Bioresource Technology 369 (2023) 128451

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech

Artificial intelligence technologies in bioprocess: Opportunities and challenges

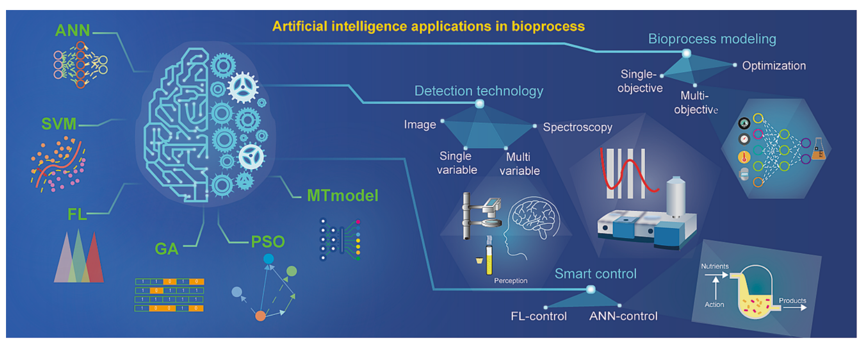
Yang Cheng [a,b], Xinyu Bi [a,b], Yameng Xu [a,b], Yanfeng Liu [a,b], Jianghua Li [a,b], Guocheng Du [a,b], Xueqin Lv [a,b], Long Liu [a,b,*]

[a] *Key Laboratory of Carbohydrate Chemistry and Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, China*
[b] *Science Center for Future Foods, Ministry of Education, Jiangnan University, Wuxi 214122, China*

HIGHLIGHTS

- AI has been extensively applied in bioprocess modeling and optimization.
- Convenient, accurate and rapid monitoring technologies are emerging.
- Advanced control strategies based on AI technologies have been proposed.
- Integrating AI with GEMs and CFD will further improve bioprocess.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Artificial intelligence
Bioprocess modeling
Rapid detection
Real-time monitoring
Smart control

ABSTRACT

Bioprocess control and optimization are crucial for tapping the metabolic potential of microorganisms, and which have made great progress in the past decades. Combination of the current control and optimization technologies with the latest computer-based strategies will be a worth expecting way to improve bioprocess further. Recently, artificial intelligence (AI) emerged as a data-driven technique independent of the complex interactions used in mathematical models and has been gradually applied in bioprocess. In this review, firstly, AI-guided modeling approaches of bioprocess are discussed, which are widely applied to optimize critical process parameters (CPPs). Then, AI-assisted rapid detection and monitoring technologies employed in bioprocess are summarized. Next, control strategies according to the above two technologies in bioprocess are analyzed. Lastly, current research gaps and future perspectives on AI-guided optimization and control technologies are discussed. This review provides theoretical guidance for developing AI-guided bioprocess optimization and control technologies.

* Corresponding author at: Key Laboratory of Carbohydrate Chemistry and Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, China.
E-mail address: longliu@jiangnan.edu.cn (L. Liu).

https://doi.org/10.1016/j.biortech.2022.128451
Received 30 October 2022; Received in revised form 1 December 2022; Accepted 3 December 2022
Available online 9 December 2022

1. Introduction

As the global transition from a fossil-based to a bio-based economy continues, the number of industrial bioprocesses employed to produce biofuels, materials, and healthcare products is steadily growing (Mol et al., 2021). Bioprocess is any process that uses living cells or their components to produce value-added products (Patel et al., 2020). However, low bioconversion and productivity are usually bioprocess limitations (Xu et al., 2021). Furthermore, the strain's performance of the bioprocess is limited by its metabolic characteristics and external environmental conditions. For instance, one big challenge called the "scale-up effect" existed during the transfer from the lab scale to the industrial scale, which may be accompanied by the poor performance of bioprocess (Wang et al., 2020), resulting in reduced economic benefits. Therefore, bioprocess often need to be further optimized to achieve better performance and control to eliminate environmental conditions' negative impact.

In the past few decades, significant progress has been made in bioprocess control and optimization. For example, orthogonal experimental design (OED) and response surface methodology (RSM) were extensively applied to determine the best parameters to achieve a satisfactory performance of bioprocesses (Feng et al., 2020; Sharma et al., 2021). Besides, machine learning (ML) algorithms have been gradually employed to study the nonlinear relationship between variables in bioprocess. For example, an artificial neural network (ANN) was applied to model and optimize poly (3-hydroxybutyrateco-3-hydroxyvalerate) production through fermentation (Zafar et al., 2012); Radial basis function neural work (RBF-ANN) and particle swarm optimization (PSO) was used in the optimization of hyaluronic acid production (Liu et al., 2009). Additionally, real-time monitoring technologies are also developing rapidly. For example, a well-established inline biomass concentration detection technology was achieved by alternating electric fields produced using a probe (Justice et al., 2011). Proteins marked with fluorophores for optical transduction have been used to take advantage of binding-based sensors for substrates (Ge et al., 2004). Despite the notable developments in bioprocess control and optimization, some potential problems will limit its further applications. For instance, it is difficult to accurately predict complex nonlinear systems using shallow ML or standard modeling techniques. Additionally, the added value of some inline detection technology and equipment, such as viable cell biosensors and *in situ* microscopy, is not justified except in research (Reardon, 2021). Furthermore, advanced control strategies like adaptive control and model-based control find limited industrial applications thus far (Rathore et al., 2021).

Artificial Intelligence (AI) is a computer science that attempts to imitate human thinking to solve problems (Hamet and Tremblay, 2017). AI programs can make independent decisions through pre-determined rules or data pattern identification (Khanijou et al., 2022). Recently, an increasing number of AI technologies have been applied in bioprocess optimization and control to enhance the performance of bioprocess. For instance, a hybrid multi-objective strategy was used to simultaneously optimize the biomass and production yield of microorganisms for industrial production (Saini et al., 2021). Furthermore, deep learning approaches were applied to predict critical process parameters (CPPs) according to online sensor data in anaerobic digestion (AD) (Jia et al., 2022); image recognition technology is also applied for the rapid detection of compost maturity (Xue et al., 2019). These studies show that AI has a good bioprocess optimization and control prospect. This review first discusses recent contributions of AI-guided modeling and optimization technologies. After that, AI-assisted rapid detection and monitoring technology applications were introduced and analyzed. Advanced control technologies based on the previously mentioned technologies were also summarized. Finally, existing problems are summarized and the future directions are prospected.

2. AI-guided modeling and optimization of bioprocess

Bioprocess involves complex, multiscale and high-dimensional conversion processed across time and space, which is the consequence of the interaction of many process parameters (Wang et al., 2020). Therefore, by developing theory-driven or data-driven models, it is possible to anticipate the viability of a bioprocess and its ability to reduce workload. In the past few decades, many classical methods, such as OED and RSM, have been applied in bioprocess modeling. However, because of the complexity of the bioprocess, it is usually impossible to build accurate models using this method. Therefore, AI has been introduced as a facilitating approach to build a data-driven model to describe bioprocess accurately.

2.1. Artificial neural network (ANN)

With the typical characteristics of strong adaptability and fault tolerance, ANNs are good at dealing with complex nonlinear problems. The ANN consists of input layer, hidden layer, and output layer, which are very close to the human brain's neural network from an information processing perspective (Andrade Cruz et al., 2022). As shown in Fig. 1, the neurons in each layer are linked to the next layer according to penalty function (Guo et al., 2021). The performance of an ANN is mainly impacted by structure and learning algorithms. The ANN model can be divided into feedforward neural network (FFNN), convolution neural network (CNN), and recurrent neural network (RNN) based on its structure. Moreover, the ANN of each structure can be further classified based on the learning algorithm and the mode of the neurons. For example, a back-propagation neural network (BP-ANN) is an FFNN trained by a supervised learning algorithm named back-propagation. Likewise, multilayer perceptron neural networks (MLP-ANN) and RBF-ANN also belong to the FFNN.

As the most widely applied ML method, ANN is often employed to optimize the pretreatment process. Since the changes in pretreatment often involve multiple scales, CPPs always are nonlinear with and critical quality attributes (CQAs). Recently, an ANN network and an RSM were employed to build the total reducing sugar predictive model to test performance of these two models in pretreatment (Singhal et al., 2018). Both RSM ($r^2 = 0.996$) and ANN ($r^2 = 0.998$) had high correlation coefficient. However, root mean square error (RMSE) (5.564) and standard error of prediction (SEP) (2.294 %) for the RSM model were larger than those for the ANN model, 3.630 and 1.908 %, indicating that the ANN provides a better approximation of pretreatment process. Additionally, ANN is used to contrast various pretreatment strategies. For example, microwave-based and steam-based ANN models were constructed to explore optimal CQAs (Moodley et al., 2019). Although these models have not been optimized to determine the best CPPs in their research, it is still an attractive idea to determine best experimental method by comparing different models *in silico*. Moreover, ANN is also widely used to optimize the fermentation process. For instance, carbon and nitrogen sources have been optimized according to an FFNN paradigm (Zafar et al., 2012). Besides, a different study focuses on realizing time-dependent fermentation control strategies for enhancing production. To achieve the objective, time was regarded as an input nod of ANN models. Furthermore, the genetic algorithm (GA) was employed to determine optimal control trajectories of fermentation parameters (Peng et al., 2013). Accurate simulation of time sequence data may be critical to the further development of smart control technologies. DNN was also employed to simulate the enrichment of microorganisms with specific functions in control of multi-species communities (co-culture) systems at different times (Treloar et al., 2020). Unlike the previous studies, reinforcement learning was directly applied to learn the relationship between control behavior and consequence. In addition, conventional kinetic model was applied to generate training and testing datasets for DNN, which indicated that hybrid model integrated with kinetic model and DNN enables the accurate simulation of long-term dynamic

