Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale, Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

A Physics-Informed Approach to Dynamic Modeling and Parameter Estimation in Biotechnology

Konstantinos Mexis[a*], Stefanos Xenios[a], Nikolaos Trokanas[a] and Antonis Kokossis[a]

[a] National Technical University of Athens, Department of Chemical Engineering, Athens, Greece
* Corresponding Author: konstantinosmexis@mail.ntua.gr.

ABSTRACT

The increasing complexity of industrial biotechnology demands advanced modeling techniques capable of capturing the intricate dynamics of bioreactors. Traditional regression-based and empirical methods often fall short when confronted with the highly nonlinear behavior and limited experimental data characteristic of bioprocesses. Addressing these challenges requires a more intelligent approach—one that leverages domain knowledge to model complex bioprocess dynamics effectively, even with sparse data, while maintaining interpretability and robustness. In this study, we introduce a process-informed, data-driven methodology for modeling the dynamics of industrial bioreactors, leveraging the capabilities of the rising field of Scientific Machine Learning (SciML). Our approach leverages Physics-Informed Neural Networks (PINNs) to seamlessly integrate domain knowledge encoded in physical laws with sparse experimental data and deep learning techniques, enabling precise simulation and modeling of bioreactor operations. The framework is validated using real-world experimental data, demonstrating its capability for state and space estimation, essential for optimizing biomanufacturing processes. Our findings underscore the potential of Physics Informed Neural Networks to address the challenges of dynamic modeling in bioprocesses, paving the way for the development of intelligent, autonomous, and data-driven solutions in industrial biotechnology. This fusion of data-driven and mechanistic modeling represents a transformative step toward enhancing the efficiency, scalability, and sustainability of modern biomanufacturing practices.

Keywords: Machine Learning, Simulation, Dynamic Modelling, Industry 4.0, Intelligent Systems

INTRODUCTION

The increasing complexity of modern industrial biotechnology underscores the urgent need for advanced modeling and simulation techniques. Many biomanufacturing processes still rely heavily on empirical knowledge and operator expertise, which limits their ability to effectively handle the increasing intricacies and dynamic nature of contemporary bioprocesses. This dependency on traditional methods constrains the scalability, efficiency, and optimization of biomanufacturing, highlighting the need for more data-driven and adaptive approaches. As bioprocess systems become more sophisticated, traditional control strategies struggle to address emerging challenges, including raw material variability, unpredictable biological dynamics, and the imperative for real-time, data-driven decision-making. The advent of Industry 4.0—characterized by digitalization, automation, and intelligent data integration—has initiated a transformative shift towards smarter, more efficient, sustainable, and adaptive bioprocessing systems [4]. Digital transformation in bioprocesses involves integrating multiple technologies, yet the inherent complexity, high dimensionality, and dynamic nature of these systems, driven by numerous interacting parameters, pose significant challenges in optimizing process efficiency, maintaining stability, and maximizing product yields [6].

Artificial Intelligence (AI) is gaining increasing recognition for its potential to revolutionize bioprocesses modeling by effectively leveraging real-world experimental data. Recently, a growing number of AI technologies have been applied in bioprocess optimization and control to enhance the performance of bioprocess. These advancements demonstrate that AI holds great promise for enhancing bioprocess efficiency, scalability and adaptability, making it a valuable tool for addressing the challenges faced in modern biomanufacturing [6]. However, conventional machine learning approaches, such as neural networks, face significant limitations in this domain due to their reliance on large datasets to achieve optimal performance. In bioreactor applications, the availability of data is often sparse and limited, posing a critical challenge for purely data-driven methodologies. Moreover, traditional neural networks operate as black-box models, lacking the ability to incorporate physical knowledge of the underlying bioprocess, which is essential for ensuring reliability and interpretability in complex biomanufacturing systems.

To address these challenges, in this study, we introduce a novel approach that leverages the principles of Scientific Machine Learning (SciML) to overcome the challenges of bioreactor modeling. Specifically, we employ Physics-Informed Neural Networks (PINNs), a cutting-edge framework that integrates physical knowledge directly into the machine learning process. Unlike traditional neural networks, PINNs embed governing equations, such as mass balance, energy balance, and reaction kinetics, as constraints within the learning architecture. This enables the model to respect the underlying physics of the bioprocess while learning from limited experimental data. In addition to enabling accurate dynamic modeling with limited data, the Physics-Informed Neural Network framework offers the capability to perform parameter estimation directly within its architecture. Traditional methods for parameter estimation in bioprocesses often require extensive experimental data or rely on iterative optimization techniques that can be computationally expensive and time-consuming. In contrast, PINNs naturally incorporate parameter estimation by treating unknown process parameters as trainable variables within the network. We validate our approach using a dynamic model of a batch bioreactor and real-life experimental data. Our demonstration highlights the effectiveness of PINNs in achieving accurate dynamic modeling and parameter estimation, even with very limited experimental data.

