



OPEN Hybrid-driven modeling using a BiLSTM–AdaBoost algorithm for diameter prediction in the constant diameter stage of Czochralski silicon single crystals

Yu-Yu Liu^{1,2}, Ding Liu^{1,2}✉, Shi-Hai Wu^{1,2} & Yi-Ming Jing^{1,2}

During the preparation of electronic-grade silicon single crystals (SSC), accurately predicting the crystal diameter is crucial for obtaining high-quality crystals. In this paper, a hybrid-driven modeling method integrating Bidirectional Long Short-Term Memory network (BiLSTM) and Adaptive Boosting (AdaBoost) algorithm is proposed, aiming to improve the accuracy and stability of crystal diameter prediction in the medium diameter stage of the SSC growth by the Czochralski (CZ) method. First, the initial prediction of SSC diameter is performed using a mechanistic model and the prediction error is calculated; then, the time series data are processed using a BiLSTM network to generate the predicted values at each time point. Subsequently, the prediction results of the BiLSTM network are weighted and fused by the AdaBoost algorithm to obtain the final time series prediction output, and the prediction performance is further enhanced by iterative optimization. Compared with the traditional mechanistic model or a single data-driven model, this hybrid model retains the explanatory nature of the mechanistic model while also ensuring the accuracy of the data-driven model, which effectively overcomes the challenges posed by complex coupling and nonlinear problems.

Keywords Czochralski silicon single crystals, Mechanistic modeling, Hybrid drive modeling, Long short-term memory network, Adaptive boosting algorithms

Electronic-grade silicon is an outstanding semiconductor material that allows efficient regulation of conductivity between conductors and insulators, and it is widely used in integrated circuits, automobiles, communications, and other fields. Electronic-grade SSC are commonly prepared in industrial production by the CZ method, which ensures uniform crystal growth by heating high-purity polysilicon to near 1400 °C, melting it into a liquid state, and then immersing SSC seed crystals into the molten silicon, pulling it upward slowly and rotating the seed crystals^{1,2}. During this process, the solution temperature and lifting speed are precisely controlled to make the silicon atoms orderly arranged according to a certain crystalline phase, forming a high-quality SSC structure.

In the semiconductor SSC growth process, the main factors influencing diameter variations include the heater power and the crystal lifting speed³. Notably, although adjusting the crystal pulling rate can quickly alter the crystal diameter, its adjustable range is relatively limited. Moreover, frequent pulling rate adjustments can cause large fluctuations at the growth interface, increasing the likelihood of crystal breakage and defects. In contrast, controlling the crystal diameter by adjusting the heater power is a slower process, but it has a wider adjustment range and is less likely to cause instability at the growth interface. Therefore, constructing a SSC growth model that accurately describes the relationship between heater power and crystal diameter, and designing an efficient diameter control system have become important issues in realizing high-quality SSC production.

One type of model commonly used in crystal diameter prediction is mechanistic modeling. Mechanistic modeling is an effective method to understand and analyze the dynamic properties of complex systems by quantitatively describing and predicting system behaviors and processes through scientific principles and physical laws. This modeling approach is widely used in the fields of physics, chemistry, biology, and engineering, which helps to deeply resolve the intrinsic mechanisms of a system. In their study, Winkler et

¹School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China. ²Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing, Xi'an University of Technology, Xi'an 710048, China. ✉email: liud@xaut.edu.cn

al.^{4,5} took the relationship between the variation in the crystal radius at the meniscus and the tilt angle of the crystal as a starting point to construct fluid-dynamic and geometric models for the CZ crystal growth process, thereby circumventing the complexity of thermodynamic modeling. Subsequently, Rahmanpour et al.^{6,7} and Bukhari et al.⁸ thoroughly investigated and extended this work, and proposed a centralized parametric model (expressed in state-space form) incorporating the crystal growth kinetics and temperature kinetics, which further incorporated a simplified heat transfer model on the basis of the original hydrodynamic and geometrical model. In the context of the 300 mm CZ SSC growth process, Zheng et al.⁹ developed a centralized parameter model that integrates energy balance, mass balance, hydrodynamics, and geometric equations to enhance the understanding of CZ kinetics. This model was utilized for predicting the crystal radius and growth rate. For their investigation into the distribution parameter model concerning the CZ SSC growth process, Ng et al.^{10–12} introduced a model that incorporates both the crystal radius and temperature variation, allowing for a more precise representation of system dynamics and temperature fluctuations within the crystal. In general, for lumped-parameter and distributed-parameter models to be applicable to real-world silicon single crystal growth control systems, strict assumptions regarding model accuracy must be satisfied. Moreover, in many of these mechanistic models of crystal growth, numerous parameters are either unknown or lack sufficient precision, and some practical parameter values may vary significantly across different operating conditions. For example, in the case of distributed-parameter models for the CZ crystal growth process, a primary challenge lies in the need to numerically solve complex partial differential equations. This requires powerful solvers and substantial computational resources, which are often impractical on the limited computing hardware available for process control in industrial environments. Additionally, information derived from numerical models (as well as the inverse solutions required for control purposes) may not be obtainable in real time. Furthermore, considerable effort is needed to adapt distributed-parameter models to actual crystal growth system designs and to determine the physical parameters and boundary conditions required by the solvers. In summary, mechanistic modeling approaches for CZ silicon single crystal growth inevitably face certain limitations, including high modeling costs, the need for extensive domain knowledge, the difficulty of obtaining key parameters online, and excessive model simplification that compromises accuracy in representing the system.

Another commonly adopted strategy is data-driven modeling. As computer hardware continues to advance, data analysis and machine learning approaches have increasingly become essential tools for investigating complex systems. Liang et al.^{13,14} introduced a T-S fuzzy temperature identification algorithm aimed at modeling temperature changes during the growth of CZ SSC. Experimental results indicated that this fuzzy inference model achieves high accuracy in temperature prediction. Subsequently, Zhao et al.¹⁵ combined a T-S fuzzy model with an improved ant lion optimization (ALO) algorithm to establish a thermal model relating heating power to liquid surface temperature in the priming stage of CZ crystal growth. By applying the improved ALO algorithm to determine globally optimal parameters for the T-S fuzzy model, this method further enhanced the accuracy of predicting the liquid surface temperature. To expand deep learning applications in CZ SSC, Zhang et al.¹⁶ developed a data-driven model using a BP neural network to predict crystal diameter based on experimental data. Simulation outcomes showed not only high predictive accuracy but also strong stability, suggesting this approach can reliably guide crystal dimension control. Advancing these efforts, Zhang et al.¹⁷ introduced a deep learning model linking heater power to crystal diameter via an long short-term memory (LSTM) architecture, particularly effective for analyzing sequential data. Results from multiple simulations demonstrated that this LSTM-based model outperformed traditional techniques such as BP neural networks and support vector machines in predicting crystal diameter, indicating that LSTM integration can lead to more precise control of the crystal growth process. Although data-driven modeling offers advantages such as high prediction accuracy, ease of implementation, low cost, and no need to consider the inherent complexities of the crystal growth process, the resulting models are often black-box in nature, with limited interpretability and a lack of insight into the underlying dynamics of the CZ crystal growth process.

To harness the benefits of both mechanistic and data-driven models, we merge them in a way that the resultant model not only elucidates the fundamental physical processes but also maintains high predictive accuracy. Kato et al.¹⁸ developed a gray-box model aimed at forecasting the temperature gradient parameters of crystals at the solid–liquid interface, utilizing a mechanistic framework of the CZ SSC growth process, thereby establishing a groundwork for investigating the integration of data-driven approaches with mechanistic models in hybrid modeling techniques related to CZ SSC. Ren et al.¹⁹ introduced a Just-In-Time Learning–Stacked AutoEncoder–Extreme Learning Machine hybrid model, grounded in the concept of integrating data-driven and mechanistic elements, demonstrating its effectiveness in accurately predicting melt temperature and crystal growth rates. Wan et al.²⁰ proposed the mechanism–data-driven hybrid variable weighted stacked autoencoder random forest hybrid model, which was based on the mechanistic framework and industrial data, aimed at facilitating online monitoring of V/G. The distinctive feature of this hybrid-driven modeling strategy lies in its incorporation of the mechanistic model's explanatory capabilities alongside the generalization strengths of data-driven models, enhancing the model's robustness and adaptability, thus ensuring it better mirrors actual operational conditions. Although previous studies have explored data-driven approaches to predict crystal diameter using operational data, these efforts have primarily focused on establishing a direct mapping between process parameters (e.g., heater power) and crystal diameter, without considering the physical constraints provided by mechanistic models. The key distinction of this study lies in the seamless integration of mechanistic models—derived from energy balance, fluid dynamics, and geometric equations—with real production data through the BiLSTM–AdaBoost framework. Unlike purely data-driven methods, our hybrid model employs data-driven corrections to explicitly compensate for the residuals of the mechanistic model, thereby achieving both physical consistency and adaptability to process variations. This approach bridges the gap between theoretical modeling and industrial applicability, which has not been systematically addressed in prior work.

The main contributions of this study are as follows:

1. Application of Hybrid Modeling in Crystal Growth Prediction:

To address challenges such as nonlinearity, time delay, and difficulty in direct measurement of crystal diameter during the growth of electronic-grade semiconductor SSCs, a hybrid modeling approach combining mechanistic and data-driven models is proposed. The mechanistic model captures the dynamic behavior of crystal growth, while the data-driven model provides dynamic compensation for prediction results. This approach enhances prediction accuracy while maintaining model interpretability, offering theoretical support for quality control in SSC growth.

2. BiLSTM-AdaBoost-Based Prediction Algorithm:

To overcome the limitations of traditional single-model approaches, an ensemble learning framework combining Bidirectional Long Short-Term Memory (BiLSTM) networks and Adaptive Boosting (AdaBoost) is designed. Multiple comparative experiments (see Chapter 3) demonstrate that this algorithm significantly improves prediction accuracy and robustness compared to conventional methods, providing a new solution for parameter prediction in complex industrial processes.

3. Interdisciplinary Methodological Advancement in Semiconductor Crystal Growth:

By integrating control theory and machine learning, the proposed hybrid modeling framework offers a new perspective for research on semiconductor SSC growth. The method can be extended to related applications such as crystal defect detection and growth rate optimization, serving as a valuable reference for promoting intelligent development in semiconductor material manufacturing.

The structural content of this paper is organized as follows: Section "Proposed approach": A brief description of the mechanistic modeling of the SSC growth process and an in-depth introduction to the implementation details of the BiLSTM and Adaboost. Section "Heater to Crucible Thermal Radiation Transfer Rate for": Through simulation experiments, the effectiveness and superiority of crystal diameter prediction based on BiLSTM and Adaboost algorithm are fully demonstrated, and the model performance is comprehensively discussed and analyzed. Concluding part: to summarize and conclude the work done in this paper, emphasizing the importance and practical application value of the research.

