), 118-129. DOI: 10.13228/j. boyuan. issn1006-9356.20250285.

[3] Wang, J., et al. CO reduction in sintering flue gas by CFD-ML for process parameters optimization. Journal of Cleaner Production, 2025.

[4]. Long, H., Ding, L., Zhao, H., et al. (2023). "Progress in CO emission reduction in typical steel production flue gases. " Steel, 58(08), 1-12+24. DOI: 10.13228/j. boyuan. issn0449-749x. 20230132.

[5] Pei, Y., Liao, J., Zhang, J., et al. (2019). "Discussion on CO reduction in iron ore sintering process. " Sintering and Pelletizing, 44(01), 69-73. DOI: 10.13403/j. sjqt. 2019.01.016.

[6] Li, Q., Li, G., Yin, G., et al. (2019). "Study and practice of CO emission reduction on a 550 m² sintering machine at Zhongtian Steel. " Sintering and Pelletizing, 44(04), 70-73. DOI: 10.13403/j. sjqt. 2019.04.065.

[7] Li, J., Xu, R., Zhang, W., et al. (2025). "Current status and prospects of CO emission reduction in sintering flue gas. " China Metallurgy, 35(07), 109-120. DOI: 10.13228/j. boyuan. issn1006-9356.20250135.

[8] Yang, C., et al. Prediction and optimization of flue pressure in sintering process based on machine learning algorithms., 2025.

[9] Zhu, T. Y., et al. Numerical simulation of CO emission in a sintering pot under sintering flue gas recirculation (SFGR). Chemical Engineering Journal, 2023.

[10] Guo, X., et al. Optimizing sintering air volume for enhanced lean gas (CO, H_2, CH_4) generation in sintering flue gas. Scientific Reports, 2025.

[11] Sun, J., et al. Intensive carbon combustion in a sintering packed bed via surface steam spraying. Fuel, 2023.

[12] Liu, S., et al. A Prediction System of Burn-through Point Based on GBDT Algorithm and Decision Rules. ISIJ International, 2019.

[13] Chen, X., et al. Burn-through point prediction and control based on multi-period dynamic spatio-temporal extraction., 2025.

[14] Li, X., et al. Prediction model of burn-through point with data correction based on feature matching and TCN-GRU., 2024.

[15] Li, Y., et al. Sintering Quality Prediction Model Based on Semi-Supervised Dynamic Feature Extraction Framework. Sensors, 2022.

[16] Li, Y., et al. A Soft Sensor Model of Sintering Process Quality Index Based on Multi-Source Data Fusion., 2023.

[17] Xu, Z., et al. Burn-through point prediction based on deep learning feature selection and ensemble learning., 2023.

[18] Yan, F., et al. Data-driven modelling methods in sintering process: current status and prospects. Canadian Journal of Chemical Engineering, 2023.

[19] Hu, M., et al. Review of Intelligent Modeling for Sintering Process Under Variable Operating Conditions. Processes, 2025.