METHODS AND MATERIALS

Dynamic Model

As a pilot study, we expanded upon the work of [1], which successfully demonstrated the production of recombinant nanobodies with scorpion antivenom activity in Escherichia coli. Nanobodies (VHH), derived from camelid antibodies, were produced intracellularly in E. coli WK6 through fed-batch cultivation, with production yields significantly improved using dynamic modeling techniques [2]. In our study, experimental data from the batch phase of the bioreactor were utilized to develop a predictive model using Physics-Informed Neural Networks (PINNs). The bioreactor is assumed to be an infinitely mixed culture without any transfer limitation. The mass balances for biomass and glucose are given by Eq. (1) and (2):

$$\frac{\mathrm{dX}}{\mathrm{dt}} = \frac{\mu_{max} \cdot S}{K_S + S} \cdot X \tag{1}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{XS}} \cdot \frac{\mu_{max} \cdot S}{K_S + S} \cdot X \tag{2}$$

where X and S are the concentrations of biomass and glucose (in g/L), μ_{max} is the maximum specific growth rate (in h^{-1}), K_S is the substrate saturation coefficient (in g/L) and Y_{XS} is the yield coefficient of substrate (in $g \cdot cdw/gS$).

Scientific Machine Learning

Scientific Machine Learning (SciML) is an innovative and rapidly evolving field that merges principles from machine learning (ML) with scientific computing to tackle complex challenges across various domains, including physics, biology, and engineering [5]. This interdisciplinary approach leverages the strengths of both data-driven methods and traditional physics-based models, aiming to enhance predictive capabilities and facilitate data-driven discoveries. This integration enables more accurate simulations and predictions, particularly when the under-lying physics of a system is only partially known or too complex to model explicitly.

By incorporating scientific knowledge, SciML techniques require less data to achieve accurate and robust predictions, which is especially beneficial when collecting large amounts of data is expensive or impractical. Furthermore, these methods maintain interpretability by embedding physical principles and domain knowledge into the models, allowing for better understanding of the underlying processes and greater trust in the predictions they provide.

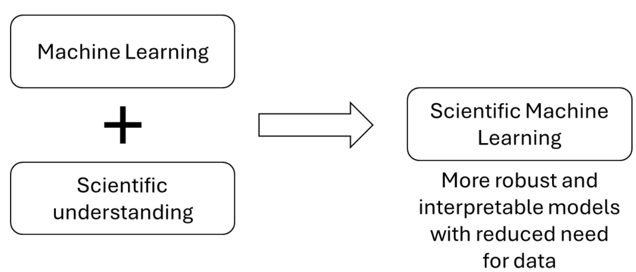

Figure 1: Scientific Machine Learning

Physics-Informed Neural Network

Physics-Informed Neural Networks (PINNs) represent a significant advancement in the intersection of machine learning and computational physics, providing a framework for solving complex mathematical models governed by partial differential equations (PDE).

Introduced in [3], PINNs leverage the capabilities of deep learning to incorporate physical laws directly into the learning process, enabling the networks to approximate solutions to both forward and inverse problems effectively. This approach allows one to automatically find data-driven solutions for PDEs or ODEs. In this paper, we consider parametrized nonlinear system of ODEs given by:

$$\mathcal{F}\left(t, \boldsymbol{y}(t), \frac{dy(t)}{dt}; \boldsymbol{\lambda}\right) = 0 \tag{3}$$

where $\boldsymbol{y}$ represents the state of the dynamic system (i.e., the latent ODE solution) and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_2]$ are the ODE parameters. In the context of PINNs, a fully connected feed-forward neural network with multiple hidden layers is employed to approximate the solution of the differential equations. The network takes time t as input and outputs:

$$\boldsymbol{y} = f_w(t) \tag{4}$$

where f_w represents the mapping function learned by the neural network, parametrized by adaptive weights $\boldsymbol{w}$. This setup requires the network to learn to compute the solution of the given ODEs. The neural network in a PINN framework is trained by minimizing a mean squared error (MSE) cost function, typically defined as:

$$MSE = \mathcal{L}_{data} + \mathcal{L}_{physics} \tag{5}$$

where $\mathcal{L}_{data}$ represents the data-based loss term, and $\mathcal{L}_{physics}$ enforces the physical consistency of the solution. These loss terms are formulated as follows:

$$\mathcal{L}_{data} = \frac{1}{N_y}\sum_{i=1}^{N_y}\frac{1}{N_t}\sum_{j=1}^{N_t}\left\|y_i(t_j) - y_i^*\right\|_2^2 \tag{6}$$

$$\mathcal{L}_{physics} = \frac{1}{N_y}\sum_{i=1}^{N_y}\frac{1}{N_{\mathcal{F}}}\sum_{j=1}^{N_{\mathcal{F}}}\left\|\mathcal{F}(y_i(t_j))\right\|_2^2 \tag{7}$$

where, N_t,$N_{\mathcal{F}}$ and N_y correspond to the number of training data points, the number of collocation points and the number of outputs of the neural network respectively; y_i is the i-th output of the neural network; y_i^* represents the desired i-th output, considering the j-th data pair $(t_j, y_i^*(t_j))$. The data-based loss term penalizes discrepancies between the network's predictions and the desired outputs from the training data. The physics-based loss term enforces the alignment of the network's outputs with the governing differential equations of the physical system by minimizing the residuals of these equations at a finite set of randomly sampled collocation points. Together, these terms allow the network to approximate the solution of the system's dynamics while adhering to physical laws. By incorporating prior knowledge through the governing equations, the PINN framework reduces the reliance on large datasets, enabling it to perform accurate dynamic modeling and parameter estimation even with sparse data. This integration of theoretical models with experimental data enhances the efficiency and robustness of process modeling, making PINNs particularly suitable for complex, data-limited systems. PINNs can also be employed in a inverse setup to perform parameter estimation of the governing ODEs. In this configuration, unknown parameters within the ODEs are treated as additional trainable variables within the neural network. By incorporating these parameters into the optimization process, the network learns not only the dynamic behavior of the system but also estimates the values of the unknown parameters. This process enables PINNs to infer parameters even in cases where traditional methods struggle due to sparse or noisy data.

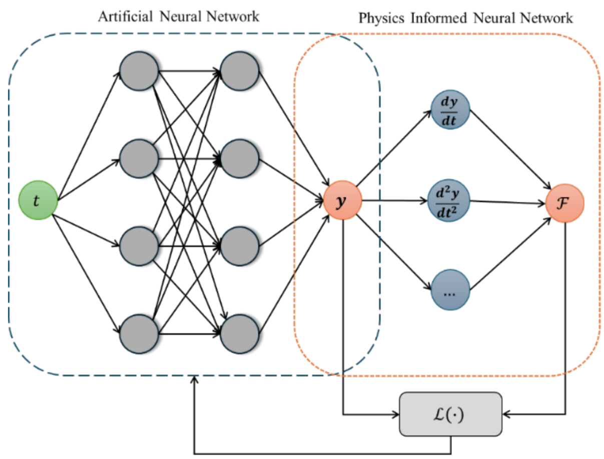

Figure 2: Physics Informed Neural Network

RESULTS

Leveraging experimental data from [1] and the dynamic model described previously, we applied our approach to a batch bioreactor, showcasing the ability of Physics-Informed Neural Networks (PINNs) to accurately capture the system's state-space dynamics and unlock predictive insights, thereby advancing data-driven intelligence in bioprocess modeling. The dataset consists of measurements of biomass and glucose concentrations over time in a batch bioreactor (**Figure 3**).