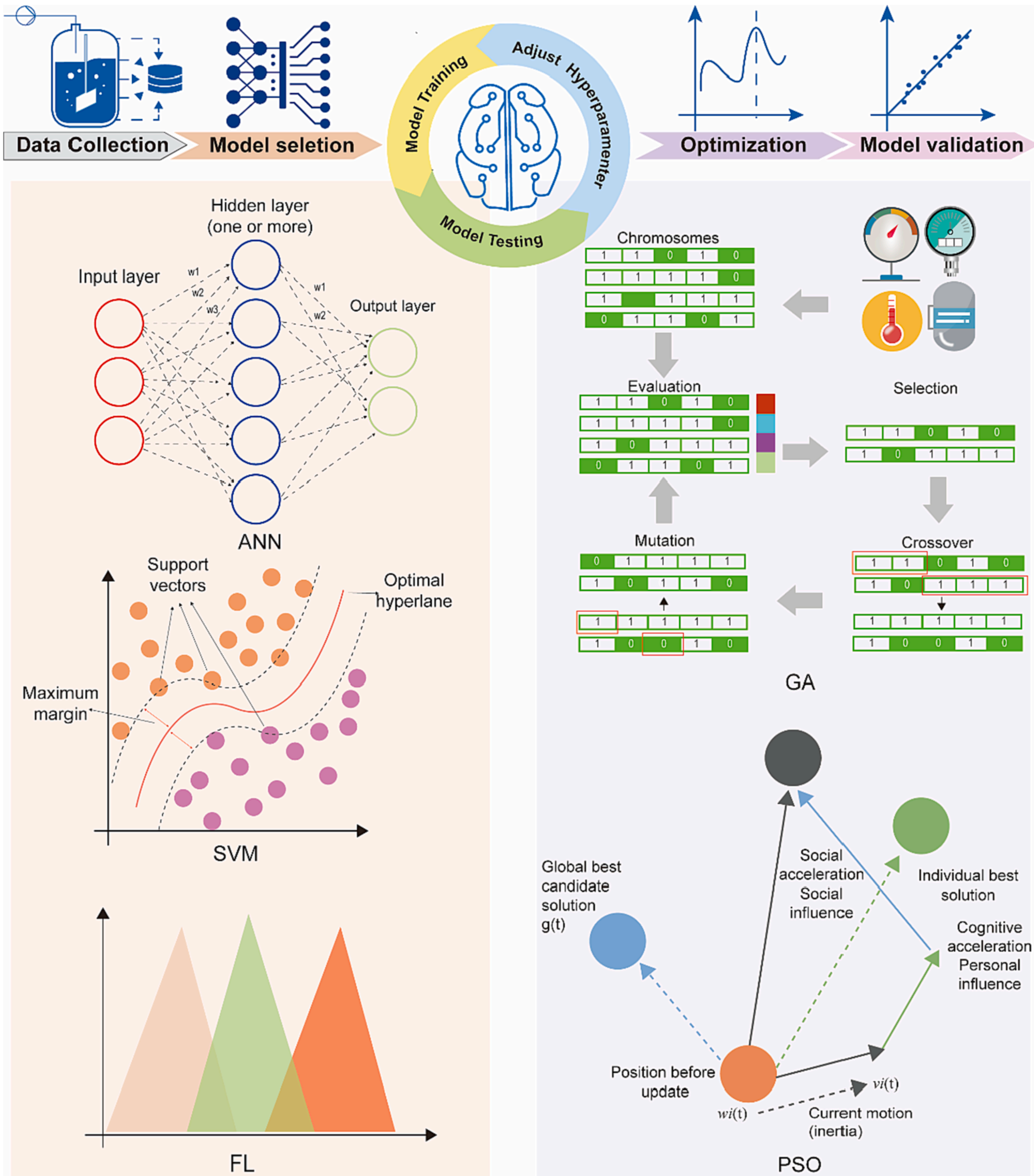


Fig. 1. Bioprocess modeling and optimization procedure and schematic diagram of artificial neural network (ANN), support vector machine (SVM), Fuzzy logic (FL), genetic algorithm (GA) and Particle swarm optimization (PSO).

enrichment of microbial communities in co-culture systems (Xu et al., 2022). More than this, kinetic models could be further exploited by ANN. In a recent study, pyrolysis kinetics were generated by Chemistry-Informed Neural Networks (CINN), where a database containing thermogravimetric analysis measurements was regarded as training data (Xing et al., 2022). ANN has also been applied in a downstream process to maximize the extraction efficiency of biodiesel by optimizing the CPPs (Vinoth Arul Raj et al., 2021).

ANNs have many successful applications in bioprocessing, but there are also significant limitations. ANN is a “black box” model, which cannot provide the basis for modeling (Kannangara et al., 2018). ANNs use experimental data to adjust model parameters through repeated iterations rather than understanding the essence behind the change. This limitation is surely unfriendly to researchers even though the ANN model always leads to a better fit than conventional methods (Ascher et al., 2022; Jadhav et al., 2021). Another issue is choosing the correct number of neurons for the hidden layer. Generally, an increase in the number of neurons in ANN often result in better learning performance. However, too many neurons may result in an overfitting phenomenon (Alrashed et al., 2018). Therefore, there is a trade-off between learning and generalization abilities of ANN.

2.2. Support vector machine (SVM)

SVM is an advanced supervised ML algorithm first reported by Cortes and Vapnik (Cortes and Vapnik, 1995). Input parameters are mapped into multi-dimensional feature space during data processing based on a nonlinear mapping kernel function. Subsequently, the optimal hyperplane maximizes the distance between the nearest data points and the hyperplane (Fig. 1). The sample point closest to the hyperplane is called the support vector (Guo et al., 2021). SVM uses the structural risk minimization (SRM) approach, which is opposed to the empirical risk minimization (ERM) principle applied in ANN (Sakr et al., 2016). Therefore, the model trained by SVM is less likely to be overfitted when the train set is limited. Currently, SVM has been considered as one of the most powerful tools for solving the issues of limited dataset fitting and classification.

Benefiting from the SRM principle, SVM is good at handling nonlinear, multi-parameter, and small data size problems. For instance, the SVM-based model exhibited more accurate predictions than ANN-based model in a small dataset (Alejo et al., 2018). Furthermore, Zhang et al. used SVM and BP-ANN to construct the lipid production models (Zhang et al., 2020). The SVM model can still successfully fit the data set, even when the dataset only has 77 samples. Additionally, to further improve SVM performance in a limited dataset, the hybrid model was constructed by integrating classical methods and SVM. For example, OED factors were used to design and arrange the training dataset. LS-SVM was used to model the anaerobic fermentation of corn stalks (Dong and Chen, 2019). The cumulative biogas production of the optimal conditions increased by 14.13 % compared to optimal conditions for OED through optimization guided by SVM. Moreover, since it is difficult for human to understand the prediction or simulation mechanism of SVM, several methods have been applied to determine the impact of each CPPs on CQAs. For example, Partial Dependence Analysis (PDA) and SHapley Additive exPlanations (SHAP) and were conducted to evaluate influence of temperature, moisture content and pH on maturity of compost (Ding et al., 2022); permutation variable importance (PVI) was also employed to determine the influence of acetate, acetate, butyrate on H_2 production process (Hosseinzadeh et al., 2022). There are also a few similar studies on weight analysis (Tang et al., 2021; Zhou et al., 2022).

SVM outperforms ANN on small samples, potentially due to ANN's greater propensity for overfitting in small datasets. Overfitting will lead to poor generalization capability of the model. K-fold cross-validation (KCV) is employed widely in determining the model's hyper-parameters. K-CV can also be used to compare the generalization capability of the models. Although quite a few studies do not pay attention to investigate the generalization ability of models, the performance of models on different datasets should be carefully analysed. In addition, the SVM model's prediction performance heavily depends on kernel function selection (Andrade Cruz et al., 2022). However, the selection of kernel function mainly depends on the predicting experience. The computational cost increases rapidly while data sets become larger because the kernel matrix grows with the data size in quadratic form (Cervantes et al., 2020). Furthermore, SVM is sensitive to missing values and noisy data. Therefore, unlike other ML strategies, data set quality is more important than size when using SVM modeling.

2.3. Fuzzy intelligence (FL)

Unlike the "black box" mode of ML, the operation process of fuzzy theory is closer to human thinking. For example, fuzzy intelligence can describe the complex correlation between multiple growth parameters and the target indicators (Bhola et al., 2017). The fuzzy reasoning process is defined by membership functions and "if-then" rules (Mullai et al., 2022). These techniques convert digital data into text data, including low, moderate, and high (Fig. 1). Fuzzy logic (FL) based algorithms are robust and perform well in small sample datasets.

Due to the FL's distinctive capacity for logical thinking, several different bioprocesses have been effectively used with it. For instance, the prediction of metabolic flow in *Chlorella* sp. was successfully predicted based on an FL model (Bhola et al., 2017). In this study, three fuzzy rules were produced, and the data set is well-fitted by simulation of the TL model. A different study used FL, a GA, to determine the best nutrient composition for producing phycobiliproteins from cyanobacteria (Kumar Saini et al., 2020). Moreover, in FL applications, a set of "If-then" rules is produced using the professional knowledge of users, which is time-consuming and laborious. Therefore, an Adaptive Network-based Fuzzy Inference System (ANFIS), integrated with the learning ability of ANN and the reasoning process of FL, was proposed (Akinade and Oyedele, 2019). To compare ANFIS and ANN, an ANFIS and an ANN were constructed to predict the impacts of four input parameters on water flux in Osmotic Membrane Bioreactor, respectively (Hosseinzadeh et al., 2020). The RMSE in ANFIS models was significantly lower than in ANN models. Similarly, ANFIS and perceptron neural network were also applied in prediction of biochar yield, both methods has satisfied fitting result (Akinade and Oyedele, 2019).

Despite its ability to cope with large data sets produced by complex nonlinear processes, ANFIS still suffers some drawbacks. For example, each output of multiple output models needs an ANFIS model, significantly increasing the modeling process's workload. Besides, the prediction or simulation mechanism of ANFIS are not as easy to understood as kinetic methods because the control rules are generated by ANN.

