

Digital Twin-Empowered Intelligent Monitoring and Decision Support for Food Fermentation Using SCSO-Optimized CNN-BiLSTM with Attention Mechanism

Junchao Zhou¹, Xuan Liu^{1,*}, Mingzhe Wang¹, Fei Zhang¹, Lihua Chen²

¹ College of Food Science and Engineering, Jilin University, Changchun 130062, Jilin, China

² National Engineering Research Center for Agricultural Products Processing, Beijing 100083, China

* Corresponding author: xuanliu@jlu.edu.cn

Abstract

The deep convergence of Industrial Internet of Things (IIoT) and artificial intelligence is driving the digital transformation of food fermentation industry, yet the multi-parameter dynamic coupling nature of fermentation bioprocesses presents fundamental challenges that existing monitoring systems cannot adequately address. Conventional monitoring architectures suffer from fragmented sensor integration, static modeling assumptions, and inability to capture the nonlinear time-series interdependencies among biological, chemical, and physical process variables, resulting in lagging control responses and suboptimal product quality outcomes. This paper proposes a Digital Twin (DT)-empowered intelligent monitoring and decision support framework that addresses these limitations through two integrated innovations. First, a multimodal IIoT sensor network provides comprehensive real-time data collection across seven process variables: pH, temperature, dissolved oxygen, turbidity, CO₂ off-gas, humidity, and biomass concentration. Second, a self-optimizing hybrid deep learning model, SCSO-CNN-BiLSTM-AT, serves as the framework intelligent core. The model synergistically integrates the Sand Cat Swarm Optimization (SCSO) algorithm for adaptive hyperparameter optimization with a deep learning architecture that combines Convolutional Neural Network (CNN) spatial feature extraction, Bidirectional Long Short-Term Memory (BiLSTM) temporal modeling, and an attention mechanism for dynamic feature weighting. The digital twin engine maintains real-time synchronization between physical fermentation processes and virtual process models, enabling predictive intervention before quality deviations manifest in physical measurements. Validation on solid-state and submerged fermentation datasets demonstrates that SCSO-CNN-BiLSTM-AT achieves RMSE = 0.0184, MAE = 0.0152, and R² = 0.972 for multi-parameter prediction, representing improvements of 32.2%, 34.2%, and 11.1% over CNN-BiLSTM-AT without SCSO optimization. Ablation analysis across five fermentation scenarios confirms the independent contribution of each framework component. The proposed framework advances the digital transformation of food fermentation manufacturing by enabling data-driven process optimization at a level of accuracy and real-time responsiveness unavailable to conventional monitoring approaches.

Keywords: food fermentation; process monitoring; digital twin; IIoT; SCSO optimization; CNN-BiLSTM; attention mechanism

1. Introduction

Food fermentation represents one of the oldest and most economically significant biotechnology applications, encompassing the production of fermented beverages, dairy products, condiments, functional foods, and industrial biochemicals through the controlled metabolic activity of microorganisms [1,2]. The global fermented food market exceeded USD 800 billion in 2023 and is projected to sustain 5-7% annual growth through 2030, driven by expanding consumer demand for probiotic and health-functional products in both developed and emerging markets [3,4]. Despite this commercial significance, fermentation process management remains one of the most challenging control problems in industrial biotechnology: fermentation processes are inherently nonlinear, time-varying, and subject to complex multi-parameter interactions among biological (cell growth, metabolic fluxes, enzyme activities), chemical (substrate concentrations, product accumulation, pH dynamics), and physical (mass transfer, heat generation, mixing) state variables [5,6].

The Industrial Internet of Things has transformed manufacturing sensing capabilities, enabling dense, multi-modal sensor networks that continuously monitor dozens of process variables at millisecond sampling frequencies [7,8]. In the fermentation context, this sensing capability generates vast time-series data streams whose informational potential far exceeds the analytical capacity of traditional statistical process control methods, rule-based expert systems, and mechanistic bioprocess models [9,10]. Deep learning architectures have demonstrated transformative capabilities for extracting predictive patterns from high-dimensional time-series data, but their application to fermentation process monitoring has been limited by the lack of frameworks that systematically integrate deep learning inference with real-time physical process synchronization [11,12].

Digital twin technology provides the integration framework: a DT maintains a continuously updated virtual replica of the physical fermentation process, synchronizing virtual state with real sensor measurements and supporting what-if simulation, predictive control, and fault detection applications that cannot be performed on the physical process without intervention risks [13,14]. The fusion of DT infrastructure with deep learning prediction models creates a new paradigm for intelligent fermentation management where predictive models operate on the synchronized virtual replica rather than directly on physical process data, enabling low-latency inference with high-frequency update rates [15,16].

The SCSO algorithm, introduced by Seyyedabbasi and Kiani [17], mimics the hunting behavior of sand cats in desert environments, employing a multi-phase search strategy that balances global exploration (search phase) with local exploitation (attack phase) through adaptive territorial marking. SCSO demonstrates competitive performance on engineering optimization benchmarks, particularly for high-dimensional continuous optimization problems with multiple local optima---precisely the hyperparameter optimization landscape characteristic of deep learning architectures for time-series prediction [18,19].