Proposed approach

Figure 1 illustrates the proposed hybrid-driven modeling method for diameter prediction in the constant diameter stage of CZ SSC using BiLSTM and the AdaBoost algorithm. The method comprises two modules: The first module is about the mechanistic modeling of the SSC growth process. Based on the mechanistic model established by the hydrodynamic model, energy balance model and geometric model of the CZ SSC growth process, the historical data of the actual SSC growth process is inputted into the mechanistic model to obtain the preliminary diameter prediction data and the error data of the diameter.

The second module takes the error data in Module 1 and the historical data from the actual SSC growth process as inputs, sets the data of the corresponding step as the training samples, and sets them in proportion to the corresponding training and testing sets. Then, the data are normalized and modeled using BiLSTM network as a weak regressor to store the prediction results of each time step and calculate the weights. Finally, the weights of all the weak regressors are weighted and combined, and the final diameter time series prediction results are obtained after inverse normalization.

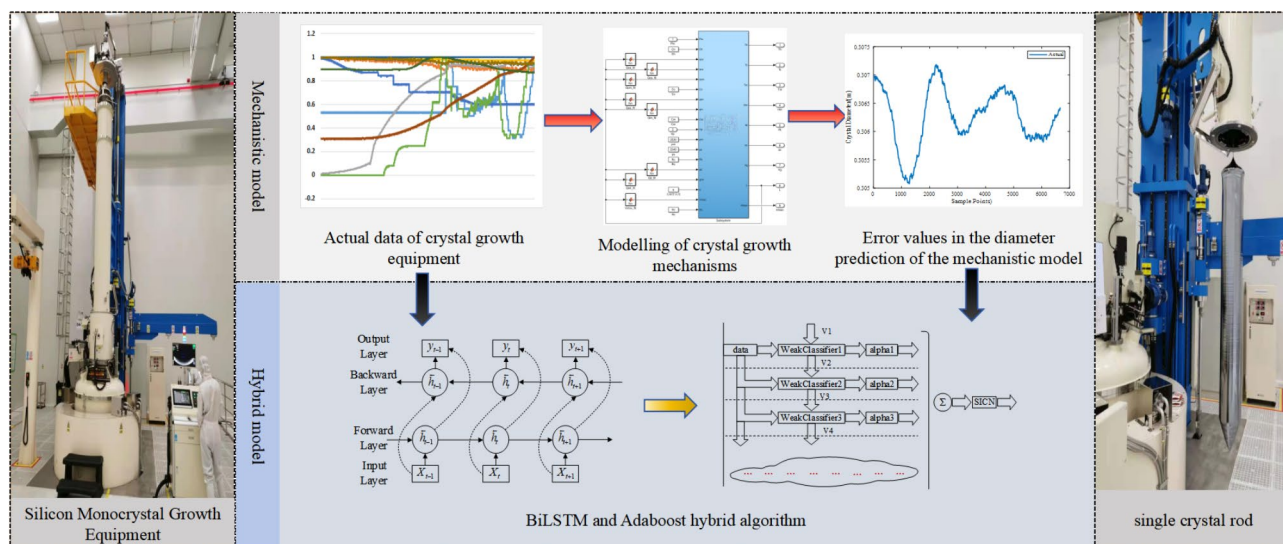


Fig. 1. Overall scheme of hybrid modeling.

Mechanistic modeling of SSC growth systems

Mechanistic modeling is a scientific and highly explanatory modeling method for scenarios with a good understanding of the internal laws of the system. For the large time lag and nonlinear process such as the modeling of the SSC preparation process of the direct drawing method, the mechanistic modeling is used with a good explanatory nature.

T_h represents the heater temperature, which controls the distribution of the thermal field and influences the silicon melt temperature and its melting state. T_c denotes the crucible temperature, which is crucial for ensuring the uniformity and stability of the melt. T_m indicates the melt temperature, a critical thermal parameter that impacts the melting, solidification, and the presence of crystal defects in silicon. H_m refers to the melt height, which plays a significant role in maintaining thermal balance and the stability of crystal diameter. R_i signifies the crystal radius, determining the diameter of the crystal rod and serving as an important indicator of process stability. α_c represents the crystal tilt angle; excessive deviation in the crystal axis offset angle adversely affects crystal quality. V_g denotes the crystal growth rate, which influences the rate of crystal formation; both excessively fast and slow growth can compromise quality and efficiency. L indicates the crystal length, which reflects the stability of growth and production efficiency. H_{men} represents the height of the meniscus, the height difference between the melt and the crystal, which affects both the growth shape and quality. V_p signifies the crystal pulling rate, impacting the balance between crystal quality and growth rate. P_{in} represents the heater power, which regulates temperature and the thermal field, thereby indirectly influencing melt conditions and crystal growth.

Thermal transfer model

It is assumed that the heat released by the graphite heater is completely absorbed by the quartz crucible, with no energy loss. Utilizing the law of conservation of energy, we can develop a heat transfer model for the growth of silicon single crystals^{7,21}. Figure 2 shows the crystal growth model.

1. Equation for heater temperature as a function of time $\dot{T}_h$:

$$\dot{T}_h = \frac{1}{C_h} (P_{in} - q_{hc}) \quad (1)$$

where the heater heat capacity C_h is:

$$C_h = S_h \times \rho_h \times V_h \quad (2)$$

2. Heater volume V_h is:

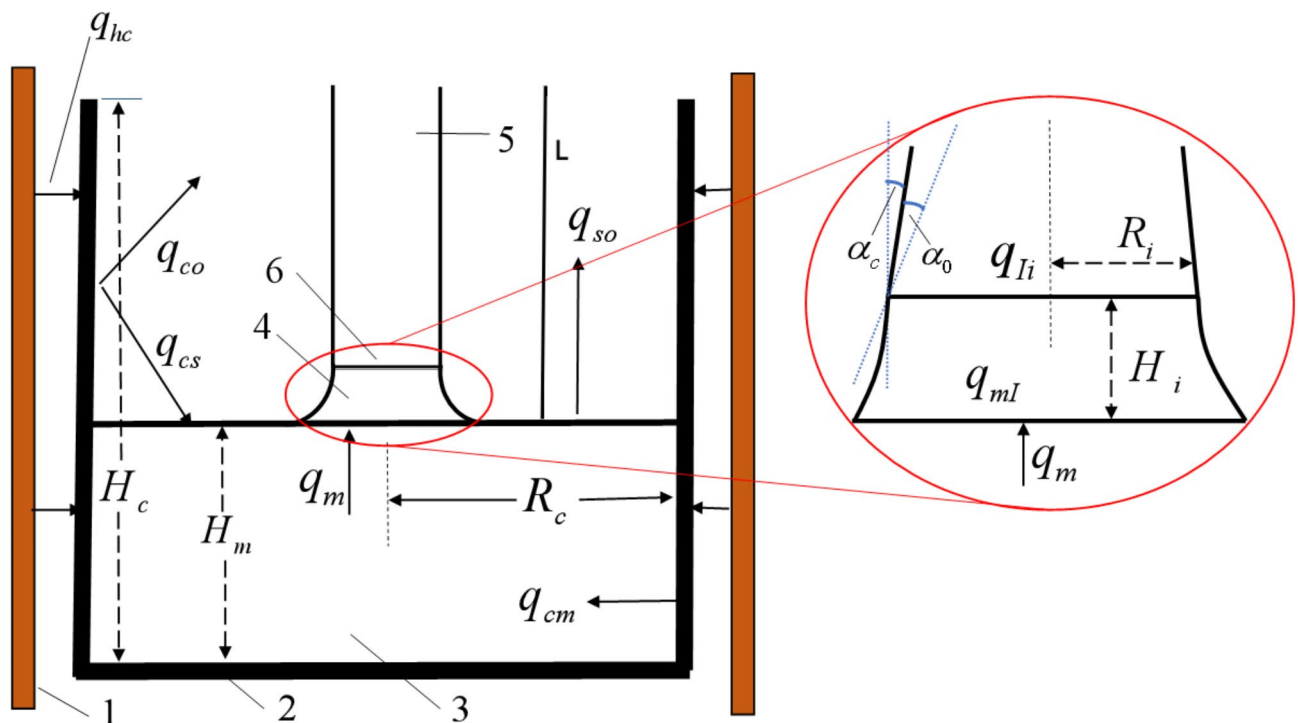


Fig. 2. Model of crystal growth set parameters.

$$V_h = \pi(R_{ho}^2 - R_{hi}^2) \times H_h \quad (3)$$

3. Heater to Crucible Thermal Radiation Transfer Rate q_{hc} for:

$$q_{hc} = A_c \times \sigma \times (T_h^4 - T_c^4) \quad (4)$$

The power of the heater is denoted as P_{in} , the specific heat capacity as S_h , the density as ρ_h , the outer diameter as R_{ho} , the inner diameter as R_{hi} , the height as H_h , the surface area of the crucible as A_c , the Stefan-Boltzmann constant as σ , the heater temperature as T_h , and the crucible temperature as T_c .

4. Equation for the variation of crucible temperature with time $\dot{T}_c$:

$$\dot{T}_c = \frac{1}{C_c} (q_{hc} - q_{co} - q_{cs} - q_{cm}) \quad (5)$$

where the crucible heat capacity C_c is calculated as:

$$C_c = S_c \times \rho_c \times V_c \quad (6)$$

5. Crucible volume V_c is calculated as:

$$V_c = \pi R_c^2 \times H_c \quad (7)$$

where q_{co} is the thermal radiation transfer rate from the crucible to the environment, q_{cs} is the thermal radiation transfer rate from the crucible to the melt, q_{cm} is the thermal conduction transfer rate from the crucible to the melt, S_c is the specific heat capacity of the crucible, ρ_c is the density of the crucible, and H_c is the height of the crucible;

6. Based on the energy and mass conservation of the melt, the equation for the change in melt temperature over time as $\dot{T}_m$:

$$\dot{T}_m = \frac{1}{C_m} (q_{cm} + q_{cs} - q_{so} - q_m) - \frac{\dot{H}_m T_m}{H_m} \quad (8)$$

7. where the method for calculating the melt heat capacity as C_m :

$$C_m = S_m \times \rho_m \times V_m \quad (9)$$

8. The method for calculating the melt volume as V_m :

$$V_m = \pi R_c^2 \times H_m \quad (10)$$

where q_{so} is the heat transfer rate from the melt to the surroundings, q_m is the heat transfer rate from the melt to the curved surface, $\dot{H}_m$ is the rate of change of melt height, T_m is the melt temperature, H_m is the melt height, S_m is the melt heat capacity, ρ_m is the melt density, and R_c is the crucible radius.

(4) At the solid–liquid interface of crystal growth, the growth rate of the crystal can be determined by calculating the difference between the heat flux from the melt to the growth interface (Q_{mI}) and the heat flux from the growth interface to the crystal (Q_{Ii}). This difference reflects the heat released by the crystal during the crystallization process. Therefore, based on the principle of energy conservation, we can derive the calculation formula for the crystal growth rate V_g ⁷.

$$V_g = \frac{Q_{Ii} - Q_{mI}}{H_f \rho_s} = \frac{q_{Ii} - q_{mI}}{H_f \rho_s \pi R_i^2} \quad (11)$$

where H_f is the latent heat of crystallization, ρ_s is the crystal density, q_{Ii} is the heat transfer rate from the solid–liquid interface to the crystal, q_{mI} is the heat transfer rate from the melt to the solid–liquid interface, and R_i is the crystal radius.