2.4. Multi-objective model

Most of the mentioned studies are conducted with a key indicator of bioprocess as the objective, such as biomass, yield, or productivity. Nevertheless, many applications still need to optimize a series of objectives in bioprocess in combination to achieve higher economic benefits. For example, biomass and cyanobacterial phycobiliproteins (PBPs) in *Nostoc* sp.CCC-403 was enhanced in a recent study (Saini et al., 2021). PBPs are high-value-added products because of their famous therapeutic properties, while cyanobacteria can be regarded as a source of biofuels. Therefore, biomass and PBPs were optimized simultaneously to achieve economic viability for industrial production. In this study, four input parameters of the experiment were determined by a central composite design principle. Therefore, the final biomass and the yield of PBPs increased by 61 % and 90 %, respectively. Similarly, key CPPs were successfully predicted by temperature, fermentation time, and pH through an ANN (Pappu and Gummadi, 2017). Differently, the biomass and byproduct concentration of previous input data points enables the model to determine the best control trajectories of fermentation parameters. Furthermore, different optimization strategies could be compared to find the optimal conditions. For instance, ML models were employed in predicting bio-oil yield and contents of oxygen and nitrogen in bio-oil (Zhang et al., 2021). The optimal parameters were determined by using both forward and reverse optimization. Moreover, due to the complex and changeable conditions in bioprocesses, the detection process of CPPs would generates data noise. In order to cope with this problem, RBF-ANN was employed in water treatment to enhance nitrogen and phosphorus removal though optimization of CPPs (Li et al., 2022).

Nevertheless, some potential problems may limit the further development of multi-objective models. At present, most of the multi-objective models are based on shallow ML, which is not suitable for dealing with complex nonlinear problems. In addition, multi-objective optimization is a major application of multi-objective model construction work. However, most relevant researches focus on the construction of model, while the methods and strategies of multi-objective optimization are always not systematically mentioned. In order to further promote the applications of multi-objective models, the above problems may need to be solved.

2.5. Genetic algorithm (GA)

GA is an optimization method to optimize hyperparameters and process parameters by simulating the evolutionary process (Renata et al., 2015). As illustrated in Fig. 1, the next generation is decided by mutation, crossover, and selection process. The selection pressure will be simulated by choosing individuals with different possibilities. The one with the highest fitness value will be selected through constant iteration and evolution.

Several solutions could be simultaneously evaluated by using this GA when dealing with complicated combinatorial problems. Additionally, GA may be easily extended to work with other models. For example, Zhang et al. used SVM and GA to determine the optimum lipid production conditions from cellulosic ethanol wastewater (Zhang et al., 2020). Moreover, GA was used for the optimization of poly(3HB-*co*-3HV) production through the simulation of ANN and RSM (Zafar et al., 2012). And BP-ANN and GA were also used to determine the best conditions for biogas production. In addition, GA was interacted with ANN to determine CPPs in enzyme process and composting process production (Dixit et al., 2022; Shi et al., 2022). Furthermore, deep learning method was combined with GA to optimize performance of microfluidic fuel cell (Nguyen et al., 2022). Except the optimization of CPPs of bioprocess, GA could also be employed to optimize hyperparameters of data-driven models. For example, six different metaheuristic algorithms was exploited to determine hyperparameters of ANN to predict the biochar production (Khan et al., 2022). More than this, GA could also be applied in optimizing the multi-objective model. For instance, GA was used to maximize the production of biomass and bioactive phycobiliproteins in *Nostoc* sp. CCC-403 (Saini et al., 2021). However, unlike the optimization result of a single objective model, multi-objective optimization would provide multiple solutions for researchers. Each solution obtained will attempt to increase the sum of multiple objectives as much as possible. Therefore, how to balance multiple objectives in multi-objective optimization is another problem needed to considered.

Although GA has been widely applied in the optimization process, the premature convergence of GA is still a problem that needs to take care of. The mutation process, crossover and selection, and population scale may result in premature convergence. Besides, the computation cost of GA is relatively high, so the time consumption should be mainly considered when dealing with complex problems. Furthermore, implementation of GA requires programming skills, because its process involves adjustment of multiple parameters.

2.6. Particle swarm optimization (PSO)

PSO is a population-based optimization technique activated by the intelligence of animal population behaviors. Similar to GA, the PSO algorithm generates a set of particles randomly, and optimization procedures are searched by updating generations. Therefore, each particle could be considered as a set of parameters. As shown in Fig. 1, a particle's movement would be affected in three directions: inertia, the global best solution, and the individually best solution. Therefore, the solution space can be effectively explored to find the optimal solution.

PSO is also quite extensible and simple to integrate with other models like the GA algorithm. For example, the radial Basis Function Neural Network was coupled with PSO to model and optimize microbial hyaluronic acid production (Liu et al., 2009). Furthermore, Mahmoodi-Babolan et al. demonstrate the coupling of the ANN technique with the PSO and compare the performance of ANN-PSO with RSM in the prediction of the adsorption of methylene blue (Mahmoodi-Babolan et al., 2019). In addition, PSO could be applied in forward and reverse optimization strategies to determine the best optimization results. For example, a multi-objective model (bio-oil yield and the contents of oxygen and nitrogen in bio-oil as outputs) was applied in forward and reverse optimization to determine the best optimal conditions of HTL for different algae (Yang et al., 2020). Although PSO has been widely used in optimization problems, there are still some limitations. For example, like the GA algorithm, PSO has the problem of premature convergence in the optimization process, which would occur when the best particles are far from the global optimum.

Both AI and conventional statistical methods have been extensively employed to model and optimize bioprocess. However, each of above mentioned mathematical model has the specific application prerequisites (Table 1). Recently, quite a few studies attempt to find the optimal algorithm for the specific cases in their research (Jacob and Banerjee, 2016; Singhal et al., 2018; Vinoth Arul Raj et al., 2021). In general, AI technologies perform better in complex nonlinear problems, such as image recognition, multi-objective optimization and reinforcement learning problems, while some limitations of conventional statistical methods result in low performance in nonlinear problems. For example, OED treats each factor as the same level; RSM requires that the best experimental condition should be contained in experimental dataset.

3. AI-assisted detection technology

3.1. Rapid detection based on spectroscopy and image

Recently, spectroscopy has been used for rapid detection in bioprocess because of its convenience and accuracy. As a classical spectral analysis technology, near-infrared reflectance spectroscopy (NIRS) has been widely employed for qualitative analysis and quantitative detection according to the information of overtones and combination bands of the fundamental functional groups (Baqueta et al., 2021; Goi et al., 2020; Liu et al., 2021). The above studies show that there is a complex nonlinear relationship between the content of compounds and their spectral data. Recently, in order to rapidly estimate the biochemical

Table 1
Advantages, limitations of the most frequently AI (artificial intelligence) and CSM (conventional statistical method) used in bioprocess.

Analysis Method	Models	Advantages	Limitations
AI	ANN	• High prediction accuracy, strong adaptive and fault tolerance.	• "Black box" nature, hyperparameters optimization is complex, prone to overfitting.
	SVM	• Excellent performance in small dataset, Strong generalization ability, good at high-dimensional problems.	• Poor efficiency in large dataset, difficulty in determining kernel function, Conventional SVM only supports binary classification problems.
	FL	• Distinctive capacity for logical thinking, robust in small sample datasets.	• The set of "If-then" rules is produced based on professional knowledge of users.
	GA	• It is extensible and can be integrated with other algorithms, belongs to heuristic algorithm and the optimization process is simple.	• Require programming skills, high computation cost, easy to fall into local optimum.
	PSO	• Easy to understand and implement, fast search and high efficiency.	• Low performance in discrete optimization problems and local optimization, easy to fall into local optimum.
CSM	OED	• Fewer experiments, good representativeness and the experimental factors are not strictly limited.	• Only suitable for experiments with fewer gradients, each factor is treated as same level.
	RSM	• Simple calculation process, the prediction model is continuous. Random error is considered.	• Best experimental conditions need to be covered in experimental points, low performance in complex nonlinear problems.

AI: Artificial intelligence; CSM: Conventional statistical method; ANN: Artificial neural network; SVM: Support vector machine; FL: Fuzzy logic; GA: Genetic algorithm; PSO: Particle swarm optimization; OED: Orthogonal experimental design; RSM: Response surface methodology.

methane potential (BMP) of feedstocks in anaerobic digestion, a multivariate regression model was established between BMP and NIRS. The result show that predicted accuracy of NIRS model based on characteristic wavelengths outperformed all regression models based on the physicochemical indexes (Yang et al., 2021). Furthermore, ANN networks based on UV–vis were also applied to detect metabolite concentrations, such as glutamine, glutamate, glucose, and viable cell concentrations (Takahashi et al., 2015).

With the rapid development of hardware, such as digital photography instruments, AI has been widely used in rapid detection approaches, such as image analysis based on machine versions (Mahlein, 2016). However, conventional image analysis needs to be divided into multistep methods according to an ML-based vision system, including feature extraction, classifier selection, and classification or regression. Therefore, these methods are not suitable for rapid detection. Recently, deep learning methods have been gradually employed in machine vision. For example, a bioreactor foam sensing approach was reported (Austerjost et al., 2021). The machine vision system based on CNN was connected to a control system of chemical foam. Therefore, image recognition technology can eliminate foam effectively. In another study, a deep learning method was employed to predict agricultural waste compost maturity (Xue et al., 2019). By studying 30,000 pictures, the accuracy of the proposed method was better than 99 % in all test sets (Fig. 2B).