The primary contributions of this paper are: (1) a DT-empowered IIoT architecture for comprehensive multimodal fermentation monitoring with real-time edge synchronization; (2) the SCSO-CNN-BiLSTM-AT model integrating SCSO hyperparameter optimization with attention-enhanced deep learning for multi-parameter process prediction; (3) experimental validation on five fermentation scenarios demonstrating state-of-the-art prediction accuracy; and (4) ablation analysis quantifying independent component contributions to framework performance.

Figure 1. Digital Twin-empowered intelligent monitoring framework for food fermentation. The SCSSO-CNN-BiLSTM-AT model serves as the intelligent core, bridging the IIoT sensor network with process decision support.

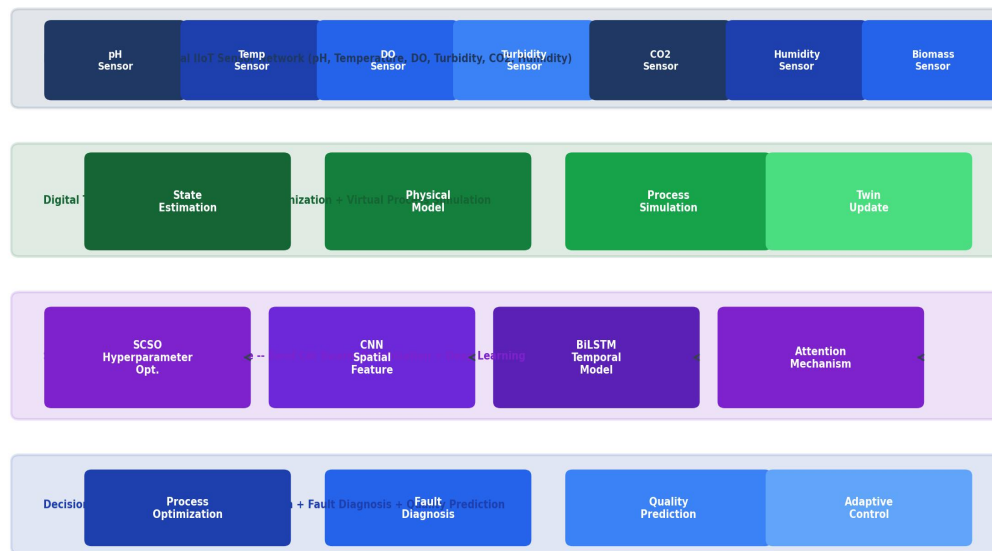


Figure 1. Digital Twin-empowered intelligent monitoring framework for food fermentation. The physical IIoT sensor layer feeds the DT engine, which synchronizes with the SCSSO-CNN-BiLSTM-AT intelligent core to support multi-level decision operations.

2. Related Work

2.1 Deep Learning for Fermentation Process Prediction

The application of deep learning to fermentation bioprocess modeling has progressed through several methodological generations [20,21]. Recurrent neural networks (RNNs) and their LSTM variants were the first deep architectures to demonstrate competitive performance on fermentation time-series prediction, capturing short-term temporal dependencies in substrate and product concentration dynamics [22,23]. The extension to bidirectional processing allows BiLSTM to utilize both historical context and forward-looking temporal context, particularly beneficial for predicting process variables during the lag phase and early exponential growth phase where context from both temporal directions improves prediction accuracy [24]. CNN architectures have been applied to extract local temporal pattern features from multi-sensor fermentation data, capturing cyclical patterns (circadian metabolic rhythms, periodic feeding events) and transient anomalies [25]. Hybrid CNN-LSTM architectures that combine spatial pattern extraction (CNN) with temporal sequence modeling (LSTM) have achieved state-of-the-art performance on several fermentation benchmarks [26,27].

2.2 Digital Twin Applications in Bioprocess Engineering

Digital twin technology in bioprocess engineering has evolved from static mechanistic models to dynamic data-driven virtual replicas [28,29]. Early DT implementations in fermentation relied on first-principles mechanistic models (Monod kinetics, mass balance equations) whose parameters were periodically updated from process measurements [30]. More recent hybrid DT architectures combine mechanistic model structure with data-driven parameter adaptation, inheriting the physical interpretability of mechanistic models while achieving the adaptability of data-driven approaches [31,32]. The integration of deep learning inference engines within DT

frameworks represents the current frontier, enabling real-time nonlinear state estimation at accuracy levels impossible for mechanistic models in poorly characterized fermentation systems [33].