Geometric mode

1. Equation for the variation of melt height with time at the solid–liquid interface $\dot{H}_m$ ²²:

$$\dot{H}_m = -\frac{\rho_s R_i^2 (V_p - \dot{H}_{men})}{\rho_m R_c^2 - \rho_s R_i^2} \quad (12)$$

where V_p represents the velocity at which the crystal rises, $\dot{H}_{men}$ indicates the rate at which the meniscus height changes, ρ_m denotes the density of the melt, and R_c refers to the radius of the crucible.

2. The equation for the height of the curved lunar surface is H_{men} ²³:

$$H_{men} = a \sqrt{\frac{1 - \sin(\alpha_0 + \alpha_c)}{1 + \frac{a}{\sqrt{2}R_i}}} \quad (13)$$

From Eq. (13) it can be calculated that the

$$\dot{H}_{men} = \frac{\partial H_{men}}{\partial R_i} \dot{R}_i + \frac{\partial H_{men}}{\partial \alpha_c} \dot{\alpha}_c \quad (14)$$

here $\dot{R}_i$ is the capillary length, also known as the Laplace constant, which is related to the surface tension $\dot{R}_i$ and melt density of the meniscus. The value of $\dot{R}_i$ is $\dot{R}_i$, $\dot{R}_i$ is the crystal growth angle, and $\dot{R}_i$ is the crystal tilt angle²⁴.

3. Relationship between the amount of change in crystal radius with time $\dot{R}_i$ and the growth rate V_g :

$$\dot{R}_i = V_g \tan \alpha_c \quad (15)$$

4. Change in angle of inclination over time $\dot{\alpha}_c$ ²⁵:

$$\dot{\alpha}_c = \frac{V_p - V_{cruc} - C_{\alpha z} V_g}{C_{\alpha n}} \quad (16)$$

Among them

$$C_{\alpha z} = 1 - \frac{\rho_s R_i^2}{\rho_m R_c^2} + \left[\left(1 - \frac{R_i^2}{R_c^2} \right) \frac{\partial H_{men}}{\partial R_i} - \frac{2R_i H_{men}}{R_c^2} - \frac{a^2}{R_c^2} \cos(\alpha_0 + \alpha_c) \right] \tan \alpha_c \quad (17)$$

$$C_{\alpha n} = \left(1 - \frac{R_i^2}{R_c^2} \right) \frac{\partial H_{men}}{\partial \alpha_c} + \frac{a^2 R_i}{R_c^2} \sin(\alpha_0 + \alpha_c) \quad (18)$$

$$\frac{\partial H_{men}}{\partial R_i} = \frac{a^2 [1 - \sin(\alpha_0 + \alpha_c)]}{2\sqrt{2}R_i^2 \left(1 + \frac{a}{\sqrt{2}R_i} \right) S_\alpha} \quad (19)$$

$$\frac{\partial H_{men}}{\partial \alpha_c} = -\frac{a \cos(\alpha_0 + \alpha_c)}{2S_\alpha} \quad (20)$$

$$S_\alpha = \sqrt{[1 - \sin(\alpha_0 + \alpha_c)] (1 + a/\sqrt{2}R_i)} \quad (21)$$

where V_{cruc} is the crucible rise rate.

5. The change in crystal length over time as $\dot{L}$:

$$\dot{L} = V_g \quad (22)$$

Long short-term memory (LSTM) network

In the diameter prediction in the constant diameter stage of direct-drawn SSC, the input data required for model construction comes from acquisition and simulation with significant time-series characteristics. The LSTM model, due to its efficient ability to capture the time dependence, is able to deeply mine the temporal correlations in the input data, while adapting to the complex relationships between different features in the multi-input task, and exhibits superior modeling results^{21,26}.

As explained in^{22,27}, the LSTM model is a widely recognized deep learning approach for handling sequential data. It includes three key components—the input gate, the forget gate, and the output gate—as well as a cell

state. By discarding past information through the forget gate, integrating new information via the input gate, and determining the current hidden state through the output gate, the LSTM dynamically updates and transmits information within its cell state. Compared to conventional recurrent neural networks (RNN), LSTM more effectively capture long-term dependencies, thereby enhancing overall model performance.

RNN

A conventional RNN is a specific kind of neural network designed for handling sequential data. In contrast to feedforward neural networks, RNNs feature an internal recurrent structure that enables them to retain and leverage information from prior states while processing input sequences. This capability allows them to effectively capture temporal dependencies present in the data, as illustrated in Fig. 3.

The basic unit of RNN combines the current input with the hidden state of the previous moment through a circular connection to form a new hidden state. Its basic formula is as follows:

$$f_t = f(W_h \cdot h_{t-1} + W_x \cdot x_t + b) \quad (23)$$

where h_t is the hidden state at the current moment, h_{t-1} is the hidden state at the previous moment, x_t is the current input, W_h and W_x are the weight matrices, b is the bias, and f is usually a nonlinear activation function.

Gating mechanism

In order to improve the shortcomings of the traditional RNN, the LSTM introduces a gating mechanism that allows the network to dynamically select and manage the information at each time step by a combination of three gates to effectively capture long term dependencies, which control the flow of information by designing a series of “gates”. These gates serve to determine which information needs to be retained, updated, or exported, thus effectively managing the cell state and hidden state. This design makes LSTM more powerful and flexible when dealing with complex sequential data.

1. Amnesia Gate

The forgetting gate determines how much old information is discarded from the cellular state:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (24)$$

The output of the forgetting gate, denoted as f_t , varies between 0 and 1. The sigmoid activation function, represented as σ , maps the input to a range of 0 to 1. Additionally, W_f refers to the weight matrix, h_{t-1} indicates the hidden state from the previous time step, x_t represents the current input, and b_f is the bias term. Finally, the output f_t governs the extent of information retained from the prior cell state, C_{t-1} .

2. input gate

The input gate determines which new information will be added to the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (25)$$

where i_t is the output of the input gate; W_i is the weight matrix and b_i is the bias term.

3. output gate

The output gate determines the hidden state of the next step :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (26)$$

$$h_t = o_t * \tanh(C_t) \quad (27)$$

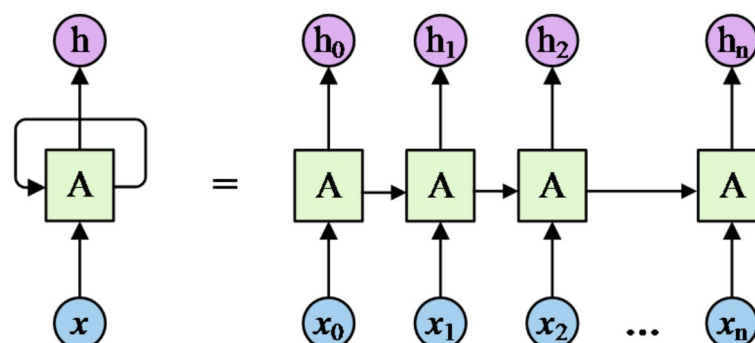


Fig. 3. Schematic diagram of RNN.

where o_t is the output of the output gate; W_O is the weight matrix and b_O is the bias term.

4. cell state

$$C_t = f_t * C_{t-1} + i_t * \overline{C_t} \quad (28)$$

where C_t is the weighted sum of the information retained by the previous cell state through the forget gate selection and the new information selected through the input gate.

At each time step, the LSTM's forget gate decides how much of the previous information to discard, based on the current input. Simultaneously, the input gate selects which essential details to incorporate. Finally, the output gate determines which data from the cell state forms the hidden state, which then becomes the model's output for the next layer. This synchronized process enables the LSTM to retain critical information over extended periods, effectively capturing long-term dependencies, as illustrated in Fig. 4.

BiLSTM

For diameter prediction in the constant diameter stage of CZ SSC, various types of data need to be linked to the past and future information related to the prediction point, and BiLSTM can better model complex feature interactions and dependencies for multiple input tasks such as simulated diameter, lifting speed, heater power, etc. in the training of the model^{28,29}.

The BiLSTM is an extended LSTM model that consists of two LSTM modules, one that reads from the forward to the time series (forward LSTM) and the other that reads from the backward to the time series (backward LSTM). This bi-directional structure enables the model to fully capture the contextual information and aims to improve the understanding of sequence data by considering both forward and backward information of the sequence.

The basic structure is a forward LSTM that needs to deal with the natural order of the sequence, from the first to the last time step ($x_1, x_2, \dots, x_T$), and a backward LSTM that needs to deal with the reverse order input sequence, from the last time step to the first time step ($x_T, x_{T-1}, \dots, x_1$). In this model, the input signal passes through the forward LSTM layer output $\vec{h}_t$ and the backward LSTM layer output $\overleftarrow{h}_t$, which together determine the values passed into the hidden layer, to obtain the output of BiLSTM y_t , which is updated with the formula.