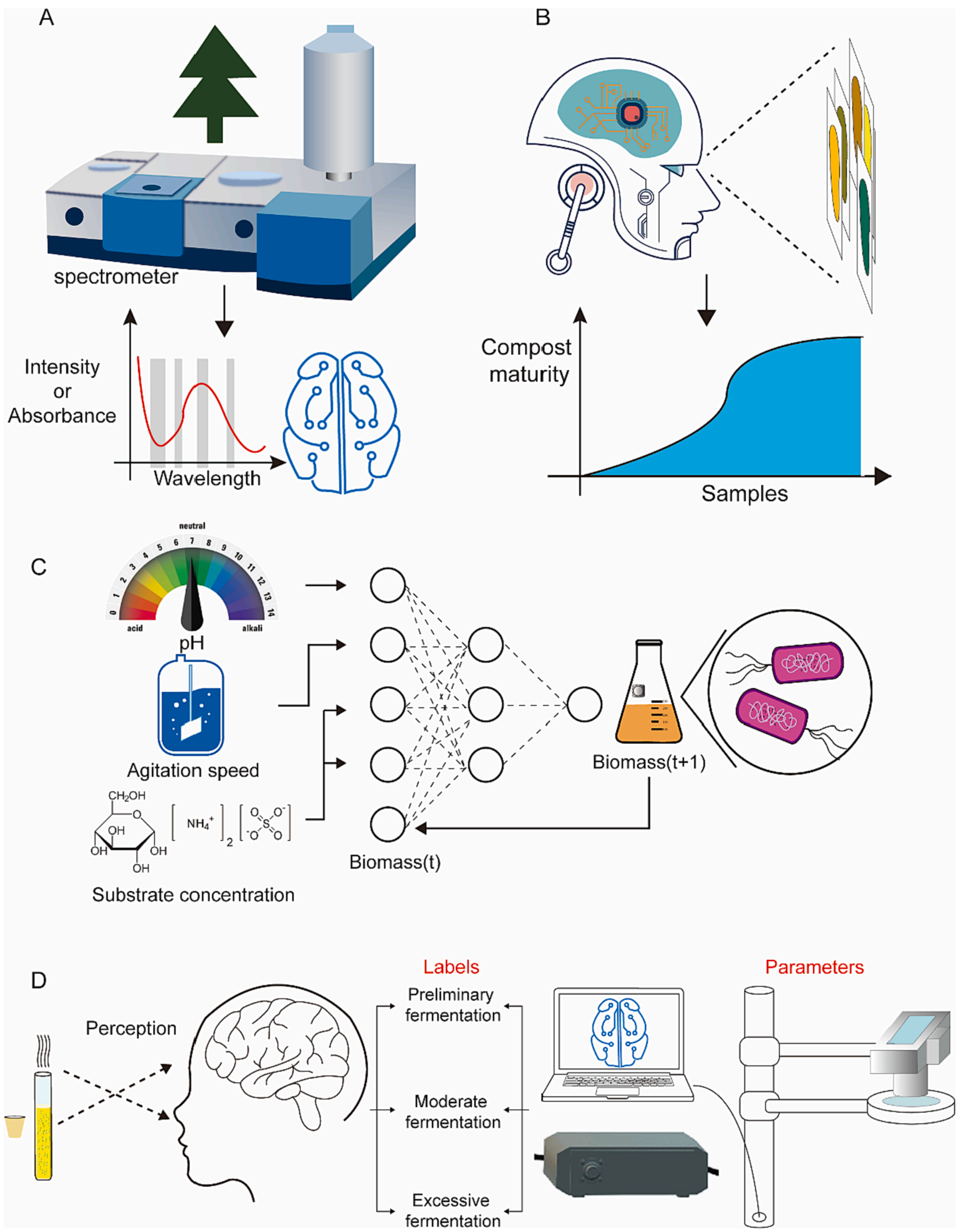


Fig. 2. Artificial intelligence (AI)-guided detection technology. (A) Rapid detection technology based on spectral data. (B) Rapid detection technology based on image analysis. (C) Single variable real-time monitoring technology based on soft sensors. (D) Multivariable real-time monitoring technology based on soft sensors.

Rapid detection technology of CPPs is an essential part in bioprocess control and optimization technology. In recent years, rapid detection technology has huge advancement due to the emergence of image recognition and AI technologies. However, there are still some limitations hindering its further development. For example, rapid detection methods to determine the content of metabolites by spectrum are still rare, which may contribute to the complexity of method construction; it is an impressive idea to apply image recognition technology in fermentation process control. However, the models are easily affected by environment. Therefore, applications of rapid detection methods would be further promoted by automatic modeling framework and equipment integrating machine vision and control system.

3.2. Real-time monitoring based on soft sensor

Real-time monitoring of the critical indicators is essential to control the bioprocess in the optimal direction (Wang et al., 2020). Nevertheless, many parameters cannot be monitored online because of the complexity of the bioprocess environment or the lack of rapid detection equipment and methods. For example, the alkalinity of the fermentation process can be detected online with additional automatic equipment and pipelines, which would increase the cost (Lahav and Morgan, 2004). Besides, because of the harsh environment of most bioprocesses, hardware-based sensors often have an accumulation of detection error (Sharma et al., 2010). Typical soft sensors are mathematical models that allow the prediction of target values via continuously measured secondary variables, such flow rates, dissolved oxygen. Recently, AI has been employed as core algorithm to exploit data-driven soft sensors. More critically, soft sensors based on AI are feasible solutions to achieve low cost and high accuracy online monitoring (Moncks et al., 2022). With the fast development of software and hardware, soft sensors have been gradually employed in various bioprocesses.

3.2.1. Single variable estimation

Single variable estimation is to estimate a critical process parameter through a soft sensor. In the upstream bioprocess, biomass concentration has been considered one of the most essential attributes besides the product titer (Brunner et al., 2020b). A few techniques were used for the online monitoring of biomass concentration, such as flow cytometry, turbidimetry, and spectroscopy. However, these methods are unreliable because the morphological change of cells would happen during the bioprocesses (Luttmann et al., 2012). Besides, changes in process characteristics and operating conditions would also affect detection (Kano and Fujiwara, 2013). Therefore, a few studies employed data-driven soft sensors to achieve stable online monitoring of biomass to cope with these problems. For example, an adaptive soft sensor was developed for online monitoring of the *Pichia pastoris* fed-batch process (Fig. 2C), which could be divided into several phases based on induction conditions (Brunner et al., 2020b). Moreover, a Multiphase ANN (MANN) was applied to predict the biomass concentration where insoluble substrates existed (Murugan and Natarajan, 2019). More specifically, the soft sensor has three nonlinear auto regressive models to capture the complete dynamics of lag, log and stationary phases of the microbe using online measurements such as pH, substrate concentration and agitation speed.

In addition to biomass concentration, the physiological state of cells and product synthesis are strongly affected by specific growth rate (SGR). Therefore, a few studies also focused on the online monitoring of SGR. A method for estimating SGR according to online measurement of oxygen uptake rate has been developed (Survyla et al., 2021). Cells are living organisms that breathe and consume food, so respiratory data can express the state of the cell culture in the bioreactor. Thus, the relationship between the oxygen uptake rate and SGR can be modeled using Luedeking/Piret-type relationships. It is worth mentioning that this method does not require a particular bioprocess model. Moreover, this method was tested on different fermentation scales with different hosts. However, changes in experimental conditions, such as agitation rates, would increase the noise in the oxygen concentration signal. Therefore, another strategy was further exploited according to the second law of thermodynamics (Allampalli et al., 2022). This research indicated a correlation between the energy used by metabolism and biomass synthesis. Therefore, quantifying metabolic heat rate could be applied to predict SGR. Furthermore, this strategy was integrated with the PID control system to maintain SGR in a narrow range.

3.2.2. Multivariable estimation

Multi-objective estimation is to estimate a few critical process parameters through a soft sensor. Because of the high complexity of the bioprocess procedure, both environment and microorganisms changed simultaneously. Therefore, a single indicator is primarily insufficient to evaluate the bioprocess comprehensively (Jandric et al., 2021). Multivariable estimation based on soft sensors has been developed to describe the bioprocess state accurately. At present, quite a few studies attempt to enhance their model performance by increasing high-accuracy sensors and expensive detection techniques. However, most of these strategies could not be applied in industrial production (Jin et al., 2020). Therefore, cost and estimation accuracy should be considered while developing a multivariable estimation method. For example, a strategy that integrates low-cost micro-NIRS with the machine vision system was employed to achieve online rapid detection of black tea fermentation (Jin et al., 2021). It is interesting here that human senses entirely determine the quality of black tea sample (Fig. 2D). Moreover, a SVM model integrated with PCA was exploited to build correlation between the manual evaluation grade and process parameters. Imitating human perception system through soft sensors has a promising application prospect, especially in food and pharmaceutical manufacturing industry. In addition, high time scale variability of bioprocesses would also result in the uncertainty of fermentation phases, to overcome the phenomenon of high time scale variability caused by variation of substrate characteristics and the "scale-up effect," a noninvasive soft sensor was exploited, which could be employed for online monitoring multiple concentrations according to particle filter estimator (Sinner et al., 2021). Moreover, to realize real-time monitoring of complex systems, deep learning was applied in learning nonlinear relationship to predict CPPs.

Despite the advances in soft sensors, this part still suffers from some limitations. For example, the use conditions of most soft sensors are strictly limited, which impedes its further applications (Brunner et al., 2020a). To cope with this problem, soft sensors could be integrated with the automatic calibration program to adapt to different detection environments. In addition to combining image recognition and shallow ML, electronic tongue and nose can be integrated with deep learning to achieve more accurate imitation of human perception system.

4. AI-guided bioprocess control

Effective control strategies can improve the yield and productivity of bioprocess effectively (Rathore et al., 2021). Unlike the chemical process, which can operate stably with long operational periods in the reactor (del Rio-Chanona et al., 2019), bioprocess have high time scale variability of different stages and phases because of the different inoculation size, seed age, or culture conditions. Therefore, it is challenging to determine current bioprocess phases precisely according to controllable CPPs and implement effective control strategies (Nikzad-Langerodi et al., 2017). Nevertheless, developing simulation technologies to predict the optimal control strategy based on CPPs is feasible. Kinetic models were first introduced in bioprocess to determine optimal control strategy (Adesanya et al., 2014; del Rio-Chanona et al., 2017). Despite the broad applications of kinetic methods, these methods still have some drawbacks. The construction of kinetic models relies on an accurate understanding of the mechanism. More critical, kinetic models are not good at dealing with nonlinear and sophisticated problems. Dynamic time warping (DTW) technology was applied to cope with these problems and to optimize the control strategy by overcoming the time scale variability (Gollmer and Posten, 1996).