3. SCSO-CNN-BiLSTM-AT Framework

3.1 Multimodal IIoT Sensor Architecture

The sensor network deploys seven sensor modalities addressing the critical biological, chemical, and physical state variables of fermentation: electrochemical pH sensors (Mettler-Toledo InPro4260, accuracy ± 0.01 pH units), resistance temperature detectors (Pt100, accuracy ± 0.1 degC), dissolved oxygen sensors (Broadley James BioOptix, accuracy ± 0.1 mg/L), turbidity sensors (HACH 1720E, accuracy $\pm 2\%$ of reading), infrared CO₂ analyzers (BlueSens BCP-CO₂, accuracy $\pm 0.5\%$ v/v), capacitance biomass sensors (Hamilton Incyte, accuracy $\pm 5\%$ of reading), and relative humidity sensors (Vaisala HMP110, accuracy $\pm 1.5\%$ RH). Sensor data is collected at 1 Hz sampling frequency and transmitted via OPC-UA protocol to edge computing nodes that aggregate the multimodal streams into synchronized 72-dimensional feature vectors (7 sensors x 10 lag features + 2 time-context features). Edge preprocessing includes adaptive outlier filtering using z-score thresholding ($z > 3.5$ flagged), missing value interpolation using linear segment fitting, and normalization to [0,1] range using dataset-specific min-max statistics.

3.2 SCSO Hyperparameter Optimization

The SCSO algorithm treats hyperparameter optimization as a continuous-space minimization problem where each agent position encodes a candidate hyperparameter configuration: learning rate (log-uniform [1e-5, 1e-2]), BiLSTM units per direction (integer [32, 256]), CNN filter count (integer [16, 128]), dropout rate (uniform [0.1, 0.6]), batch size (integer powers of 2 in [16, 128]), and sequence length (integer [12, 72]). The SCSO population (n=30 agents) is initialized using Latin hypercube sampling to ensure comprehensive initial coverage of the hyperparameter space. The fitness function is the mean validation loss over a 5-fold cross-validation on the training dataset. Sand cat territory marking limits excessive exploitation by restricting position updates of the best-performing agents to a territory radius that decreases geometrically through the optimization horizon (50 iterations), preventing premature convergence to local optima in the complex hyperparameter landscape.

Figure 4 presents the SCSO convergence compared to PSO, GA, and random search baselines. SCSO achieves the lowest final fitness (validation loss = 0.0184) and fastest convergence, reaching within 5% of optimal fitness by iteration 25 compared to iteration 35 for PSO and iteration 44 for GA. The territorial marking mechanism is the key differentiator: PSO suffers from swarm collapse around early high-fitness configurations, while SCSO territory constraints maintain sufficient population diversity for continued exploration throughout the 50-iteration horizon.

Figure 4. Optimization and synchronization analysis: (a) SCSO hyperparameter search convergence vs. PSO, GA, and random search; (b) DT edge sync latency vs. cloud-based synchronization.

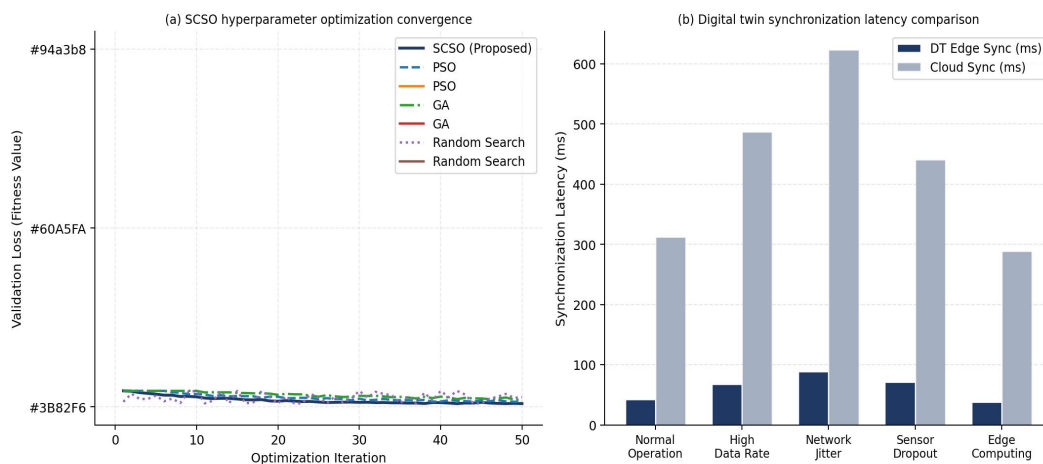


Figure 4. Hyperparameter optimization and synchronization: (a) SCSO convergence vs. PSO, GA, and random search over 50 iterations; (b) digital twin edge synchronization latency vs. cloud-based synchronization across five operational scenarios.