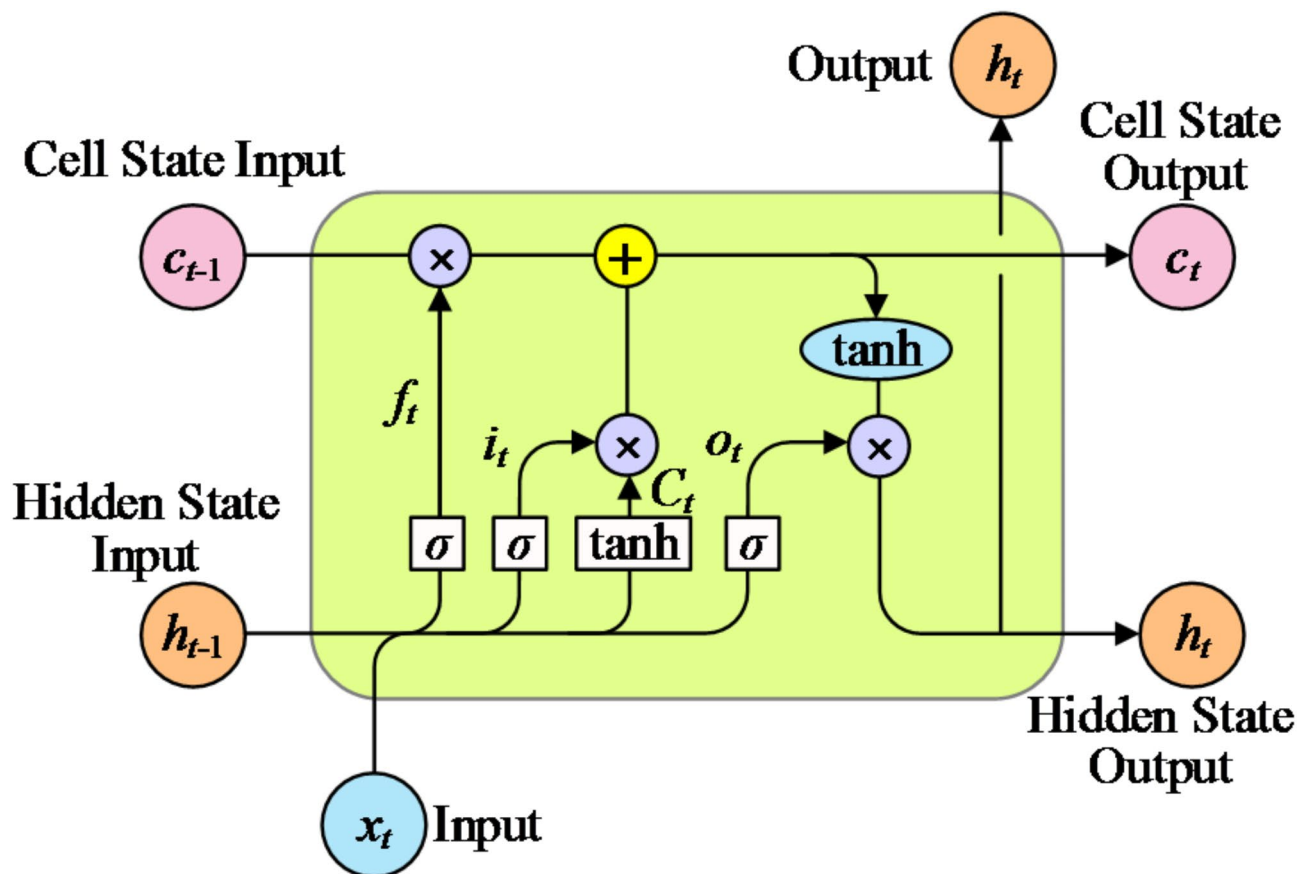


Fig. 4. Schematic diagram of LSTM network.

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \quad (29)$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t-1}) \quad (30)$$

$$y_t = \vec{W}\vec{h}_t + \overleftarrow{W}\overleftarrow{h}_t + b \quad (31)$$

The hidden states of the forward and backward LSTM are spliced and the output is obtained:

$$h_t^{BILSTM} = [h_t^{forward}, h_t^{backward}] \quad (32)$$

The output of BILSTM is used as input for subsequent tasks. Further processing can be performed using a fully connected layer, as shown in Fig. 5.

AdaBoost based prediction models

The growth process of direct-drawn SSC is complex, characterized by large hysteresis and high nonlinearity. Since a single BiLSTM model does not predict well when dealing with the multiple-input single-output case, it is particularly important to improve the model's ability to model complex multi-class data features in practical applications. Such improvement can help us to better predict the diameter of direct-drawn SSC, thus improving production efficiency and quality^{25,26,30,31}.

In the field of crystal diameter prediction, commonly used ensemble learning methods include Gradient Boosting, XGBoost, Random Forest, and AdaBoost. Based on the specific requirements of the Czochralski process for crystal diameter prediction, Table 1 summarizes and compares the advantages and disadvantages of these algorithms, providing a reference for selecting an appropriate model.

AdaBoost is a method of ensemble learning aimed at improving classifier performance. It creates a robust classifier by merging several weak classifiers, training each new model specifically on the mistakes made by the previous one. Through the modification of sample weights, AdaBoost highlights the samples that were misclassified, which ultimately leads to enhanced accuracy in the classification system. Therefore, we opt to use AdaBoost. The process of modeling can be summarized as follows:

1. Weight distribution:

The algorithm, for a given training dataset $T = \{(x_1, y_1), (x_2, y_2) \dots (x_t, y_t)\}$, initializes the weights of each sample by assigning the same weights: $w_i = 1/N$, such that the initial weight distribution of the training sample set $D_1(i)$:

$$D_1(i) = (w_1, w_2, \dots, w_N) = \left(\frac{1}{N}, \dots, \frac{1}{N}\right) \quad (33)$$

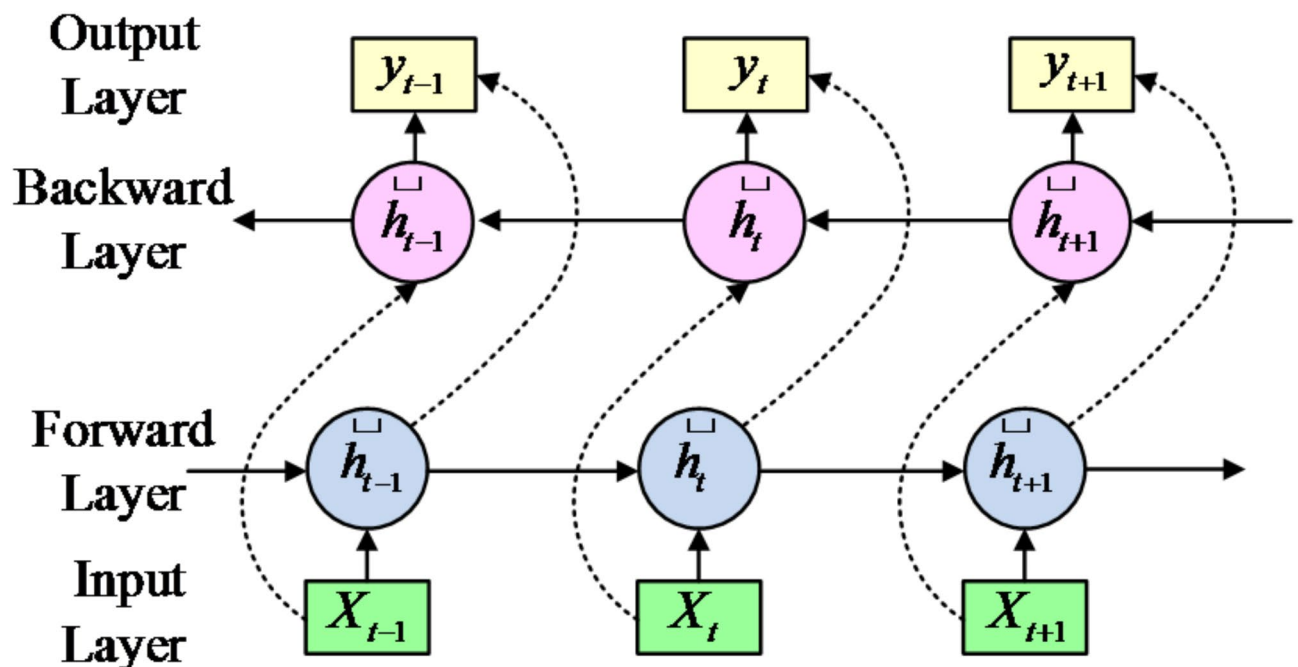


Fig. 5. Schematic diagram of BILSTM network.

Feature	AdaBoost	Gradient Boosting	XGBoost	Random Forest
Mechanism	Sequentially trains weak learners, reweights misclassified samples	Sequentially adds weak learners to minimize loss function	Optimized Gradient Boosting with regularization	Builds independent trees, aggregates via voting/averaging
Adaptability to Misclassified Samples	High, focuses on difficult-to-predict samples	Moderate, depends on loss function design	Moderate to high, requires tuning	Low, no targeted reweighting mechanism
Integration with Mechanistic Model	Strong, sequential adjustments align with mechanistic errors	Moderate, overall optimization may dilute mechanistic input	Moderate, similar to Gradient Boosting	Weak, independent trees cannot directly leverage mechanistic output
Computational Complexity	Moderate, fewer hyperparameters	High, requires tuning multiple hyperparameters	Very high, additional regularization and optimization	Low to moderate, simple hyperparameters
Suitability for Time-Series Data	High, iterative correction fits time-series nature	Moderate, relies on data preprocessing	Moderate, similar to Gradient Boosting	Low, lacks temporal dependency
Relevance to This Study	High, simple and synergistic with hybrid framework	Moderate, added complexity with limited gains	Moderate, high optimization cost	Low, limited ability to utilize mechanistic model

Table 1. Advantages and disadvantages of each algorithm.

where $D_1(i)$ is the weight of sample N in round t .

2. Training a weak classifier

After completing several iterations, a weak classifier, denoted as H_t , is trained at round t utilizing the existing sample weights $D_1(i)$. The weak classifier h , which exhibits the smallest error rate at this point, is chosen as the t th basic classifier H_t . Subsequently, the weak classifier $H_t : X \rightarrow \{-1, 1\}$ is computed, which incurs an error on the distribution D_t .

$$e_t = P(H_t(x_i) \neq y_i) = \sum_{i=1}^N w_{ti} I(H_t(x_i) \neq y_i) \quad (34)$$

where the error rate of $H_t(x)$ on the training set e_t is the sum of the weights of the samples misclassified by $H_t(x)$.

3. Calculate the weights of weak classifiers

Calculate the weights of the classifier H_t based on the current error rate of the weak classifier:

$$a_t = \frac{1}{2} \ln\left(\frac{1 - e_t}{e_t}\right) \quad (35)$$

where a_t denotes the degree of contribution of weak classifiers.

4. Update sample weights

Update the weight distribution of the training samples D_{t+1} :

$$D_{t+1} = \frac{D_t(i) \exp(-a_t y_i H_t(x_i))}{Z_t} \quad (36)$$

where Z_t is a normalization constant $Z_t = \sqrt{e_t(1 - e_t)}$.

5. Combinatorial weak classifier

Combine each weak classifier by its weight a_t with the following formula:

$$f(x) = \sum_{t=1}^T a_t H_t(x) \quad (37)$$

$$H_{final} = \text{sign}(f(x)) = \text{sign}\left(\sum_{t=1}^T a_t H_t(x)\right) \quad (38)$$

where T is the total number of weak classifiers and the sign function is used to transform the output into a classification result.

AdaBoost is a powerful ensemble learning method that can effectively enhance model performance by combining multiple weak classifiers. Figure 6 illustrates the AdaBoost training process, which includes key steps such as initialization of sample weights, training and weighting of weak classifiers, and the integration of the final

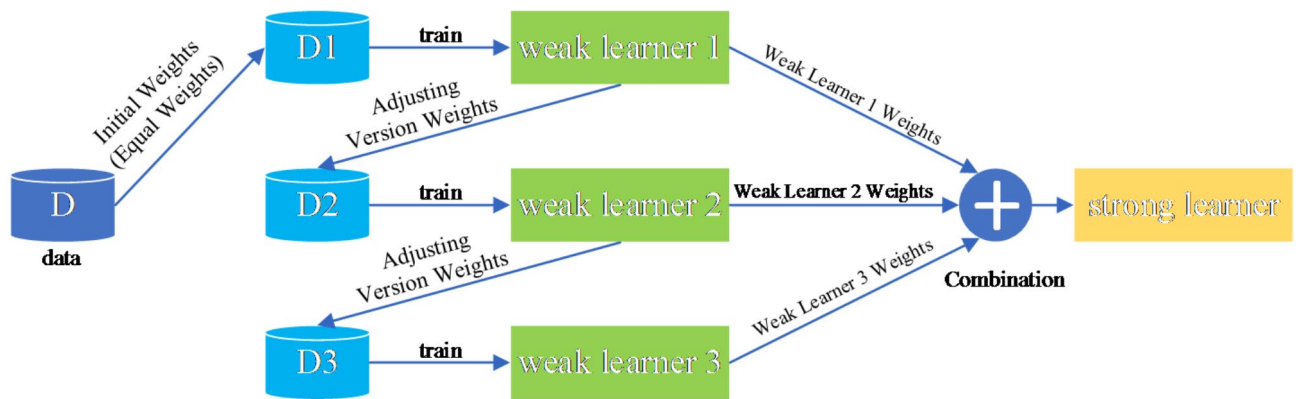


Fig. 6. ADABOOST training flow.

model. To evaluate the model's effectiveness, the previously mentioned metrics provide comprehensive insights into various aspects of performance. By appropriately adjusting parameters and model design, AdaBoost can achieve excellent results in many practical applications.