Additionally, phase changes of bioprocess could be detected by a sliding window-based approach (Maiti et al., 2009). However, DTW is only limited to applications with a few CCPs because noncorrelation parameters cannot exist in high-dimensional space simultaneously. A sliding window-based approach could cope with high-dimensional problems, but the enormous computational cost associated with optimizing hyperparameters and model refitting might limit its development. Nevertheless, it is an alternative approach to employing data-driven methods and AI technologies. Furthermore, these methods could determine control strategy without extensive knowledge of the bioprocess systems (Rathore et al., 2021). A summary of AI technologies applied in bioprocess control is shown in Table 2.

4.1. Neural network-based control

Due to its powerful data pattern identification ability, ANN has already been applied in bioprocess modeling. At present, quite a few works are limited to enhancing the ability to simulate bioprocess. However, it is also essential to control and optimize bioprocess in a real-time approach through dynamic simulation of models (del Rio-Chanona et al., 2019). A few ANN-based control strategies were conducted based on time as one of the inputs (Peng et al., 2013) to achieve this objective. For instance, after the modeling procedure, an optimization process is employed to optimize process parameters at a different time to obtain the optimal control trajectories. To further exploit the simulation capability of time-series data, RNN was employed to determine the best optimization procedure for the production of microalgal lutein (del Rio-Chanona et al., 2019). In this research, RNN and kinetic model both significantly improved lutein intracellular content. However, the RNN model performed best in total yield optimization, indicating that ML is more suitable for dealing with nonlinear problems.

Table 2
Artificial intelligence (AI) technologies applied in bioprocess control strategies.

Control strategy	Technique	Task description	Reference
Open-loop control	RNN	• RNNs was applied to determine the optimal control trajectories	(Peng et al., 2013)
Feedback control strategy	FL model	• Glucose consumed was regarded as an input, and the feeding operations controlled by the FL model were determined timely.	(Tai et al., 2016)
Open-loop control	Positive and negative fuzzy rules	• Optimal control strategies were determined by positive and negative fuzzy rules.	(Birle et al., 2016)
Feedback control strategy	FL integrated with soft sensors	• The biomass detected by soft sensors was coupled with classical kinetic model to implement dynamic control strategies.	(Escalante-Sánchez et al., 2018)
Open-loop control	RNN	• Total production of microalgal lutein was significantly increased under control of a RNN model.	(del Rio-Chanona et al., 2019)
Open-loop control	ANN in combination with kinetic model	• Visualization of continuous future trajectories of bioprocess was achieved by refitting the kinetic model using discrete data points predicted by data-driven methods.	(Zhang et al., 2019)
Closed-loop strategy	Reinforcement learning in combination with DNN	• Reinforcement learning method was implemented in the co-culture fermentation process to learn how to maintain specific population levels for maximizing production.	(Treloar et al., 2020)
Closed-loop strategy	DNN	• Self-learning could be achieved in the online learning phase within the closed-loop strategy.	(Natarajan et al., 2021)

RNN: Recurrent neural network; FL: Fuzzy logic; ANN: Artificial neural network; co-culture: Control of multi-species communities; DNN: Deep neural network.

Moreover, a hybrid modeling framework based on a kinetic model and ANN was proposed for online simulation and bioprocess optimization (Zhang et al., 2019). Interestingly, the visualization of continuous future trajectories of bioprocess in this framework was achieved by refitting the kinetic model using discrete data points predicted by data-driven methods. Nevertheless, these methods are open-loop control strategies that cannot take adequate measures when random disturbances occur (Rathore et al., 2021). Therefore, a deep neural network was employed to construct an online feedback control approach to cope with this problem (Natarajan et al., 2021). Furthermore, as a closed-loop strategy, self-learning could be achieved in the online learning phase within the closed loop. Consequently, the control trajectory, which could be used to attain the product's maximum yield, was demonstrated.

In addition, multi-species microbial communities could be employed for biomanufacturing within the control strategy guided by ANN (Treloar et al., 2020). Recently, a reinforcement learning framework combined with neural networks was applied to control co-cultures system within bioreactors. The co-cultures system consisted of two auxotrophs dependent on two different nutrients, with competition over a common carbon source. Control actions were achieved by adjusting the concentration of auxotrophic nutrients flowing into the reactor. Despite the intricate details of the interactions in co-culture systems, the reinforcement learning method was directly implemented in the co-culture fermentation process to learn how to maintain specific population levels for maximizing production, which exhibits that reinforcement learning strategy will have a fundamental role to play in multi-species microbial communities co-cultures.

4.2. Fuzzy logic-based control

ML-based control strategies have been gradually applied because of their powerful and precise simulation ability. However, ambiguity often exists in quite a few bioprocesses where critical indexes could not be obtained precisely (da Ros et al., 2013), such as enzymatic hydrolysis (Tai et al., 2016). In contrast to ML, artificial experience and reasoning are the basis for fuzzy intelligence. Fuzzy logic-based control strategies would generate approximate approaches rather than precise control schemes or trajectories, which have been applied in bioprocess to improve their performance. In a few studies, proper feeding operations were determined using FL-based control strategies (Fig. 3) (Rathore et al., 2021). For instance, an FL feedback control system was exploited to control the feeding of enzymatic hydrolysis of lignocellulosic biomass (Tai et al., 2016). Glucose consumed in real-time during fermentation was considered an input, and the feeding operations controlled by the FL feedback control system were determined timely.

Moreover, a soft-sensor-based FL control system was proposed (Escalante-Sánchez et al., 2018). In this work, a soft sensor capable of predicting biomass was exploited. The biomass detected by soft sensors was coupled with glucose content based on the classical cell growth kinetic model to implement real-time control strategies. In addition, the concept of fuzzy control was further explored. The control concept adopted in most studies mentioned above can be mainly concluded as positive fuzzy rules, expressed as "if A, then take measures." However, negative fuzzy rules could also be applied to generate FL models. For

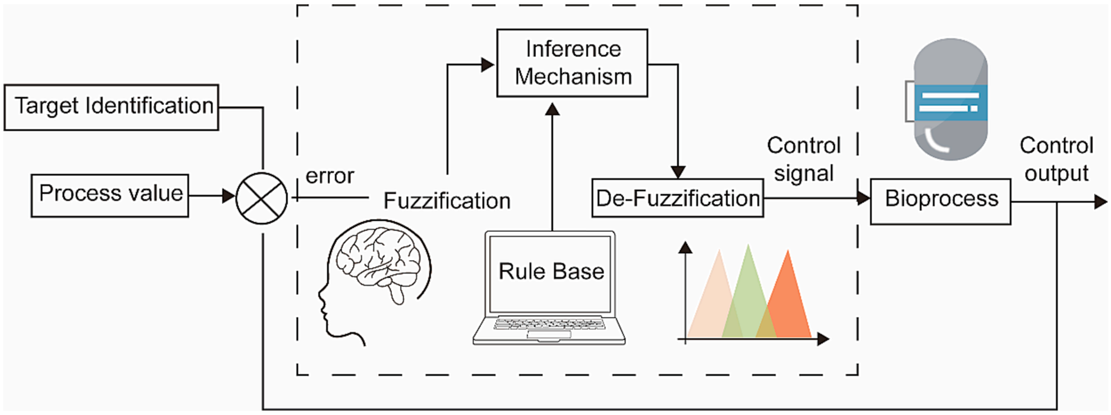


Fig. 3. Fuzzy logic (FL)-based control system.

example, a combination of positive and negative rules was employed to build integrated FL control strategies for yeast propagation (Birle et al., 2016). Under the guidance of a reference trajectory, RMSE generated by the integrated FL models could be reduced by 62.8 % compared to conventional models. Furthermore, a general control strategy based on FL models has been exploited for feed phase identification (Nikzad-Langerodi et al., 2017). Multivariate classifiers were trained to distinguish between different feed phases by combining the classifiers with a oneway switch.

5. Challenges and perspectives

This critical review presents the applicability of AI technologies, which can be used as a guidance for model selection in the bioprocess control and optimization. However, despite the advantages that AI technologies provide, there are some potential challenges and research gaps which may limit its further application in bioprocess.

Potential challenges and solutions in bioprocess modeling and optimization are listed as follows. First, how to obtain the available data effectively is an important problem. Nowadays, a growing number of bioprocesses studies have been conducted, however, it is quite difficult to integrate their data for further analysis because experimental conditions in their researches vary widely. To cope with this problem, more detailed message of experimental data should be provided. Moreover, the database for collecting and managing bioprocess data can be established according to the detailed message. Second, there are quite a few algorithms that require programming skills because some parameters need to be adjusted, such as GA and DNN. Therefore, the graphical user interface of AI toolbox could be exploited to further promote AI applications in bioprocess. Third, most ML algorithms are criticized for their "black box" characteristics. It's unfriendly for researchers to comprehend the simulation mechanisms of these well-performed models, which limits the further development. In a few studies, simulations by changing inputs could be employed to glimpse the inside of the "black box" model (Moreira et al., 2021). Moreover, integrating ML models with conventional mechanism methods, such as kinetic models, could strengthen the interpretability of those black box model (Andrade Cruz et al., 2022). Additionally, compared with single objective optimization, multi-objective optimization is a more complicated process. However, quite a few studies focus on the construction of model, while the methods and strategies of multi-objective optimization are always not systematically mentioned. Therefore, strategies and methods of multi-objective optimization should be studied in future.

In addition, AI-assisted detection technology also faces some challenges, for example, qualitative and quantitative detection methods based on spectroscopy are still rare; image recognition technology and soft sensors are easily affected by experimental conditions; the prediction results of some CPPs are not accurate enough. To cope with these problems, automatic modeling framework could be exploited to accelerate the development of detection methods. Soft sensors and image recognition technology could be integrated with automatic calibration program to enhance their stability under different conditions. Furthermore, advanced equipment such as electronic nose can be combined with deep learning to achieve more accurate predictions.