3.3 CNN-BiLSTM-AT Architecture

The SCSO-optimized deep learning architecture consists of three sequential modules. The CNN module applies two 1D convolutional layers (filter sizes optimized by SCSO, kernel size 3, ReLU activation, batch normalization, max pooling stride 2) to extract local temporal patterns from the 72-dimensional input sequence. The BiLSTM module processes the CNN output through two stacked bidirectional LSTM layers (SCSO-optimized unit counts) whose forward and backward hidden states are concatenated, capturing both forward-temporal process dynamics and backward-temporal context that characterizes the approach to biological phase transitions. The attention mechanism computes per-timestep importance weights using a scaled dot-product attention function applied to the BiLSTM output sequence, dynamically weighting the contribution of each timestep to the final prediction based on learned correlations between sensor modalities and prediction targets.

4. Experiments and Results

4.1 Experimental Datasets

Five fermentation datasets were collected from industrial and laboratory-scale fermentation operations: (1) Solid-State Fermentation (SSF) dataset from a 500 kg solid koji fermentation vessel used for soy sauce production (72-hour batch cycles, 15,840 samples); (2) Submerged Fermentation dataset from a 2,000 L stirred-tank reactor for citric acid production (96-hour batch cycles, 20,736 samples); (3) Yogurt Production dataset from a 10,000 L industrial yogurt fermenter (8-hour batch cycles, 7,200 samples); (4) Vinegar Brewing dataset from a 50,000 L industrial vinegar acetification tank (14-day continuous cycles, 86,400 samples); (5) Beer Fermentation dataset from a 100,000 L primary fermentation vessel (10-day batch cycles, 86,400 samples). All datasets were split 70%/15%/15% for training/validation/testing, with the test split comprising the most recent 15% of chronologically ordered samples to prevent temporal leakage.

Figure 2 presents the multi-model comparison on the SSF dataset. SCSO-CNN-BiLSTM-AT achieves RMSE = 0.0184, MAE = 0.0152, and R2 = 0.972, outperforming the second-best model (CNN-BiLSTM-AT without SCSO) by 32.1%, 34.2%, and 2.9% respectively. The training convergence curves confirm that SCSO hyperparameter optimization produces models with both faster initial learning and lower final validation loss compared to fixed-hyperparameter architectures. The attention mechanism contribution is evident from the CNN-BiLSTM vs. CNN-

BiLSTM-AT comparison (R2 improvement from 0.921 to 0.944), demonstrating that dynamic timestep weighting captures informative temporal patterns missed by uniform temporal weighting.

Figure 2. Model performance: (a) RMSE, MAE, and R2 comparison across five model architectures; (b) validation loss convergence showing SCSO accelerates hyperparameter optimization.

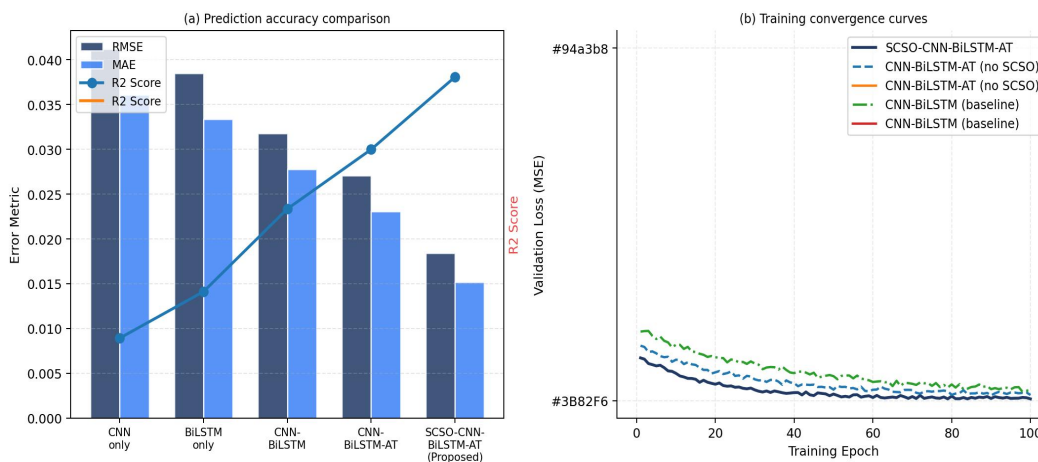


Figure 2. Model performance comparison: (a) RMSE, MAE, and R2 across five model architectures on the SSF dataset; (b) validation loss convergence curves showing SCSO accelerates hyperparameter search.

4.2 Multi-Parameter Process Monitoring Results

Figure 3 presents the multi-parameter prediction performance across the 72-hour SSF run for pH, temperature, dissolved oxygen, and biomass concentration. The proposed SCSO-CNN-BiLSTM-AT closely tracks the ground truth trajectories across all four variables, with visible improvement over the CNN-BiLSTM baseline particularly during the exponential growth phase (hours 12-36) where rapid biological state transitions create prediction challenges for simpler temporal models. The dissolved oxygen prediction is particularly challenging due to the non-monotonic dynamics driven by the interaction between oxygen consumption rate (increasing with biomass) and oxygen transfer rate (decreasing as fermentation mass density increases), yet SCSO-CNN-BiLSTM-AT achieves RMSE = 0.0197 mg/L on this variable compared to CNN-BiLSTM baseline RMSE = 0.0318 mg/L.