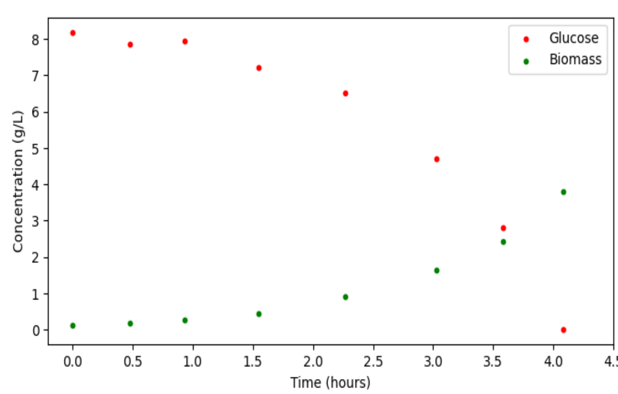

Figure 3: Experimental data for batch phase

The results demonstrate that this physics-informed framework effectively overcomes the key challenges of conventional dynamic modeling methods, particularly those arising from sparse and limited experimental datasets. The PINN architecture utilized in this study is a fully connected feed-forward neural network comprising one input layer, three hidden layers (with 64, 1024 and 1024 neurons) and one output layer. The $ReLU$ activation function was applied to the input and final hidden layers, while the $Tanh$ activation function was used for the intermediate layers, a combination that enhanced the network's accuracy and overall performance. The network was trained using the $Adam$ optimizer with a learning rate of 0.0001 over $30{,}000$ epochs, with early stopping implemented to prevent overfitting.

The network's single input was the time (t) variable, and its output layer provided predictions for Biomass (X) and Glucose (S). To capture the system's underlying dynamics, 50 collocation points were used during training. These points enforced the governing ODEs, ensuring that the PINN not only fit the available data but also adhered to the physical laws governing the system. The PINN was trained by solving the optimization problem:

$$\underset{\boldsymbol{w},\boldsymbol{p}}{minimize}\, \mathcal{L} \tag{8}$$

$$\mathcal{L} = \mathcal{L}_{data} + \mathcal{L}_{physics} \tag{9}$$

where

$$\mathcal{L}_{data} = \frac{1}{2}\cdot\frac{1}{N_y}\left(\sum_{j=1}^{N_y}\left\|X(t_j) - X^*\right\|_2^2 + \sum_{j=1}^{N_y}\left\|S(t_j) - S^*\right\|_2^2\right) \tag{10}$$

and

$$\mathcal{L}_{physics} = \frac{1}{2}\cdot\frac{1}{N_{\mathcal{F}}}\left(\sum_{j=1}^{N_{\mathcal{F}}}\left\|\frac{dX(t_j)}{dt} - \frac{\mu_{max}\cdot S(t_j)}{K_S + S(t_j)}\cdot X(t_j)\right\|_2^2 + \right.$$
$$\left. + \sum_{j=1}^{N_{\mathcal{F}}}\left\|\frac{dS(t_j)}{dt} + \frac{1}{Y_{XS}}\cdot\frac{\mu_{max}\cdot S(t_j)}{K_S + S(t_j)}\cdot X(t_j)\right\|_2^2\right) \tag{11}$$

In the equations (8)-(11), $\boldsymbol{w}$ are the neural network's trainable weights and $\boldsymbol{p}$ are the unknown kinetic parameters of the dynamic model. $\mathcal{L}_{data}$ represents the data-driven loss function, measuring how well the predictions of the model match the experimental data at specific time points. Finally, $\mathcal{L}_{physics}$ represents the physics-driven loss function, which ensures that the model adheres to the physical laws governing the bioprocess system. By minimizing the loss function ($\mathcal{L}$) with respect to both the network weights ($\boldsymbol{w}$) and the unknown kinetic parameters ($\boldsymbol{p}$) of the dynamic model, we simultaneously train the neural network and perform parameter estimation for the dynamic system. Addressing collinearity between network weights and kinetic parameters is essential to ensure that the estimated parameters accurately represent the underlying system dynamics, rather than being influenced by artifacts of the optimization process. Techniques such as regularization, orthogonalization, or introducing additional constraints can help mitigate this issue and improve the robustness of the parameter estimation process.

Parameter Estimation

We first demonstrate the capability of the PINN to accurately estimate the unknown kinetic parameters of the dynamic model using only a sparse set of measurements. Estimating these parameters is a critical step in effectively modeling and understanding the complex dynamics of bioprocesses, as they govern key aspects such as microbial growth rates and substrate consumption.

By employing the PINN in an inverse configuration, we successfully performed parameter estimation, with the resulting kinetic parameter estimates summarized in next table (**Table 1**).

Table 1: Estimated kinetic parameters

Kinetic parameter	Value
μ_{max}	0.7267
K_S	0.1634
Y_{XS}	0.3983

As a reference, the dynamic model calibrated parameters reported in [1] are presented in **Table 2**.