Moreover, the bioprocess control strategy based on AI technologies has not been widely applied in industry level (Rathore et al., 2021), which may be attribute to some relevant studies still in their infancy. For example, reinforcement learning applied in co-cultures systems has been limited in maintaining the balance between two species. Besides, lack of corresponding hardware is also a limiting factor. Therefore, the relevant studies should be more closely combined with industrial production, and jointly promote the application of AI in industrial production through the development of software and hardware.

Some emerging technologies may further promote the bioprocess control and optimization. There are two different scales of complexity in biological processes in bioreactors. The first is the cell scale's complex metabolic flux of cells' response to the external environment and the second is the complexity of turbulent flow in the reactor, including the complex transfer characteristics of mixing, mass transfer, heat transfer, etc. Unfortunately, due to the limitations of online monitoring technology, CPPs have far been predicated on the second level and nothing is known about the intracellular metabolic status in real-time. Therefore, applying state-of-art technologies like real-time omics with AI to industrial production may be able to control and optimize bioprocess from a cellular perspective. Additionally, the model built entirely through AI technology cannot enable researchers to understand the mechanism due to the technical characteristics of AI. However, the mechanism model can essentially characterize the laws of biological processes. Therefore, combining AI technology and mechanism models like genome-scale metabolic models (GEMs) may bring new insights into this field. Moreover, computational fluid dynamics (CFD) (Kannangara et al., 2018), a combination of mathematics, fluid mechanics, and computer technology, could be applied to the generation of fluid information in the fermentation process. Coupling the flow field information obtained by CFD models with the metabolic information provided by GEMs can help to comprehend how turbulence and environmental gradients affect cell performance in bioprocess, which will guide scaling-up production ultimately (Fig. 4). Furthermore, bioprocess optimization and control powered with AI will be further upgraded by adopting automated processes that use robots and high-throughput system.

6. Conclusions

This review covered the applicability of AI technologies (ANN, SVM, FL, GA and PSO) in bioprocess control and optimization, as well as their advantages and limitations. The technologies of modeling and

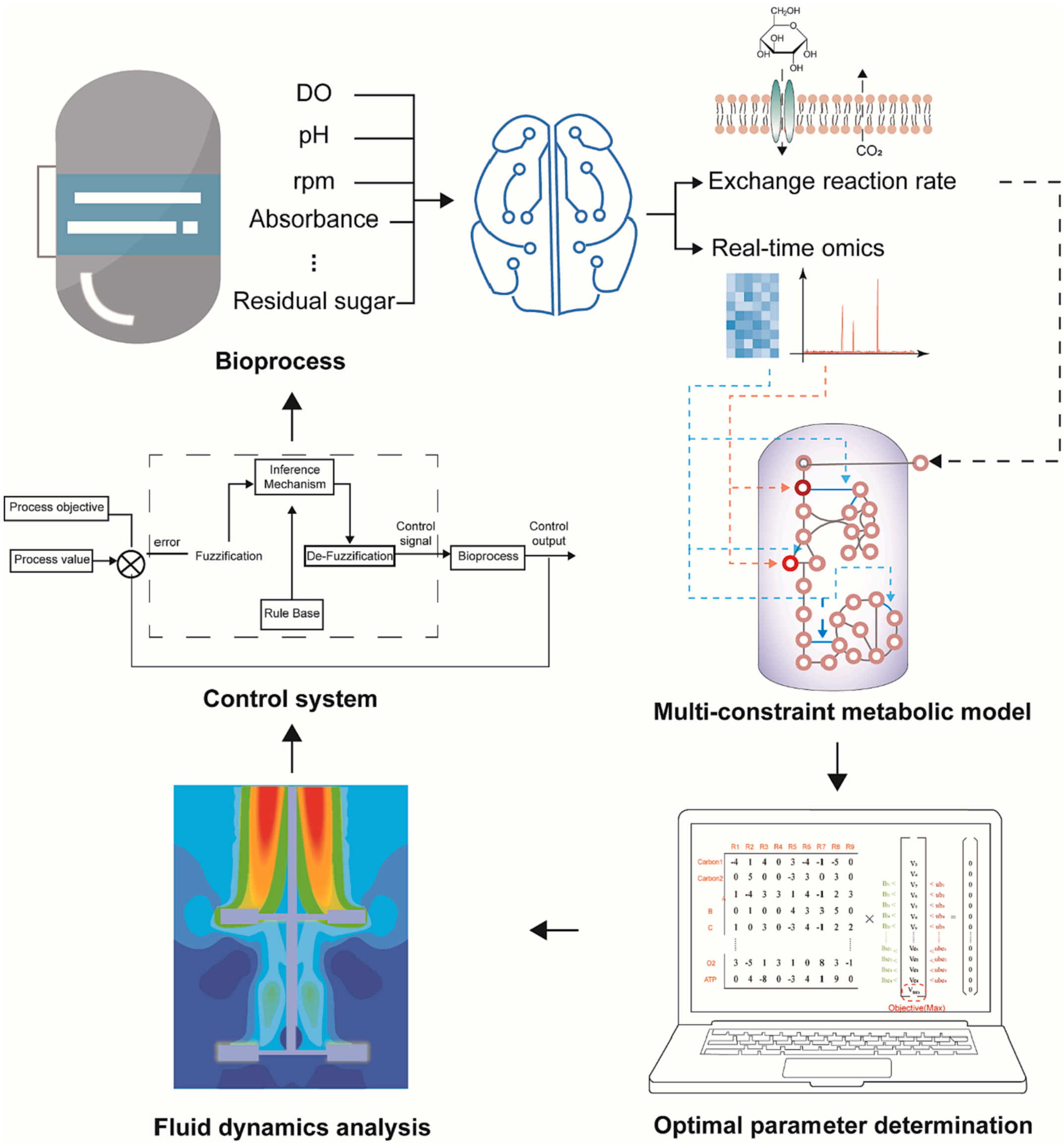


Fig. 4. Prospects for artificial intelligence (AI)-based bioprocess control and optimization technology.

optimizing the whole bioprocess have matured. In order to further enhance bioprocess, machine vison, spectroscopy and soft sensors have been gradually employed to achieve more accurate monitoring of bioprocess. Meanwhile, smart control strategies have been proposed based on advanced monitoring and modeling technologies. Moreover, by integrating GEMs and CFD, the applications of AI in bioprocess will be further promoted.

CRediT authorship contribution statement

Yang Cheng: Conceptualization, Investigation, Writing – original draft. **Xinyu Bi:** Supervision, Investigation, Writing – original draft. **Yameng Xu:** Supervision, Writing – review & editing. **Yanfeng Liu:** Supervision, Project administration, Writing – review & editing. **Jianghua Li:** Writing – review & editing. **Guocheng Du:** Writing – review & editing. **Xueqin Lv:** Supervision, Project administration, Writing – review & editing. **Long Liu:** Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was financially supported by the Key Research and Development Program of China (2020YFA0908300), the National Natural Science Foundation of China (32070085, 32021005) and the Fundamental Research Funds for the Central Universities (JUSRP222007, JUSRP62200, JUSRP121010).

References

Adesanya, V.O., Davey, M.P., Scott, S.A., Smith, A.G., 2014. Kinetic modelling of growth and storage molecule production in microalgae under mixotrophic and autotrophic