The advantages of AdaBoost are particularly evident in this study. Its iterative reweighting mechanism prioritizes samples with higher prediction errors, which is crucial for addressing the highly nonlinear and lag-dominated dynamic characteristics of the CZ crystal growth process. Unlike Bagging-based methods (e.g., Random Forest), which treat all samples equally, AdaBoost explicitly focuses on the residual errors of preceding weak learners, enabling targeted compensation for systematic biases in the mechanistic model's predictions. Additionally, compared to methods like Gradient Boosting or XGBoost, which rely on gradient descent to optimize complex loss functions, AdaBoost's additive model structure is more suitable for real-time control in industrial crystal growth environments. Its simplicity allows for seamless and efficient integration with BiLSTM outputs without introducing excessive computational overhead. Given that the features generated by BiLSTM inherently possess temporal dependencies and noise, AdaBoost effectively combines diverse weak regressors—even those with moderate individual accuracy—into a robust ensemble model. This helps mitigate the risk of overfitting, particularly in scenarios with limited training data, which is a common constraint in industrial settings.

Hybrid driven modeling approach based on BiLSTM and Adaboost algorithm

The proposed hybrid modeling approach integrates the BiLSTM algorithm and the Adaboost algorithm, where the former bi-directionally processes the data related to the prediction of the input crystal diameter through two independent LSTM networks in both the forward and backward directions, thus capturing the information of the past and future directions in the input sequences; and the latter iteratively trains multiple weak classifiers based on the data generated by the BiLSTM algorithm and gradually optimizes the overall prediction performance by assigning different weights to the results of each classifier. The latter iteratively trains multiple weak classifiers based on the data generated by the BiLSTM algorithm, and assigns different weights to the results of each classifier to gradually optimize the overall prediction performance.

As shown in Fig. 7, the steps to realize the prediction of crystal diameter during SSC growth are as follows:

1. Data Acquisition: Collecting data from the actual SSC growth process as well as the crystal growth model constructed based on the first-principles model under different working conditions in multiple furnaces to enrich the sample data set of the experiment.
2. mechanistic model simulation: Through Simlink simulation, input the lifting speed and heater power under actual working conditions, and get the prediction data of the diameter of the mechanistic model with large error. The diameter predicted by the purely mechanistic model only reaches about 70% of the actual crystal diameter.
3. Data preprocessing: the data to be put into the model training, the missing values, outliers processing, and filter out the random noise in the data, normalize the data, divided into training set and test set.
4. Training the BiLSTM Model: A bidirectional BiLSTM network is established to create the BiLSTM model. This network encompasses the following layers: an input layer, a bi-directional LSTM layer, a flattening layer, a fully connected layer, a ReLU activation layer, and an output layer. The data gathered in the earlier steps, including variables such as lifting speed, heater power, the diameter actually measured, and the discrepancy between the simulated and real diameters, are fed into the model. Subsequently, hyperparameters—including the learning rate, the number of training epochs, and the gradient threshold—are fine-tuned to optimize model training. The predictive outputs for the input samples from the trained individual BiLSTM model serve as a weak learner for the subsequent stage of the AdaBoost algorithm.
5. Adaboost Integrated Learning: This approach develops multiple weak classifiers by utilizing features produced by BiLSTM as input. Initially, each sample is assigned equal weights. The features generated by BiLSTM are iteratively trained to build the weak classifiers; if a sample is correctly classified, its weight is decreased for the next training set, while incorrectly classified samples see their weights increased. The adjusted

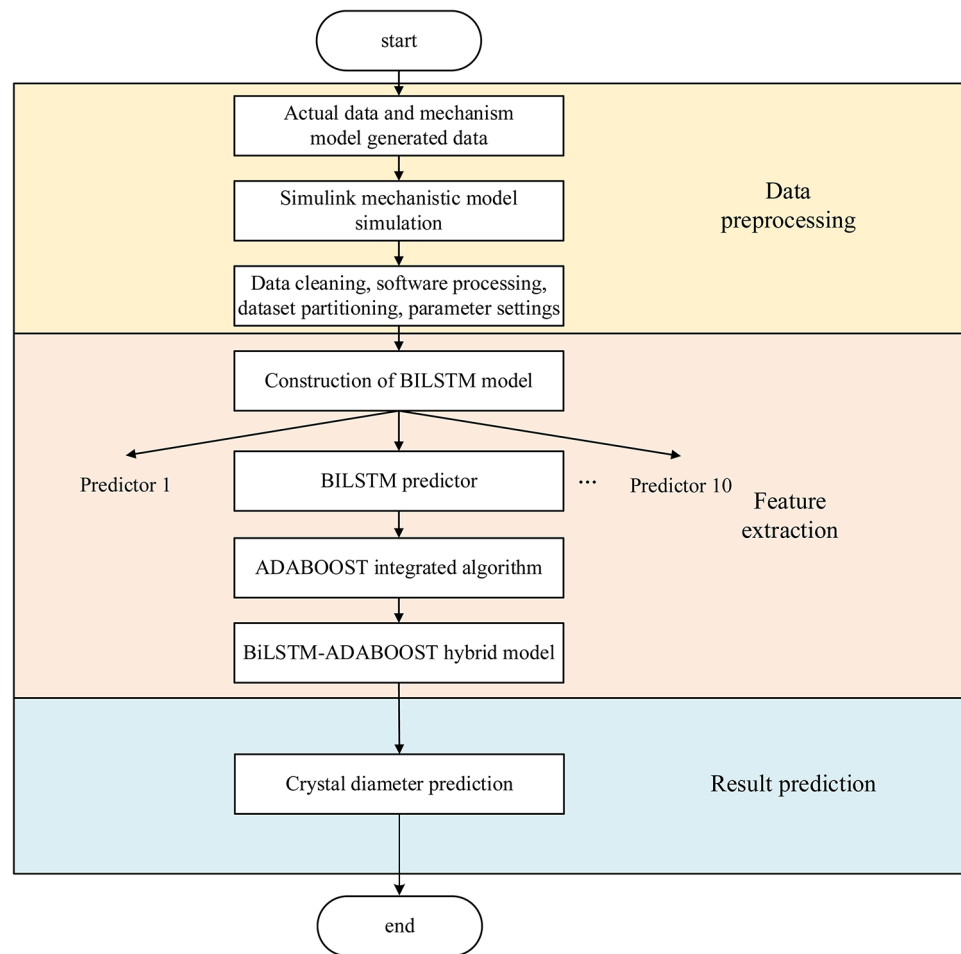


Fig. 7. Hybrid modeling predictions based on BiLSTM-ADABOOST.

sample set is then employed to train the subsequent classifier. Ultimately, all weak classifiers' outputs are amalgamated to create a robust classifier. After the completion of training for each weak classifier, the weights of those with lower classification error rates are elevated, while weights for classifiers with higher error rates are diminished. The final bias compensation is based on the difference between the predicted actual diameter and the diameter modeled mechanically.

6. The prediction of crystal diameter during SSC growth can be realized by inputting the measured data into the constructed module of hybrid modeling after summing up the difference with the difference prediction value.

The bidirectional structure of BiLSTM can capture both past and future information simultaneously (such as the lag effects of pulling speed and heating power), thereby enhancing the feature extraction capability for time series data. AdaBoost, through iterative weighting of weak learners (BiLSTM), reduces the influence of high-error samples, decreases overfitting, and improves robustness. The hybrid modeling approach combines a mechanistic model with a data-driven model—the former provides physical constraints, while the latter corrects residual errors—achieving a balanced trade-off between interpretability and accuracy.

BiLSTM-AdaBoost feature importance analysis

In the BiLSTM-AdaBoost hybrid-driven modeling proposed in this paper, the base learner is a deep network (BiLSTM) rather than a decision tree, so it cannot directly output “feature importance” metrics like tree-based models. Therefore, the Permutation Feature Importance method is used to compare feature importance. The specific steps are as follows:

First, the model's prediction error (MSE) on the unperturbed data is calculated as the baseline performance. Then, the values of a single feature are randomly permuted (i.e., shuffled), disrupting the association between that feature and the target variable, while keeping all other features unchanged. The prediction error (MSE) of the model on the perturbed data is then computed and compared with the baseline error. If the error increases significantly after the perturbation, it indicates that the feature is important for the model's prediction. Conversely, if there is little to no change, the feature is considered unimportant.

For the initial dataset D , the mean squared error (MSE) of model f_k is:

$$MSE_k = \frac{1}{M} \sum_{i=1}^M (f_k(X_i) - y_i)^2 \quad (39)$$

Feature j is permuted to generate the perturbed dataset $D^{(j)}$, and the mean squared error (MSE) of model f_k on $D^{(j)}$ is:

$$MSE_k^{(j)} = \frac{1}{M} \sum_{i=1}^M \left(f_k \left(X_i^{(j)} \right) - y_i \right)^2 \quad (40)$$

The importance score of feature j for model f_k is:

$$\Delta MSE_k^{(j)} = \frac{1}{M} \left[MSE_K - MSE_K^{(j)} \right] \quad (41)$$

where X_i is the dataset input, y_i is the dataset output, and f_k is the network model used—in this paper, it refers to the k -th BiLSTM model.

After integrating k weak learners, the global importance of feature j is:

$$P_j = \sum_{k=1}^K \left[\left(\frac{1}{2} \ln \left(\frac{1 - \varepsilon_k}{\varepsilon_k} \right) \right) \cdot \left(\Delta MSE_k^{(j)} \right) \right] \quad (42)$$

where ε_k is the weighted classification error, which is normalized to obtain a percentage form.

As shown in Fig. 8, The global feature importance scores are 0.1714, 0.2312, 0.1389, 0.2415, and 0.2169, respectively, indicating that Table 2 (heating power of the heater), Feature 4 (the discrepancy between the actual crystal radius and that predicted by mechanical models), and Feature 5 (radius measurement value) play a more significant role in model training. This is consistent with the previously mentioned Pearson correlation analysis and also aligns with real-world conditions.