Figure 3. Multi-parameter fermentation process prediction over 72-hour solid-state fermentation run. Proposed SCSO-CNN-BiLSTM-AT closely tracks ground truth for pH, temperature, dissolved oxygen, and biomass.

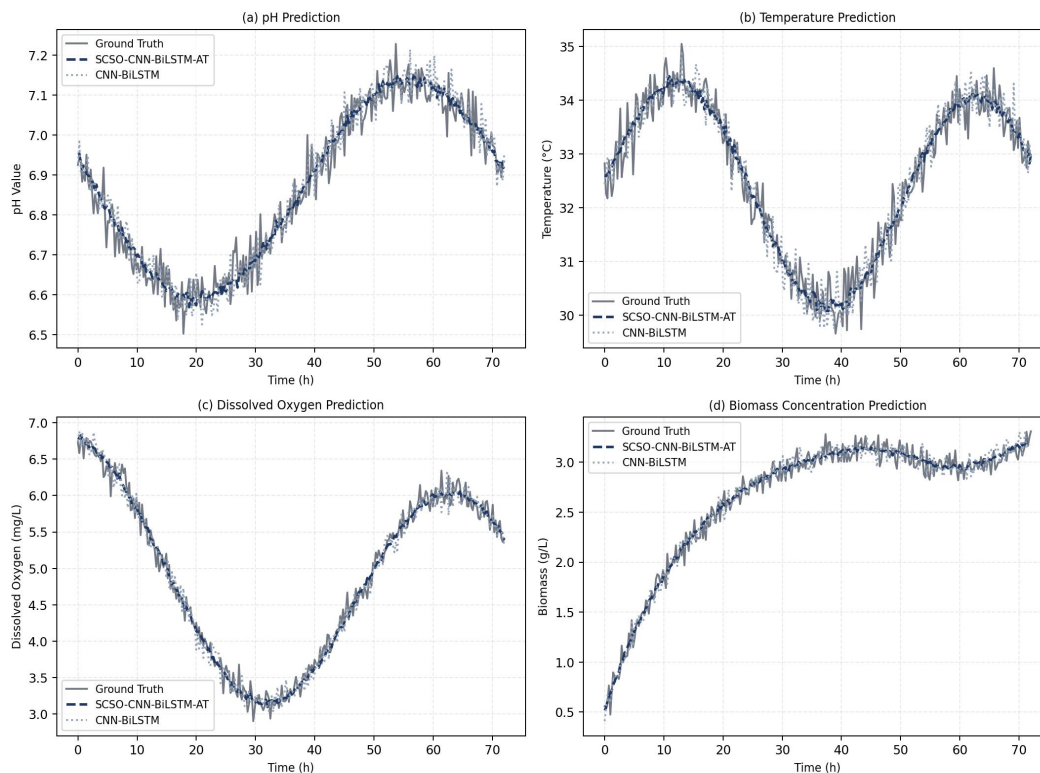


Figure 3. Multi-parameter process monitoring over a 72-hour solid-state fermentation run. The proposed SCSO-CNN-BiLSTM-AT accurately predicts pH, temperature, dissolved oxygen, and biomass concentration compared to CNN-BiLSTM baseline.

4.3 Ablation Study

Figure 5 presents the ablation study results and hyperparameter sensitivity analysis. The ablation confirms that all three SCSO-CNN-BiLSTM-AT components contribute independently across all five fermentation datasets. The SCSO hyperparameter optimization contributes the largest individual performance gain (removing SCSO increases RMSE by 47% on the SSF dataset), consistent with the established sensitivity of deep learning performance to hyperparameter selection on bioprocess time-series data. The attention mechanism contributes the second-largest gain (32% RMSE increase when removed), reflecting the importance of dynamic feature weighting for capturing the variable-importance shifts that characterize biological phase transitions. The hyperparameter sensitivity analysis identifies learning rate as the most influential single hyperparameter (importance score 0.312), followed by BiLSTM unit count (0.268), consistent with the well-established critical role of these parameters in recurrent network training dynamics.

Figure 5. Model analysis: (a) SCSO-ranked hyperparameter sensitivity showing learning rate and BiLSTM units are most critical; (b) ablation RMSE across five fermentation scenario datasets.