conditions. Bioresour. Technol. 157, 293–304. https://doi.org/10.1016/j.biortech.2014.01.032.
Akinade, O.O., Oyedele, L.O., 2019. Integrating construction supply chains within a circular economy: An ANFIS-based waste analytics system (A-WAS). J. Clean. Prod. 229, 863–873. https://doi.org/10.1016/j.jclepro.2019.04.232.
Alejo, L., Atkinson, J., Guzmán-Fierro, V., Roeckel, M., 2018. Effluent composition prediction of a two-stage anaerobic digestion process: machine learning and stoichiometry techniques. Environ. Sci. Pollut. Res. Int. 25, 21149–21163. https://doi.org/10.1007/s11356-018-2224-7.
Allampalli, P., Rathinavelu, S., Mohan, N., Sivaprakasam, S., 2022. Deployment of metabolic heat rate based soft sensor for estimation and control of specific growth rate in glycoengineered Pichia pastoris for human interferon alpha 2b production. J. Biotechnol. 359, 194–206. https://doi.org/10.1016/j.jbiotec.2022.10.006.
Alrashed, A.A.A.A., Karimipour, A., Bagherzadeh, S.A., Safaei, M.R., Afrand, M., 2018. Electro- and thermophysical properties of water-based nanofluids containing copper ferrite nanoparticles coated with silica: Experimental data, modeling through enhanced ANN and curve fitting. Int. J. Heat Mass Transf. 127, 925–935. https://doi.org/10.1016/j.ijheatmasstransfer.2018.07.123.
Andrade Cruz, I., Chuenchart, W., Long, F., Surendra, K.C., Renata Santos Andrade, L., Bilal, M., Liu, H., Tavares Figueiredo, R., Khanal, S.K., Fernando Romanholo Ferreira, L., 2022. Application of machine learning in anaerobic digestion: Perspectives and challenges. Bioresour. Technol. 345, 126433.
Ascher, S., Wang, X., Watson, I., Sloan, W., You, S., 2022. Interpretable machine learning to model biomass and waste gasification. Bioresour. Technol. 364, 128062 https://doi.org/10.1016/j.biortech.2022.128062.
Austerjost, J., Söldner, R., Edlund, C., Trygg, J., Pollard, D., Sjögren, R., 2021. A machine vision approach for Bioreactor Foam Sensing. SLAS. Technol. 26, 408–414. https://doi.org/10.1177/24726303211008861.
Baqueta, M.R., Coqueiro, A., Março, P.H., Valderrama, P., 2021. Multivariate classification for the direct determination of cup profile in coffee blends via handheld near-infrared spectroscopy. Talanta 222, 121526. https://doi.org/10.1016/j.talanta.2020.121526.
Bhola, V., Swalaha, F.M., Nasr, M., Bux, F., 2017. Fuzzy intelligence for investigating the correlation between growth performance and metabolic yields of a *Chlorella* sp. exposed to various flue gas schemes. Bioresour. Technol. 243, 1078–1086. https://doi.org/10.1016/j.biortech.2017.07.031.
Birle, S., Hussein, M.A., Becker, T., 2016. Incorporation of negative rules and evolution of a fuzzy controller for yeast fermentation process. Bioprocess Biosyst. Eng. 39, 1225–1233. https://doi.org/10.1007/s00449-016-1601-1.
Brunner, V., Klöckner, L., Kerpes, R., Geier, D.U., Becker, T., 2020a. Online sensor validation in sensor networks for bioprocess monitoring using swarm intelligence. Anal. Bioanal. Chem. 412, 2165–2175. https://doi.org/10.1007/s00216-019-01927-7.
Brunner, V., Siegl, M., Geier, D., Becker, T., 2020b. Biomass soft sensor for a Pichia pastoris fed-batch process based on phase detection and hybrid modeling. Biotechnol. Bioeng. 117, 2749–2759. https://doi.org/10.1002/bit.27454.
Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 408, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018.
da Ros, S., Colusso, G., Weschenfelder, T.A., de Marsillac Terra, L., de Castilhos, F., Corazza, M.L., Schwaab, M., 2013. A comparison among stochastic optimization algorithms for parameter estimation of biochemical kinetic models. Appl. Soft Comput. 13, 2205–2214. https://doi.org/10.1016/j.asoc.2013.01.019.
del Rio-Chanona, E.A., Ahmed, N.R., Zhang, D., Lu, Y., Jing, K., 2017. Kinetic modeling and process analysis for *Desmodesmus* sp. lutein photo-production. AIChE J 63 (7), 2546–2554.
del Rio-Chanona, E.A., Ahmed, N.R., Wagner, J., Lu, Y., Zhang, D., Jing, K., 2019. Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses. Biotechnol. Bioeng. 116, 2971–2982. https://doi.org/10.1002/bit.27131.
Ding, S., Huang, W., Xu, W., Wu, Y., Zhao, Y., Fang, P., Hu, B., Lou, L., 2022. Improving kitchen waste composting maturity by optimizing the processing parameters based on machine learning model. Bioresour. Technol. 360, 127606 https://doi.org/10.1016/j.biortech.2022.127606.
Dixit, M., Kumar Gupta, G., Yadav, M., Chhabra, D., Kumar Kapoor, R., Pathak, P., Bhardwaj, N.K., Shukla, P., 2022. Improved deinking and biobleaching efficiency of enzyme consortium from *Thermomyces lanuginosus* VAPS25 using genetic Algorithm-Artificial neural network based tools. Bioresour. Technol. 349, 126846 https://doi.org/10.1016/j.biortech.2022.126846.
Dong, C., Chen, J., 2019. Optimization of process parameters for anaerobic fermentation of corn stalk based on least squares support vector machine. Bioresour. Technol. 271, 174–181. https://doi.org/10.1016/j.biortech.2018.09.085.
Escalante-Sánchez, A., Barrera-Cortés, J., Poggi-Varaldo, H.M., Ponce-Noyola, T., Baruch, I.S., 2018. A soft sensor based on online biomass measurements for the glucose estimation and control of fed-batch cultures of Bacillus thuringiensis. Bioprocess Biosyst. Eng. 41, 1471–1484. https://doi.org/10.1007/s00449-018-1975-3.
Feng, X., Chen, Y., Lv, J., Han, S., Tu, R., Zhou, X., Jin, W., Ren, N., 2020. Enhanced lipid production by *Chlorella pyrenoidosa* through magnetic field pretreatment of wastewater and treatment of microalgae-wastewater culture solution: Magnetic field treatment modes and conditions. Bioresour. Technol. 306, 123102 https://doi.org/10.1016/j.biortech.2020.123102.
Ge, X., Tolosa, L., Rao, G., 2004. Dual-labeled glucose binding protein for ratiometric measurements of glucose. Anal. Chem. 76, 1403–1410. https://doi.org/10.1021/ac035063p.
Goi, A., Manuelian, C.L., Righi, F., de Marchi, M., 2020. At-line prediction of gelatinized starch and fiber fractions in extruded dry dog food using different near-infrared spectroscopy technologies. Animals 10, 862. https://doi.org/10.3390/ani10050862.
Gollmer, K., Posten, C., 1996. Supervision of bioprocesses using a dynamic time warping algorithm. Control Eng. Pract. 4, 1287–1295. https://doi.org/10.1016/0967-0661(96)00136-0.
Guo, H., Wu, S., Tian, Y., Zhang, J., Liu, H., 2021. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. Bioresour. Technol. 319, 124114 https://doi.org/10.1016/j.biortech.2020.124114.
Hamet, P., Tremblay, J., 2017. Artificial intelligence in medicine. Metabolism 69, S36–S40. https://doi.org/10.1016/j.metabol.2017.01.011.
Hosseinzadeh, A., Zhou, J.L., Altaee, A., Baziar, M., Li, X., 2020. Modeling water flux in osmotic membrane bioreactor by adaptive network-based fuzzy inference system and artificial neural network. Bioresour. Technol. 310, 123391 https://doi.org/10.1016/j.biortech.2020.123391.
Hosseinzadeh, A., Zhou, J.L., Altaee, A., Li, D., 2022. Machine learning modeling and analysis of biohydrogen production from wastewater by dark fermentation process. Bioresour. Technol. 343, 126111 https://doi.org/10.1016/j.biortech.2021.126111.
Jacob, S., Banerjee, R., 2016. Modeling and optimization of anaerobic codigestion of potato waste and aquatic weed by response surface methodology and artificial neural network coupled genetic algorithm. Bioresour. Technol. 214, 386–395. https://doi.org/10.1016/j.biortech.2016.04.068.
Jadhav, D.A., Carmona-Martínez, A.A., Chendake, A.D., Pandit, S., Pant, D., 2021. Modeling and optimization strategies towards performance enhancement of microbial fuel cells. Bioresour. Technol. 320, 124256 https://doi.org/10.1016/j.biortech.2020.124256.
Jandric, Z., Tchaikovsky, A., Zitek, A., Causon, T., Stursa, V., Prohaska, T., Hann, S., 2021. Multivariate modelling techniques applied to metabolomic, elemental and isotopic fingerprints for the verification of regional geographical origin of Austrian carrots. Food Chem. 338, 127924 https://doi.org/10.1016/j.foodchem.2020.127924.
Jia, R., Song, Y.-C., Piao, D.-M., Kim, K., Lee, C.-Y., Park, J., 2022. Exploration of deep learning models for real-time monitoring of state and performance of anaerobic digestion with online sensors. Bioresour. Technol. 363, 127908 https://doi.org/10.1016/j.biortech.2022.127908.
Jin, G., Wang, Y., Li, L., Shen, S., Deng, W.-W., Zhang, Z., Ning, J., 2020. Intelligent evaluation of black tea fermentation degree by FT-NIR and computer vision based on data fusion strategy. LWT 125, 109216. https://doi.org/10.1016/j.lwt.2020.109216.
Jin, G., Wang, Y., Li, M., Li, T., Huang, W., Li, L., Deng, W.-W., Ning, J., 2021. Rapid and real-time detection of black tea fermentation quality by using an inexpensive data fusion system. Food Chem. 358, 129815 https://doi.org/10.1016/j.foodchem.2021.129815.
Justice, C., Brix, A., Freimark, D., Kraume, M., Pfromm, P., Eichenmueller, B., Czermak, P., 2011. Process control in cell culture technology using dielectric spectroscopy. Biotechnol. Adv. 29, 391–401. https://doi.org/10.1016/j.biotechadv.2011.03.002.
Kannangara, M., Dua, R., Ahmadi, L., Bensebaa, F., 2018. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. Waste Manag. 74, 3–15. https://doi.org/10.1016/j.wasman.2017.11.057.
Kano, M., Fujiwara, K., 2013. Virtual sensing technology in process industries: Trends and Challenges Revealed by Recent Industrial Applications. J. Chem. Eng. Jpn 46, 1–17. https://doi.org/10.1252/jcej.12we167.
Khan, M., Ullah, Z., Mašek, O., Raza Naqvi, S., Nouman Aslam Khan, M., 2022. Artificial neural networks for the prediction of biochar yield: A comparative study of metaheuristic algorithms. Bioresour. Technol. 355, 127215 https://doi.org/10.1016/j.biortech.2022.127215.
Khanijou, J.K., Kulyk, H., Bergès, C., Khoo, L.W., Ng, P., Yeo, H.C., Helmy, M., Bellvert, F., Chew, W., Selvarajoo, K., 2022. Metabolomics and modelling approaches for systems metabolic engineering. Metab. Eng. Commun. 15, e00209.
Kumar Saini, D., Yadav, D., Pabbi, S., Chhabra, D., Shukla, P., 2020. Phycobiliproteins from *Anabaena variabilis* CCC421 and its production enhancement strategies using combinatory evolutionary algorithm approach. Bioresour. Technol. 309, 123347 https://doi.org/10.1016/j.biortech.2020.123347.
Lahav, O., Morgan, B., 2004. Titration methodologies for monitoring of anaerobic digestion in developing countries—a review. J. Chem. Technol. Biotechnol. 79, 1331–1341. https://doi.org/10.1002/jctb.1143.
Li, Y., Nuamah, L.A., Pu, Y., Zhang, H., Norgbey, E., Nwankwegu, A.S., Banahene, P., Bofah, R., 2022. A radial basis function neural network based multi-objective optimization for simultaneously enhanced nitrogen and phosphorus removal in a full-scale integrated surface flow treatment wetland–pond system. Bioresour. Technol. 344, 126336 https://doi.org/10.1016/j.biortech.2021.126336.
Liu, J., Jin, S., Bao, C., Sun, Y., Li, W., 2021. Rapid determination of lignocellulose in corn stover based on near-infrared reflectance spectroscopy and chemometrics methods. Bioresour. Technol. 321, 124449 https://doi.org/10.1016/j.biortech.2020.124449.
Liu, L., Sun, J., Xu, W., Du, G., Chen, J., 2009. Modeling and optimization of microbial hyaluronic acid production by *Streptococcus zooepidemicus* using radial basis function neural network coupling quantum-behaved particle swarm optimization algorithm. Biotechnol. Prog. 25, 1819–1825. https://doi.org/10.1002/btpr.278.