Weight distribution:

Industrial experiment simulation

In the previous section, we developed a hybrid BiLSTM-ADABOOST driven model designed to more accurately predict the diameter of crystals during the growth of CZ SSC. In this section, we compare the performance of this model with other data-driven modeling methods.

Indicators for evaluating the performance of each model

To enable a comprehensive comparison of model performance, this study employs five evaluation metrics, summarized in Table 3. MSE measures the average of squared differences between actual and predicted values, highlighting the accuracy of the predictions. RMSE further indicates the magnitude of these errors. MAE tracks average absolute differences between observed and predicted outcomes, reflecting the model's reliability. MAPE captures forecasting accuracy as a percentage error. Lastly, R2 gauges the proportion of variance in the dependent variable explained by the independent variables, delivering a thorough assessment of the model's explanatory power.

Forecast results and analysis

In this instance, the SSC diameter during the CZ SSC growth process is established as the target variable. Utilizing data on the growth of the 12 inch SSC collected from the TDR-180 single crystal furnace between October 13 and October 20, 2020, the predictive capability of the proposed model is assessed by continuously sampling data throughout the constant diameter growth phase of the crystals. Figure 9 Actual diameter measurements in the constant diameter stage of the crystal.

In the experimental processing section, the core method for selecting 2,000 samples from 6,000 original time-series data points is based on a sliding window extraction mechanism and a data timeliness filtering principle. Using the sliding window approach, the window moves with a step size of 2, meaning each adjacent sample is spaced by 2 time points, thereby avoiding excessive data overlap. After theoretically generating approximately 3,000 candidate samples, the most recent 2,000 windowed samples at the end of the time series are retained with priority. This aims to optimize computational efficiency (reducing memory usage by 33%), enhance learning of recent dynamic patterns, and avoid interference from outdated data.

The dataset is partitioned using a non-random chronological split strategy (the first 80% for training, the last 20% for testing), strictly preserving temporal causality to prevent future data leakage, ensuring that model evaluation aligns with the physical constraints of real-world forecasting scenarios. This strategy adheres to common standards in the time-series prediction field, balancing engineering feasibility with academic rigor.

Handling outliers

When analyzing experimental data, a small number of data points may significantly deviate from the expected data distribution. If these outliers are not identified and removed, they can adversely affect the accuracy of To properly handle outliers in the dataset, this study adopts the 3-sigma rule. Let the actual measured value

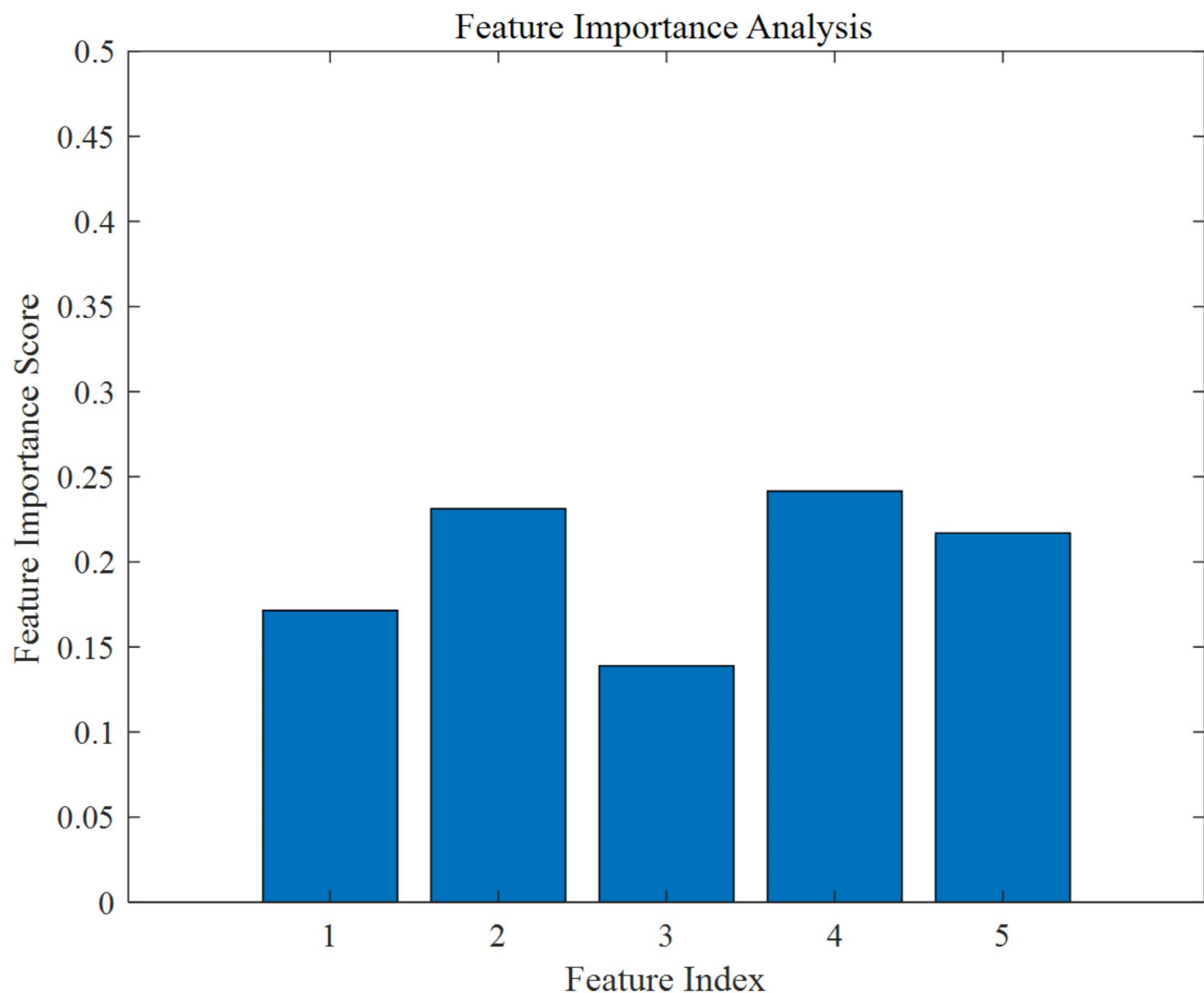


Fig. 8. Feature importance analysis.

Serial number	Feature name	Value
1	The lifting speed of the apparatus	0.1714
2	Heating power of the heater	0.2312
3	Crucible lifting speed	0.1389
4	The discrepancy between the actual crystal radius and that predicted by mechanical models	0.2415
5	Radius measurement value	0.2169

Table 2. Analysis of parameters for different feature values.

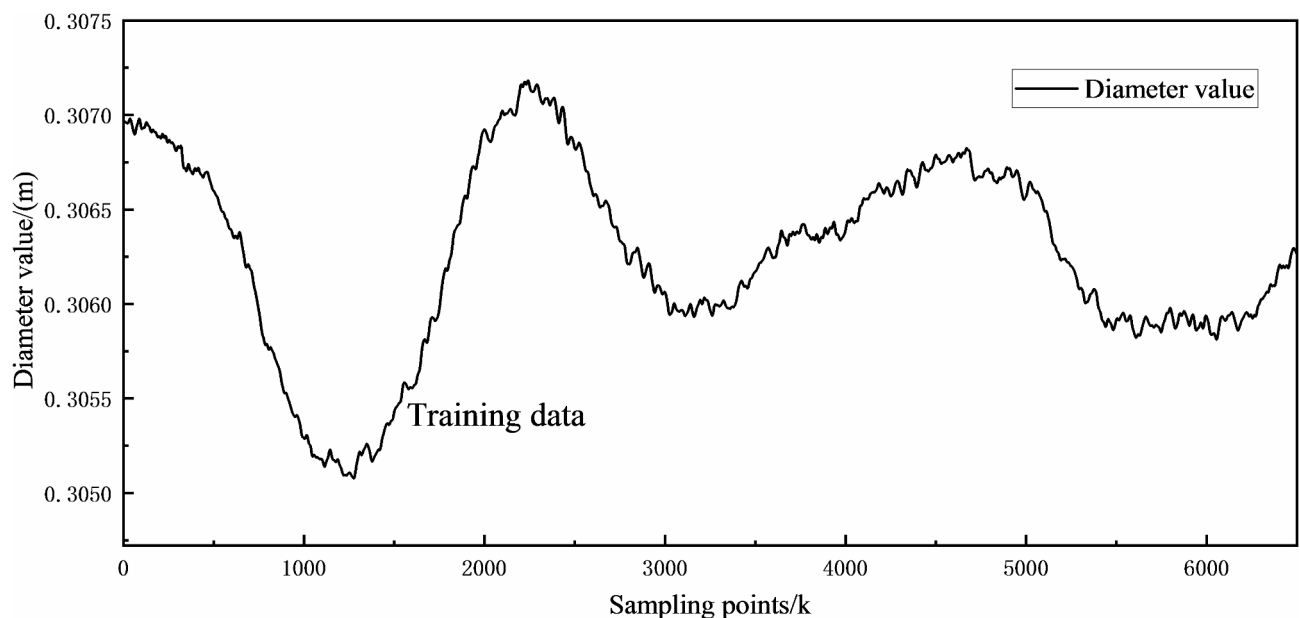
be $x_1, x_2, \dots, x_n$, the arithmetic mean be $\bar{x} = 1/n \cdot (\sum_{i=1}^n x_i)$, the absolute error be $\Delta x_i = x_i - \bar{x}$, and the standard error be $\hat{\sigma} = [\sum \Delta x_i^2 / (n - 1)]^{1/2}$. If the absolute error x_d of the measured value Δx_i satisfies condition $|\Delta x_i| > 3\hat{\sigma}$, then x_d is identified as an outlier and should be eliminated.

Data normalization processing

In data-driven soft sensing modeling, the use of auxiliary variables is essential. Due to potential differences in the numerical scales of various auxiliary variables, directly using raw values for computation may prolong the network's learning time. Therefore, it is particularly important to perform normalization on both input and output data. Specifically, the normalization is carried out using the formula shown in Eq. (39).

$$\hat{x}_i = \frac{x_i - \bar{x}}{x_{\max} - x_{\min}} \quad (39)$$

Norm	Define	Formulas
MSE	Mean square	$MSE = \frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2$
RMSE	Root mean square	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2}$
MAE	Mean absolute value error	$MAE = \frac{1}{N} \sum_{i=1}^N y(i) - \hat{y}(i) $
MAPE	Average absolute percentage error	$MAPE = \frac{1}{N} \sum_{i=1}^N \left \frac{y(i) - \hat{y}(i)}{y(i)} \right \times 100\%$
R2	Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^N (y(i) - \hat{y}(i))^2}{\sum_{i=1}^N (y(i) - \bar{y})^2}$

Table 3. Model performance evaluation indicators.**Fig. 9.** Actual diameter measurements in the constant diameter stage of the crystal.