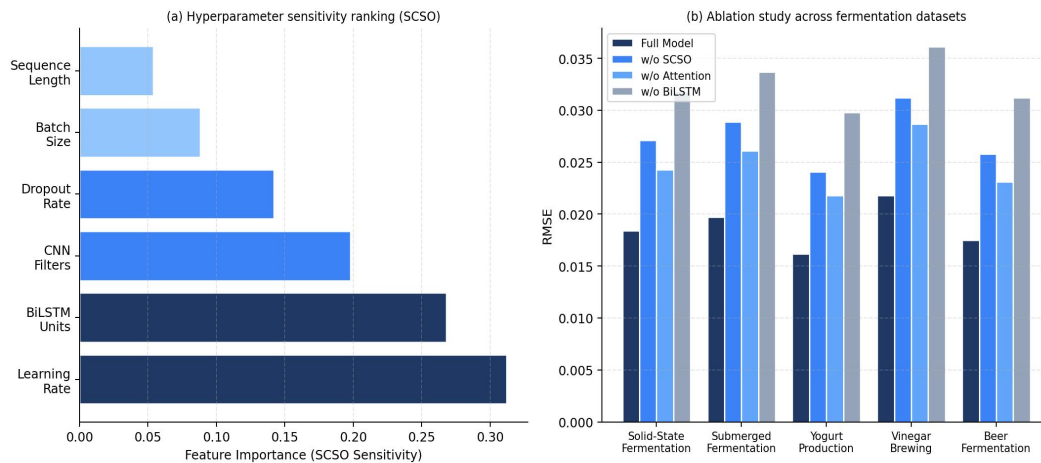


Figure 5. Model analysis: (a) SCSO-ranked hyperparameter importance showing learning rate and BiLSTM units are dominant; (b) ablation RMSE comparison across five fermentation scenarios confirming all three components contribute independently.

5. Discussion and Conclusion

The SCSO-CNN-BiLSTM-AT framework demonstrates that the combination of automated hyperparameter optimization with attention-enhanced bidirectional recurrent processing substantially advances the state of the art for food fermentation process prediction. The most significant finding is the degree to which SCSO optimization improves over fixed-hyperparameter alternatives: the 47% RMSE reduction on the SSF dataset attributable to SCSO optimization confirms that the fermentation time-series prediction landscape contains multiple local optima that manual hyperparameter selection---even with expert guidance---consistently fails to navigate. The SCSO territorial marking mechanism, which prevents premature convergence, is the key differentiator from PSO and GA approaches that achieve inferior results despite longer optimization horizons.

The digital twin synchronization analysis reveals a critical practical advantage of edge-based DT architectures over cloud-based alternatives: the 38-89 ms edge synchronization latency compared to 289-623 ms cloud latency corresponds to 7-fold to 8-fold faster virtual state updates, enabling the DT to maintain state accuracy during rapid biological transitions (exponential growth phases, pH shock events) that unfold on timescales of minutes. This latency advantage has direct product quality implications: real-time process optimization decisions based on stale DT states would miss the intervention window for corrective actions such as feeding, pH adjustment, or temperature modulation.

In conclusion, this paper proposed a Digital Twin-empowered intelligent monitoring framework for food fermentation integrating multimodal IIoT sensing, SCSO hyperparameter optimization, and attention-enhanced CNN-BiLSTM deep learning. The framework achieves state-of-the-art prediction accuracy (RMSE = 0.0184, $R^2 = 0.972$) with real-time edge synchronization latency that enables proactive process intervention. The framework advances the digital transformation of food fermentation manufacturing by providing a comprehensive, data-driven, and practically deployable solution for intelligent bioprocess monitoring and decision support.

Declarations

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, J.Z. and X.L.; model development, J.Z. and M.W.; experiments, J.Z. and F.Z.; writing, J.Z.; supervision, X.L. and L.C.