Table 2: Model parameters reported in literature

Kinetic parameter	Min. value	Max. value
μ_{max}	0.67	0.87
Y_{XS}	0.29	0.47

These findings highlight the ability of PINNs to accurately estimate unknown kinetic parameters, even when experimental data is sparse, showcasing their potential for applications in bioprocess modeling and control. In contrast to traditional parameter estimation approaches, which often require the full trajectory of data—available only at the end of the process—PINNs can achieve accurate parameter estimation using only the available data points at any time during the process. This allows for a more adaptive modeling approach, particularly in bioprocesses where real-time data collection is limited or costly. As a result, kinetic parameters can now be described as dynamic functions of time rather than fixed constants, providing a more realistic and accurate representation of the evolving nature of complex bioprocesses.

Simulation

In this paragraph, we demonstrate the ability of PINNs to perform accurate simulations of the bioprocess. By leveraging both sparse experimental data and the governing equations of the system, the PINN effectively captures the complex dynamics of the process. Accurate simulation is vital for understanding the system's behavior, predicting outcomes under various operating conditions, and optimizing bioprocess performance. This capability is particularly important in biomanufacturing, where precise control of dynamic variables such as biomass and substrate concentrations are crucial for achieving consistent product quality and yield. Unlike conventional regression approaches, which rely solely on fitting data and often require large datasets to achieve acceptable accuracy, PINNs integrate prior physical knowledge into the learning process. This integration significantly reduces the dependency on extensive datasets, making PINNs especially well-suited for applications where experimental data is limited or costly to obtain. Furthermore, the physics-based loss function in PINNs ensures that the model predictions adhere to the underlying laws of the system, improving the interpretability and robustness of the simulation results.

The simulation performance of the PINN is evaluated leveraging the mean squared error (MSE) between the predicted state variables and the experimentally measured values for both biomass and glucose. This performance metric provided a quantitative assessment of the model's ability to replicate the observed dynamics of the bioreactor (**Table 3**).

Table 3: PINN simulation performance

Number of data points	Biomass (MSE)	Glucose (MSE)
8	0.012	0.018

The low MSE values indicate that the PINN is capable of accurately modeling the dynamics of the batch bioprocess. A key advantage of the PINN also lies in its interpretability, as it incorporates prior process knowledge in the form of governing ODEs directly into the modeling framework.

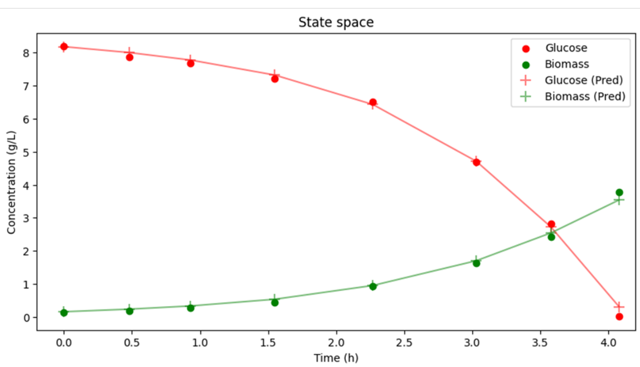

Figure 4: PINN simulation results

Figure 5 presents a comparison of the simulation results obtained by solving the system of ODEs using kinetic parameter values estimated through a traditional regression approach [1] versus the simulation results obtained through PINN. The values of kinetic parameters estimated using standard regression techniques and used for simulating the system dynamics are $\mu_{max} = 0.77$, $K_s = 0.214$ and $Y_{xs} = 0.40$. The results clearly demonstrate that the PINN provides a better fit to the experimental data points compared to the traditional regression-based approach. This improved accuracy in state estimation is achieved because PINNs integrate both sparse experimental data and fundamental process knowledge encoded in the system's governing equations. Unlike pure data-driven methods, which may struggle with limited or noisy datasets, PINNs leverage physical constrants to guyide learning, ensuring consisntesy with known bioprocess dynamics. The superiority of PINNs becomes even more evident in scenarios where fewer experimental data points are available or when modeling occurs in real-time rather than at the end of the process. Traditional regression-based methods typically require a complete dataset spanning the entire process to provide reliable parameter estimates. In contrast, PINNs can extract meaningful insights even from sparse or incomplete data by leveraging the underlying physics of the system.