In the equation: x_i and $\bar{x}$ represent the i -th variable of x and its mean value, respectively; $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of x , respectively; $\hat{x}_i$ is the normalization result.

The growth environment and process of CZ SSC is notably intricate, marked by a multitude of interrelated factors that can influence outcomes. This paper focuses on examining specific variables that pertain to the crystal radius at the constant diameter stage of growth. In a real industrial setting, various variables exert significant influence over the crystal radius. These include, among others, the speed at which the crystal is lifted, the rotation speed of the crystal, the lifting speed of the crucible, the rotation speed of the crucible, as well as the turn speed of the crystal and the rise speed of the crucible. In addition to these mechanical parameters, thermal factors such as the temperature of the liquid surface and the temperature within the hot field further contribute to the overall growth dynamics. To refine the prediction accuracy for the crystal diameter, the paper employs data gathered from practical crystal pulling experiments alongside statistical analysis through Pearson's correlation coefficient. This analysis identifies several key variables that play a crucial role in diameter prediction, namely the heating power of the heater, pulling rate, the discrepancy between the actual crystal radius and that predicted by mechanical models, as well as the measured radius itself. Subsequently, these primary influencing variables are utilized in further modeling activities aimed at improving the predictive capabilities regarding the crystal diameter, enhancing the efficiency and quality of the crystal growth process.

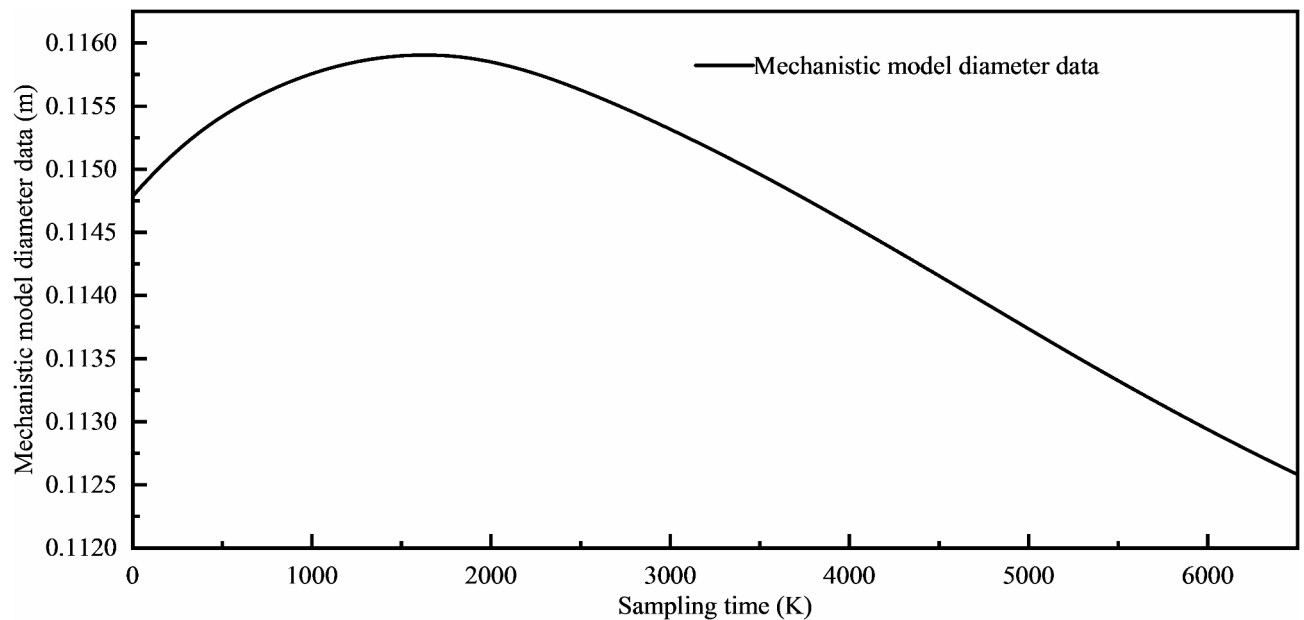
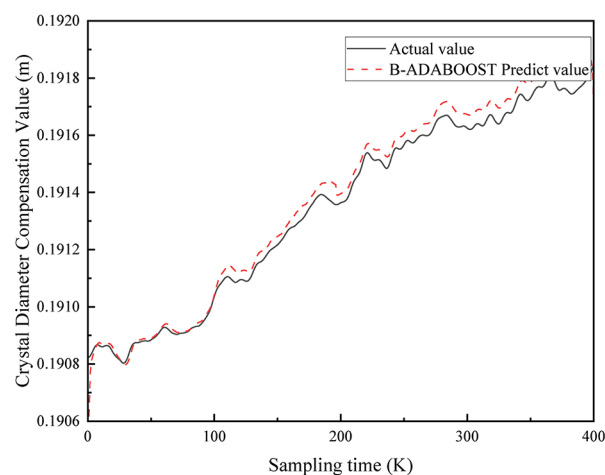
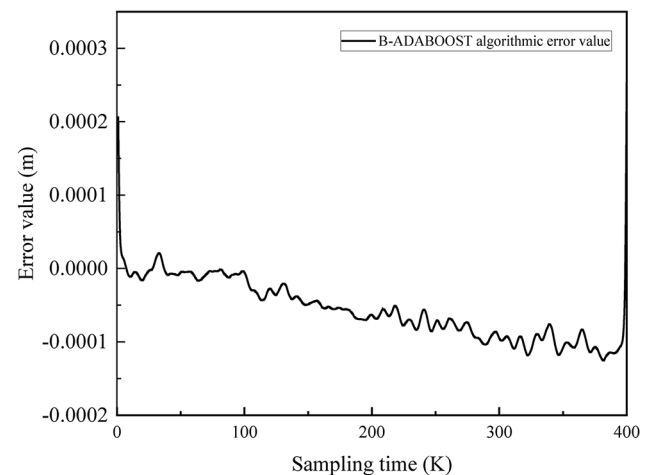


Fig. 10. Mechanistic model crystal diameter predictions.



(a) BILSTM-ADABOOST algorithm prediction results



(b) BILSTM-ADABOOST algorithm prediction error

Fig. 11. Prediction results of BILSTM-ADABOOST algorithm and prediction errors.

In this paper, hybrid modeling is used to complete the crystal radius prediction, firstly, we need to get the diameter prediction value of the mechanistic model in simulink simulation, as shown in Fig. 10, and then make difference with the actual measurement value to compensate for the error of the simple mechanistic modeling and improve the prediction accuracy, and then we will use the difference value and other variables as the inputs and outputs to train the data-driven BILSTM-ADABOOST model.

To evaluate the metrics of the BILSTM-ADABOOST model, various models including CNN, LSTM, GRU, CNN-BILSTM, CNN-LSTM, and LSTM are employed alongside BILSTM and the BILSTM-ADABOOST for diameter prediction and comparative analysis. The prediction outcomes of the BILSTM-ADABOOST algorithm, including prediction errors, are illustrated in Fig. 11. Figure 12 presents a comparison of the prediction performance of the different modeling approaches, while Fig. 13 depicts the prediction error comparison among these methods (Table 4).

Based on the data in Table 5, the comparison algorithms are CNN, GRU, CNN-BILSTM, CNN-LSTM, LSTM, BILSTM, and BILSTM-ADABOOST. From the R^2 perspective, the value increases from 56.434% (CNN) to 95.483% (BILSTM-ADABOOST). This indicates that the latter can capture data trends and features more thoroughly, achieving the highest prediction accuracy.

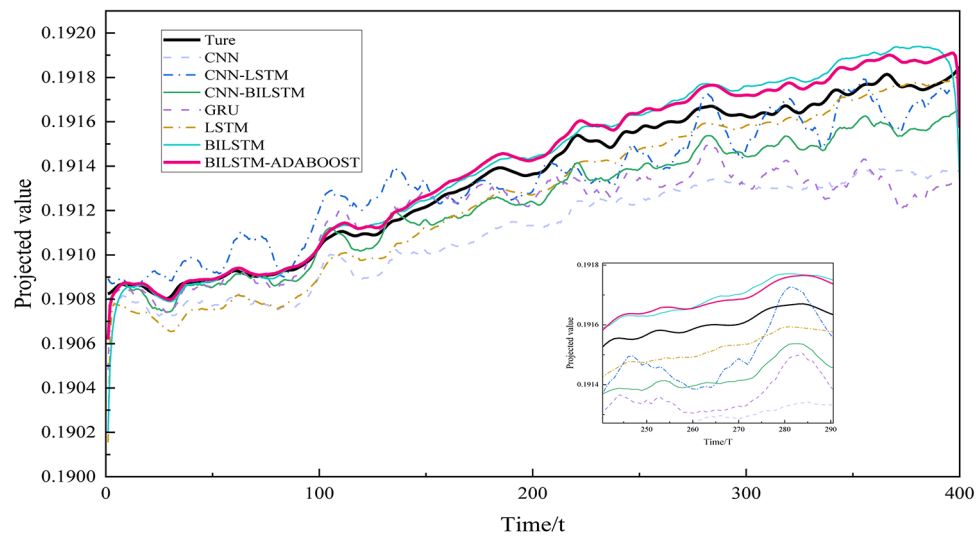


Fig. 12. Comparison of the prediction effect of each modeling method.

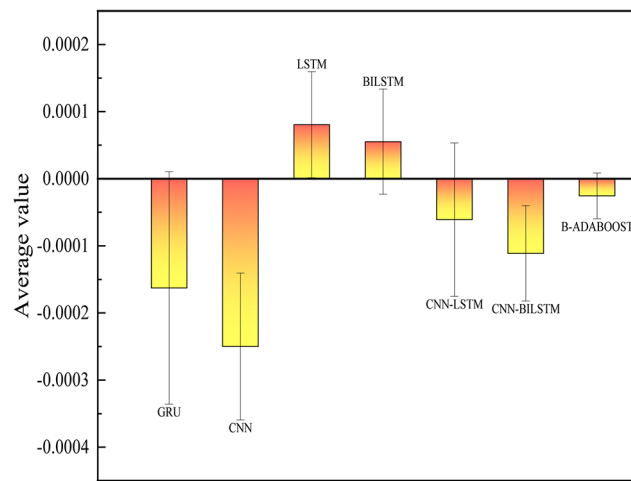


Fig. 13. Comparison of prediction errors between modelling methods.

Algorithm	Parameter settings
CNN	Kernel size = 3, Number of kernels = 128, Pooling window = 2, Dropout = 0.2, Learning rate = 0.001
LSTM	Hidden units = 128, Number of layers = 1, Dropout = 0.2, Learning rate = 0.001
GRU	Hidden units = 128, Number of layers = 1, Dropout = 0.2, Learning rate = 0.001
CNN-BiLSTM	Kernel size = 3, Number of kernels = 128, Pooling window = 2, BiLSTM hidden units = 128, Layers = 1, Dropout = 0.2, Learning rate = 0.001
CNN-LSTM	Kernel size = 3, Number of kernels = 128, Pooling window = 2, LSTM hidden units = 128, Layers = 1, Dropout = 0.2, Learning rate = 0.001
BiLSTM-AdaBoost	BiLSTM hidden units = 128, Layers = 1, Dropout = 0.2, AdaBoost weak classifiers = 50, Learning rate = 0.001

Table 4. Parameter configuration for each algorithm.