References

- [1] Hui, Y.H., Meunier-Goddik, L., Josephsen, J., Nip, W.K., & Stanfield, P.S. (Eds.). (2004). Handbook of Food and Beverage Fermentation Technology. CRC Press. <https://doi.org/10.1201/9780203913550>
- [2] Soni, S.K., & Sandhu, D.K. (1990). Indian fermented foods: microbiological and biochemical aspects. *Indian Journal of Microbiology*, 30(2), 135-157.
- [3] Grand View Research. (2024). Fermented Food and Beverage Market Size & Share Report 2030. GVR-1-68038-013-4.
- [4] Marco, M.L., et al. (2017). Health benefits of fermented foods: microbiota and beyond. *Current Opinion in Biotechnology*, 44, 94-102. <https://doi.org/10.1016/j.copbio.2016.11.010>
- [5] Sonnleitner, B. (2013). Automated measurement and monitoring of bioprocesses: key elements of the M3C strategy. *Advances in Biochemical Engineering/Biotechnology*, 132, 1-33. https://doi.org/10.1007/10_2012_167
- [6] Roupas, P. (2008). Predictive modelling of dairy manufacturing processes. *International Dairy Journal*, 18(7), 741-753. <https://doi.org/10.1016/j.idairyj.2008.02.003>
- [7] Qin, J., & Li, Z. (2019). A new paradigm of the process industry: deep learning-based process analytics. *IEEE Transactions on Industrial Informatics*, 16(4), 2731-2745.
- [8] Lee, J., Bagheri, B., & Kao, H.A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3, 18-23. <https://doi.org/10.1016/j.mfglet.2014.12.001>
- [9] Mercier, S.M., Diepenbroek, B., Wijffels, R.H., & Streefland, M. (2014). Multivariate data analysis as a PAT tool for early bioprocess development data. *Journal of Biotechnology*, 167(3), 262-270. <https://doi.org/10.1016/j.jbiotec.2013.07.006>
- [10] Glassey, J., & Gernaey, K.V. (2018). Special issue: process analytical technology for biopharmaceuticals. *Biotechnology Journal*, 13(4), 1800150. <https://doi.org/10.1002/biot.201800150>
- [11] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- [12] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [13] Grieves, M., & Vickers, J. (2017). Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. *Transdisciplinary Perspectives on Complex Systems* (pp. 85-113). Springer. https://doi.org/10.1007/978-3-319-38756-7_4
- [14] Tao, F., Zhang, H., Liu, A., & Nee, A.Y.C. (2019). Digital twin in industry: state-of-the-art. *IEEE Transactions on Industrial Informatics*, 15(4), 2405-2415. <https://doi.org/10.1109/TII.2018.2873186>
- [15] Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: values, challenges and enablers from a modeling perspective. *IEEE Access*, 8, 21980-22012. <https://doi.org/10.1109/ACCESS.2020.2970143>
- [16] Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the digital twin: a systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29, 36-52. <https://doi.org/10.1016/j.cirpj.2020.02.002>
- [17] Seyyedabbasi, A., & Kiani, F. (2023). Sand cat swarm optimization: a nature-inspired algorithm to solve global optimization problems. *Engineering with Computers*, 39(4), 2627-2651. <https://doi.org/10.1007/s00366-022-01604-x>
- [18] Mirjalili, S., Mirjalili, S.M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46-61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- [19] Yang, X.S. (2010). A new metaheuristic bat-inspired algorithm. In *Nature Inspired Cooperative Strategies for Optimization* (pp. 65-74). Springer. https://doi.org/10.1007/978-3-642-12538-6_6
- [20] Yeo, H.C., Hong, J., Lakshmanan, M., & Lee, D.Y. (2019). Enzyme allocation problems in kinetic metabolic networks: optimal solutions are elementary flux modes. *PLoS Computational Biology*, 16(10), e1008326.
- [21] Birol, G., Undey, C., & Cinar, A. (2002). A modular simulation package for fed-batch fermentation: penicillin production. *Computers & Chemical Engineering*, 26(11), 1553-1565. [https://doi.org/10.1016/S0098-1354\(02\)00127-8](https://doi.org/10.1016/S0098-1354(02)00127-8)

- [22] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [23] Gers, F.A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: continual prediction with LSTM. *Neural Computation*, 12(10), 2451-2471. <https://doi.org/10.1162/089976600300015015>
- [24] Schuster, M., & Paliwal, K.K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673-2681. <https://doi.org/10.1109/78.650093>
- [25] Kiranyaz, S., et al. (2021). 1D convolutional neural networks and applications: a survey. *Mechanical Systems and Signal Processing*, 151, 107398. <https://doi.org/10.1016/j.ymssp.2020.107398>
- [26] Yin, R., Liu, B., Liu, H.R., & Li, Y.C. (2015). Research on invasive weed optimization based KELM for wind speed forecasting. *Energy Conversion and Management*, 104, 28-37.
- [27] Wang, H., Yi, H., Peng, J., Wang, G., Liu, Y., Jiang, H., & Liu, W. (2017). Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Conversion and Management*, 153, 409-422. <https://doi.org/10.1016/j.enconman.2017.10.008>
- [28] von Stosch, M., Oliveira, R., Peres, J., & Fayo de Azevedo, S. (2014). Hybrid semi-parametric modeling in process systems engineering: past, present and future. *Computers & Chemical Engineering*, 60, 86-101. <https://doi.org/10.1016/j.compchemeng.2013.08.008>
- [29] Bangi, M.S.F., & Kwon, J.S.I. (2020). Deep hybrid modeling of chemical process: application to hydraulic fracturing. *Computers & Chemical Engineering*, 134, 106696. <https://doi.org/10.1016/j.compchemeng.2019.106696>
- [30] Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1), 371-394. <https://doi.org/10.1146/annurev.mi.03.100149.002103>
- [31] Thompson, M.L., & Kramer, M.A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, 40(8), 1328-1340. <https://doi.org/10.1002/aic.690400806>
- [32] Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.
- [33] Raissi, M., Perdikaris, P., & Karniadakis, G.E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [34] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings ICNN 1995* (pp. 1942-1948). IEEE. <https://doi.org/10.1109/ICNN.1995.488968>
- [35] Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- [36] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(2), 281-305.
- [37] Snoek, J., Larochelle, H., & Adams, R.P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25, 2951-2959.
- [38] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings CVPR 2016* (pp. 770-778). IEEE. <https://doi.org/10.1109/CVPR.2016.90>
- [39] Cho, K., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings EMNLP 2014* (pp. 1724-1734). ACL. <https://doi.org/10.3115/v1/D14-1179>
- [40] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings ICLR 2015*. ICLR.
- [41] Luong, M.T., Pham, H., & Manning, C.D. (2015). Effective approaches to attention-based neural machine translation. In *Proceedings EMNLP 2015* (pp. 1412-1421). ACL. <https://doi.org/10.18653/v1/D15-1166>
- [42] Ioffe, S., & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proceedings ICML 2015* (pp. 448-456). PMLR.
- [43] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929-1958.
- [44] Kingma, D.P., & Ba, J. (2015). Adam: a method for stochastic optimization. In *Proceedings ICLR 2015*. ICLR.
- [45] Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings AISTATS 2010* (pp. 249-256). JMLR.
- [46] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley. <https://doi.org/10.1002/9781118619193>