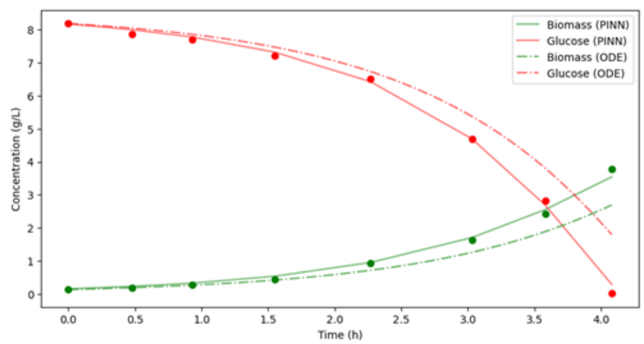

Figure 5: Simulation results of PINN and traditional regression approaches

CONCLUSIONS

Accurate modeling of complex bioprocess systems is hindered by several key challenges. One of the primary issues is the sparse availability of experimental data, as generating comprehensive datasets can be costly and time. Additionally, traditional black-box data-driven methods often lack interpretability, making it difficult to understand and the underlying predictions or to derive actionable insights. "Furthermore, the inherently complex and highly non-linear nature of bioprocesses presents additional challenges, making it difficult to develop robust and accurate models that can effectively capture the intricate and dynamic behaviors involved. n this study, we aim to address these challenges by leveraging Physics-Informed Neural Networks (PINNs), a powerful tool for modeling and simulating the dynamics of complex bioprocesses. By integrating prior knowledge in the form of governing ODEs with sparse experimental data, PINNs offer a robust framework for accurate dynamic modeling and parameter estimation.

The results obtained from a batch bioreactor system highlight the ability of PINNs to capture intricate state-space dynamics, even when data is limited, and to provide interpretable, physics-consistent predictions. Furthermore, the application of PINNs in an inverse setup for estimating kinetic parameters underscores their versatility and effectiveness in addressing one of the key challenges in bioprocess modeling. Compared to conventional regression approaches, PINNs not only reduce the dependency on extensive datasets but also ensure adherence to the physical principles of the system, enhancing both reliability and interpretability.

This integration of machine learning and scientific principles represents a significant step toward more intelligent, data-driven solutions for industrial biotechnology. Future work will focus on extending this approach to more complex bioprocess systems, incorporating real-time data, and exploring adaptive learning strategies to further enhance the scalability and applicability of PINNs in dynamic modeling and control.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme BIOINDUSTRY 4.0 (grant agreement 101094287)

REFERENCES

1. Alonso Villela, S. M., Ghezal-Kraïem, H., Bouhaouala-Zahar, B., Bideaux, C., Aceves Lara, C. A., & Fillaudeau, L. Effect of temperature on the production of a recombinant antivenom in fed-batch mode. *Applied Microbiology and Biotechnology*, 105(3), 1017–1030 (2021). https://doi.org/10.1007/s00253-021-11093-5
2. Corrales, D. C., Villela, S. M. A., Bouhaouala-Zahar, B., Cescut, J., Daboussi, F., O'donohue, M., Fillaudeau, L., & Aceves-Lara, C. A. *Dynamic Hybrid Model for Nanobody-based Antivenom Production (scorpion antivenon) with E. coli CH10-12 and E. coli NbF12-10* (pp. 145–150) (2024). https://doi.org/10.1016/B978-0-443-28824-1.50025-9
3. Raissi, M., Perdikaris, P., & Karniadakis, G. E. 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 (2019). https://doi.org/10.1016/j.jcp.2018.10.045
4. Isoko, K., Cordiner, J. L., Kis, Z., & Moghadam, P. Z. Bioprocessing 4.0: A Pragmatic review and Future Perspectives. *Digital Discovery*, *3*(9), 1662–1681. (2024). https://doi.org/10.1039/d4dd00127c
5. Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli., Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next. *J. Sci. Comput*. 92, 3 (2022) https://doi.org/10.1007/s10915-022-01939-z
6. Cheng, Y., Bi, X., Xu, Y., Liu, Y., Li, J., Du, G., Lv, X., & Liu, L. Artificial intelligence technologies in bioprocess: Opportunities and challenges. Bioresource Technology, 369, 128451 (2023). https://doi.org/10.1016/j.biortech.2022.128451
7. Khanal, S. K., Tarafdar, A., & You, S. Artificial intelligence and machine learning for smart bioprocesses. Bioresource Technology, 375, 128826 (2023). https://doi.org/10.1016/j.biortech.2023.128826