Table 5 presents the training times for each model. As shown in the table, the standalone models have relatively short training times, generally within 10 s. Due to the increased complexity of the hybrid model, its training time is comparatively longer. However, considering the slow-changing nature of semiconductor silicon single crystal production, the training time of the BiLSTM-AdaBoost model remains within practical limits. Furthermore, the training time can be reduced through model structure optimization (e.g., limiting the number of weak learners and adjusting the number of LSTM layers) as well as the use of parallel computing.

From the MSE/RMSE perspective, the BiLSTM-ADABOOST model has the lowest MSE of 5.038×10^{-9} , with a corresponding RMSE of only 7.098×10^{-5} . This shows that BiLSTM-ADABOOST has a smaller overall error

Model	R2	MSE	RMSE	MAE	MAPE	Training time
CNN	56.434%	7.450×10^{-8}	2.729×10^{-4}	2.501×10^{-4}	0.131%	8.775
GRU	58.177%	5.645×10^{-8}	2.376×10^{-4}	1.769×10^{-4}	0.093%	10.886
CNN-BILSTM	85.595%	1.740×10^{-8}	1.319×10^{-4}	1.132×10^{-4}	0.059%	22.337
CNN-LSTM	87.891%	1.315×10^{-8}	1.147×10^{-4}	9.569×10^{-5}	0.050%	20.359
LSTM	90.862%	1.065×10^{-8}	1.032×10^{-4}	9.119×10^{-5}	0.048%	9.035
BILSTM	91.786%	9.157×10^{-9}	9.569×10^{-5}	6.999×10^{-5}	0.037%	10.177
BILSTM-ADABOOST	95.483%	5.038×10^{-9}	7.098×10^{-5}	5.917×10^{-5}	0.031%	31.094

Table 5. Predictive performance metrics for each model.

in terms of numerical scale. From the MAE perspective, the absolute error decreases from 2.501×10^{-4} (CNN) to 5.917×10^{-5} (BILSTM-ADABOOST). This demonstrates that deeper or integrated deep learning methods have an advantage in reducing absolute error. From the MAPE perspective, all models have relatively small values (0.031%–0.131%), showing good overall performance. However, BILSTM-ADABOOST has the lowest MAPE of 0.031%, indicating a very small percentage error compared to the true value. This reflects higher predictive accuracy and stability.

In summary, hybrid models tend to achieve better results than single models, especially BILSTM-ADABOOST outperforms other comparative models in all indicators, suggesting that: the combination of two-way LSTM's ability to capture time-series features and AdaBoost algorithm's idea of iterative weighting of the weak learners jointly improves the model-fitting ability, significantly reduces the error, and improves the recognition accuracy.

Conclusion

This study proposes a novel hybrid-driven modeling approach for predicting the crystal diameter during the isothermal stage of CZ SSC growth. The method integrates a mechanistic model with a data-driven framework that combines BiLSTM networks and the AdaBoost algorithm. Compared to traditional mechanistic models and standalone data-driven models, the proposed hybrid model significantly improves prediction accuracy. Experimental results demonstrate that the BiLSTM-AdaBoost model achieves an R2 of 95.483%, a mean absolute error (MAE) of 5.917×10^{-5} , and a mean absolute percentage error (MAPE) of 0.031%, indicating its superior capability to capture the complex nonlinear dynamics of the CZ process. Moreover, by leveraging AdaBoost's error-driven adaptability and BiLSTM's temporal dependency modeling, the hybrid model exhibits strong robustness to noise and fluctuations in input data, making it particularly suitable for real-time industrial applications with variable data quality. The lightweight ensemble structure of AdaBoost and the parallel processing capability of BiLSTM also ensure computational efficiency, enabling real-time control of crystal diameter under high-temperature, multi-physics environments. Importantly, this hybrid approach maintains the physical interpretability of mechanistic models while incorporating the flexibility of data-driven methods, ensuring that predictions align with the underlying physical dynamics of crystal growth. Overall, the proposed method provides a robust framework for real-time diameter control in industrial SSC production, contributing to defect reduction, yield improvement, and the advancement of semiconductor manufacturing technology.

Data availability

The data related to the silicon crystal growth process involved in this study are considered confidential corporate information and trade secrets. In compliance with the confidentiality agreement signed with our collaborators, the relevant data cannot be made publicly available. However, the deep learning model code and materials used for academic research in this study can be requested by contacting the corresponding author. We will provide the relevant code after a reasonable assessment, provided it does not violate any confidentiality clauses, to support further related research. Moreover, the paper has provided detailed information on the model architecture, training process, and key parameters to enhance the reproducibility and reference value of the research.

Received: 19 January 2025; Accepted: 1 April 2025

Published online: 24 May 2025

References

1. Liu, D. *Modeling and Control of the Czochralski Silicon Single Crystal Growth Process* (Science Press, 2015).
2. Liu, D. *Numerical Simulation and Process Optimization of the Czochralski Silicon Single Crystal Growth Process* (Science Press, 2020).
3. Liu, D., Zhao, X. G. & Zhao, Y. A review of research on modeling and control of the Czochralski silicon single crystal growth process. *Control Theory Appl.* **34**(1), 1–12 (2017).
4. Winkler, J., Neubert, M. & Rudolph, J. Nonlinear model-based control of the Czochralski process I: Motivation, modeling and feedback controller design. *J. Cryst. Growth* **312**(7), 1005–1018 (2010).
5. Winkler, J., Neubert, M. & Rudolph, J. Nonlinear model-based control of the Czochralski process II: Reconstruction of crystal radius and growth rate from the weighing signal. *J. Cryst. Growth* **312**(7), 1019–1028 (2010).
6. Rahmanpour, P. & Bones, J., et al. Linear and Nonlinear State Estimation in the Czochralski Process: SDOS (2013).
7. Rahmanpour, P. et al. Nonlinear State Estimation in the Czochralski Process. *IFAC Proc. Vol.* **47**(3), 4891–4896 (2014).
8. Bukhari, H. Z., Hovd, M. & Winkler, J. Inverse response behaviour in the bright ring radius measurement of the Czochralski process I: Investigation (2021).

9. Zheng, Z., Seto, T. & Kim, S., et al. A first-principle model of 300mm Czochralski single-crystal Si production process for predicting crystal radius and crystal growth rate. *J. Crystal Growth* S226007304 (2018).
10. Ng, J. & Dubljevic, S. Optimal boundary control of a diffusion-convection-reaction PDE model with time-dependent spatial domain: Czochralski crystal growth process. *Chem. Eng. Sci.* **67**(1), 111–119 (2012).
11. Ng, J. & Dubljevic, S. Optimal control of convection-diffusion process with time-varying spatial domain: Czochralski crystal growth. *J. Process Control* **21**(10), 1361–1369 (2011).
12. Ng, J., Aksikas, I. & Dubljevic, S. Control of parabolic PDEs with time-varying spatial domain: Czochralski crystal growth process. *Int. J. Control* **86**(9), 1–12 (2013).
13. Liang, Y. M. & Liu, D. A self-organizing identification algorithm of T-S fuzzy model and its application. *Chin. J. Sci. Instrum.* **32**(9), 1941–1947 (2011).
14. Liang, Y. M. et al. Self-organizing algorithm of T-S fuzzy model based on support vector regression and its application. *Acta Autom. Sin.* **39**(12), 2143–2149 (2013).
15. Zhao, X. G., Liu, D. & Jing, K. L. Nonlinear system identification with noise based on an improved ant-lion algorithm and T-S fuzzy model. *Control Decis.* **34**(4), 759–766 (2019).
16. Zhang, X. Y. et al. Study on crystal diameter model identification method based on data-driven approach. *J. Synth. Crystals* **50**(8), 1552–1561 (2021).
17. Zhang, J., Tang, Q. & Liu, D. Research into the LSTM neural network-based crystal growth process model identification. *IEEE Trans. Semicond. Manuf.* **32**, 220 (2019).
18. Kato, S. et al. Gray-box modeling of 300 mm diameter Czochralski single-crystal Si production process. *J. Cryst. Growth* **553**, 125929 (2020).
19. Ren, J., Liu, D. & Wan, Y. Data-driven and mechanism-based hybrid model for semiconductor silicon monocrystalline quality prediction in the Czochralski process. *IEEE Trans. Semicond. Manuf.* **35**(4), 658–669 (2022).
20. Wan, Y., Liu, D. & Ren, J. C. Performance-driven semiconductor silicon crystal quality control. *J. Process Control* **120**, 68–85 (2022).
21. Rahmanpour, P., Sælid, S. & Hovd, M. Run-To-Run control of the Czochralski process. *Comput. Chem. Eng.* **104**, 353–365 (2017).
22. Mika, K. & Uelhoff, W. Shape and stability of Menisci in czochralski growth and comparison with analytical approximations. *J. Cryst. Growth* **30**(1), 9–20 (1975).
23. Huh, C. & Scriven, L. E. Shapes of axisymmetric fluid interfaces of unbounded extent. *J. Colloid Interface Sci.* **30**(3), 323–337 (1969).
24. Boucher, E. A., Evans, M. J. B. & McGarry, S. Capillary phenomena: XX. Fluid bridges between horizontal solid plates in a gravitational field. *J. Colloid Interface Sci.* **89**(1), 154–165 (1982).
25. Rahmanpour, P. & Sælid, S., et al. Nonlinear Model Predictive Control of the Czochralski Process*. *Ifac Papersonline* (2016).
26. Li, P. et al. State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network. *J. Power Sources* **459**, 228069 (2020).
27. Jingzhou, X. et al. A signal recovery method for bridge monitoring system using TVFEMD and encoder-decoder aided LSTM. *Measurement* **214**, 112797 (2023).
28. Hao, Z. et al. Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information. *Energy* **231**, 120908 (2021).
29. Liu, S., Lee, K. & Lee, I. Document-level multi-topic sentiment classification of Email data with BiLSTM and data augmentation. *Knowl. Based Syst.* **197**, 105918 (2020).
30. Thyluru, R. M. et al. Homogeneous adaboost ensemble machine learning algorithms with reduced entropy on balanced data. *Entropy* **25**(2), 245 (2023).
31. Dawei, H. et al. Demand response-oriented virtual power plant evaluation based on AdaBoost and BP neural network. *Energy Rep.* **9**(S8), 922–931 (2023).

Author contributions

Methodology, Y.-Y.L.; Software, Y.-M.J.; Validation, D.L. and Y.-M.J.; Investigation, Y.-Y.L.; Writing–review & editing, Y.-Y.L.; Supervision, S.-H.W.; Project administration, D.L.; Funding acquisition, D.L.

Funding

This research was funded by National Natural Science Foundation of China grant number 62127809.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025