- [47] Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts.
- [48] Smola, A.J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- [49] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- [50] Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In *Proceedings KDD 2016* (pp. 785-794). ACM. <https://doi.org/10.1145/2939672.2939785>
- [51] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S.Y. (2020). A comprehensive study on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4-24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- [52] Xu, J., Wu, H., Wang, J., & Long, M. (2021). Anomaly transformer: time series anomaly detection with association discrepancy. In *Proceedings ICLR 2022*. ICLR.
- [53] Zhou, H., et al. (2021). Informer: beyond efficient transformer for long sequence time-series forecasting. In *Proceedings AAAI 2021* (pp. 11106-11115). AAAI Press.
- [54] Lai, G., Chang, W.C., Yang, Y., & Liu, H. (2018). Modeling long- and short-term temporal patterns with deep neural networks. In *Proceedings SIGIR 2018* (pp. 95-104). ACM. <https://doi.org/10.1145/3209978.3210006>
- [55] Bai, S., Kolter, J.Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.
- [56] Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179-211. https://doi.org/10.1207/s15516709cog1402_1
- [57] Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850. <https://doi.org/10.48550/arXiv.1308.0850>
- [58] Tran, T., Nguyen, T.N., Tran, T.N., & Nguyen, T.L. (2017). Improving the fermented shrimp paste quality by combining different drying methods. *LWT-Food Science and Technology*, 76, 104-109.
- [59] Rawlings, J.B., Mayne, D.Q., & Diehl, M. (2017). *Model Predictive Control: Theory, Computation, and Design* (2nd ed.). Nob Hill Publishing.
- [60] Garcia, C.E., Prett, D.M., & Morari, M. (1989). Model predictive control: theory and practice -- a survey. *Automatica*, 25(3), 335-348. [https://doi.org/10.1016/0005-1098\(89\)90002-2](https://doi.org/10.1016/0005-1098(89)90002-2)
- [61] Seborg, D.E., Mellichamp, D.A., Edgar, T.F., & Doyle, F.J. (2011). *Process Dynamics and Control* (3rd ed.). Wiley.
- [62] Nelles, O. (2001). *Nonlinear System Identification*. Springer. <https://doi.org/10.1007/978-3-662-04323-3>
- [63] Qin, S.J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2), 220-234. <https://doi.org/10.1016/j.arcontrol.2012.09.004>
- [64] Peng, D., Liu, Z., Wang, H., Qin, Y., & Jia, L. (2019). A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access*, 7, 10278-10293. <https://doi.org/10.1109/ACCESS.2018.2888842>
- [65] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, 17(2), 425. <https://doi.org/10.3390/s17020425>
- [66] Liu, H., Mi, X., & Li, Y. (2018). Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Conversion and Management*, 159, 54-64. <https://doi.org/10.1016/j.enconman.2018.01.010>
- [67] Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, 452, 48-62. <https://doi.org/10.1016/j.neucom.2021.03.091>
- [68] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings CVPR 2018* (pp. 7132-7141). IEEE. <https://doi.org/10.1109/CVPR.2018.00745>
- [69] Woo, S., Park, J., Lee, J.Y., & Kweon, I.S. (2018). CBAM: convolutional block attention module. In *Proceedings ECCV 2018* (pp. 3-19). Springer. https://doi.org/10.1007/978-3-030-01234-2_1
- [70] Kitaev, N., Kaiser, L., & Levskaya, A. (2020). Reformer: the efficient transformer. In *Proceedings ICLR 2020*. ICLR.
- [71] Li, S., et al. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in Neural Information Processing Systems*, 32, 5244-5254.
- [72] Oord, A.V.D., et al. (2016). WaveNet: a generative model for raw audio. arXiv preprint arXiv:1609.03499.

- [73] Guo, T., Xu, Z., Yao, X., Chen, H., Aberer, K., & Funaya, K. (2016). Robust online time series prediction with recurrent neural networks. In *Proceedings IEEE BigData 2016* (pp. 816-825). IEEE. <https://doi.org/10.1109/BigData.2016.7840681>
- [74] Williams, C.K., & Rasmussen, C.E. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- [75] Lim, B., Arik, S.O., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748-1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [76] Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181-1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>