Luttmann, R., Bracewell, D.G., Cornelissen, G., Gernaey, K.V., Glassey, J., Hass, V.C., Kaiser, C., Preusse, C., Striedner, G., Mandenius, C.-F., 2012. Soft sensors in bioprocessing: A status report and recommendations. Biotechnol. J. 7 (8), 1040–1048.

Mahlein, A.-K., 2016. Plant disease detection by Imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. Plant Dis. 100, 241–251. https://doi.org/10.1094/PDIS-03-15-0340-FE.

Mahmoodi-Babolan, N., Heydari, A., Nematollahzadeh, A., 2019. Removal of methylene blue via bioinspired catecholamine/starch superadsorbent and the efficiency prediction by response surface methodology and artificial neural network-particle swarm optimization. Bioresour. Technol. 294, 122084 https://doi.org/10.1016/j.biortech.2019.122084.

Maiti, S.K., Srivastava, R.K., Bhushan, M., Wangikar, P.P., 2009. Real time phase detection based online monitoring of batch fermentation processes. Process Biochem. 44, 799–811. https://doi.org/10.1016/j.procbio.2009.03.008.

Mol, V., Bennett, M., Sánchez, B.J., Lisowska, B.K., Herrgård, M.J., Nielsen, A.T., Leak, D.J., Sonnenschein, N., 2021. Genome-scale metabolic modeling of P. *thermoglucosidasius NCIMB* 11955 reveals metabolic bottlenecks in anaerobic metabolism. Metab. Eng. 65, 123–134. https://doi.org/10.1016/j.ymben.2021.03.002.

Moncks, P.C.S., Corrêa, É.K., Guidoni, L.L.C., Moncks, R.B., Corrêa, L.B., Lucia Jr, T., Araujo, R.M., Yamin, A.C., Marques, F.S., 2022. Moisture content monitoring in industrial-scale composting systems using low-cost sensor-based machine learning techniques. Bioresour. Technol. 359, 127456.

Moodley, P., Rorke, D.C.S., Gueguim Kana, E.B., 2019. Development of artificial neural network tools for predicting sugar yields from inorganic salt-based pretreatment of lignocellulosic biomass. Bioresour. Technol. 273, 682–686. https://doi.org/10.1016/j.biortech.2018.11.034.

Moreira, C., Chou, Y.-L., Velmurugan, M., Ouyang, C., Sindhgatta, R., Bruza, P., 2021. LINDA-BN: An interpretable probabilistic approach for demystifying black-box predictive models. Decis. Support Syst. 150, 113561 https://doi.org/10.1016/j.dss.2021.113561.

Mullai, P., Vishali, S., Sobiya, E., 2022. Experiments and adaptive-network-based fuzzy inference system modelling in a hybrid up-flow anaerobic sludge blanket reactor to assess industrial azadirachtin effluent quality. Bioresour. Technol. 358, 127395 https://doi.org/10.1016/j.biortech.2022.127395.

Murugan, C., Natarajan, P., 2019. Estimation of fungal biomass using multiphase artificial neural network based dynamic soft sensor. J. Microbiol. Methods 159, 5–11. https://doi.org/10.1016/j.mimet.2019.02.002.

Natarajan, P., Moghadam, R., Jagannathan, S., 2021. Online deep neural network-based feedback control of a Lutein bioprocess. J. Process Control 98, 41–51. https://doi.org/10.1016/j.jprocont.2020.11.011.

Nguyen, D.D., Quy Duc Pham, T., Tanveer, M., Khan, H., Park, J.W., Park, C.W., Kim, G.M., 2022. Deep learning–based optimization of a microfluidic membraneless fuel cell for maximum power density via data-driven three-dimensional multiphysics simulation. Bioresour. Technol. 348, 126794.

Nikzad-Langerodi, R., Lughofer, E., Saminger-Platz, S., Zahel, T., Sagmeister, P., Herwig, C., 2017. Automatic feed phase identification in multivariate bioprocess profiles by sequential binary classification. Anal. Chim. Acta 982, 48–61. https://doi.org/10.1016/j.aca.2017.05.034.

Pappu, S.M.J., Gummadi, S.N., 2017. Artificial neural network and regression coupled genetic algorithm to optimize parameters for enhanced xylitol production by *Debaryomyces nepalensis* in bioreactor. Biochem. Eng. J. 120, 136–145. https://doi.org/10.1016/j.bej.2017.01.010.

Patel, A.K., Choi, Y.Y., Sim, S.J., 2020. Emerging prospects of mixotrophic microalgae: Way forward to sustainable bioprocess for environmental remediation and cost-effective biofuels. Bioresour. Technol. 300, 122741 https://doi.org/10.1016/j.biortech.2020.122741.

Peng, J., Meng, F., Ai, Y., 2013. Time-dependent fermentation control strategies for enhancing synthesis of marine bacteriocin 1701 using artificial neural network and genetic algorithm. Bioresour. Technol. 138, 345–352. https://doi.org/10.1016/j.biortech.2013.03.194.

Rathore, A.S., Mishra, S., Nikita, S., Priyanka, P., 2021. Bioprocess control: Current progress and future perspectives. Life 11, 557. https://doi.org/10.3390/life11060557.

Reardon, K.F., 2021. Practical monitoring technologies for cells and substrates in biomanufacturing. Curr. Opin. Biotechnol. 71, 225–230. https://doi.org/10.1016/j.copbio.2021.08.006.

Renata, H., Wang, Z.J., Arnold, F.H., 2015. Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. Angew. Chem.-Int. Edit. 54, 3351–3367. https://doi.org/10.1002/anie.201409470.

Saini, D.K., Rai, A., Devi, A., Pabbi, S., Chhabra, D., Chang, J.-S., Shukla, P., 2021. A multi-objective hybrid machine learning approach-based optimization for enhanced biomass and bioactive phycobiliproteins production in *Nostoc* sp. CCC-403. Bioresour. Technol. 329, 124908 https://doi.org/10.1016/j.biortech.2021.124908.

Sakr, G.E., Mokbel, M., Darwich, A., Khneisser, M.N., Hadi, A., 2016. Comparing deep learning and support vector machines for autonomous waste sorting, in: 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET). IEEE, pp. 207–212. https://doi.org/10.1109/IMCET.2016.7777453.

Sharma, A.B., Golubchik, L., Govindan, R., 2010. Sensor faults. ACM Trans. Sens. Netw. 6, 1–39. https://doi.org/10.1145/1754414.1754419.

Sharma, D., Pandey, A.K., Yadav, K.D., Kumar, S., 2021. Response surface methodology and artificial neural network modelling for enhancing maturity parameters during vermicomposting of floral waste. Bioresour. Technol. 324, 124672 https://doi.org/10.1016/j.biortech.2021.124672.

Shi, C.-F., Yang, H.-T., Chen, T.-T., Guo, L.-P., Leng, X.-Y., Deng, P.-B., Bi, J., Pan, J.-G., Wang, Y.-M., 2022. Artificial neural network-genetic algorithm-based optimization of aerobic composting process parameters of Ganoderma lucidum residue. Bioresour. Technol. 357, 127248 https://doi.org/10.1016/j.biortech.2022.127248.

Singhal, A., Kumar, M., Bhattacharya, M., Kumari, N., Jha, P.K., Chauhan, D.K., Thakur, I.S., 2018. Pretreatment of *Leucaena leucocephala* wood by acidified glycerol: optimization, severity index and correlation analysis. Bioresour. Technol. 265, 214–223. https://doi.org/10.1016/j.biortech.2018.05.084.

Sinner, P., Stiegler, M., Herwig, C., Kager, J., 2021. Noninvasive online monitoring of *Corynebacterium glutamicum* fed-batch bioprocesses subject to spent sulfite liquor raw material uncertainty. Bioresour. Technol. 321, 124395 https://doi.org/10.1016/j.biortech.2020.124395.

Survyla, A., Levisauskas, D., Urniezius, R., Simutis, R., 2021. An oxygen-uptake-rate-based estimator of the specific growth rate in *Escherichia coli* BL21 strains cultivation processes. Comp. Struct. Biotechnol. J. 19, 5856–5863. https://doi.org/10.1016/j.csbj.2021.10.015.

Tai, C., Voltan, D.S., Keshwani, D.R., Meyer, G.E., Kuhar, P.S., 2016. Fuzzy logic feedback control for fed-batch enzymatic hydrolysis of lignocellulosic biomass. Bioprocess Biosyst. Eng. 39, 937–944. https://doi.org/10.1007/s00449-016-1573-1.

Takahashi, M.B., Leme, J., Caricati, C.P., Tonso, A., Fernández Núñez, E.G., Rocha, J.C., 2015. Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes. Bioprocess Biosyst. Eng. 38, 1045–1054. https://doi.org/10.1007/s00449-014-1346-7.

Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wang, S., Chen, H., Raza Naqvi, S., 2021. Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods: Effects of pyrolysis conditions and biomass characteristics. Bioresour. Technol. 339, 125581 https://doi.org/10.1016/j.biortech.2021.125581.

Treloar, N.J., Fedorec, A.J.H., Ingalls, B., Barnes, C.P., You, L., 2020. Deep reinforcement learning for the control of microbial co-cultures in bioreactors. PLoS Comput. Biol. 16 (4), e1007783.

Vinoth Arul Raj, J., Praveen Kumar, R., Vijayakumar, B., Gnansounou, E., Bharathiraja, B., 2021. Modelling and process optimization for biodiesel production from *Nannochloropsis salina* using artificial neural network. Bioresour. Technol. 329, 124872.

Wang, G., Haringa, C., Noorman, H., Chu, J., Zhuang, Y., 2020a. Developing a computational framework to advance bioprocess scale-up. Trends Biotechnol. 38, 846–856. https://doi.org/10.1016/j.tibtech.2020.01.009.

Wang, L., Long, F., Liao, W., Liu, H., 2020b. Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. Bioresour. Technol. 298, 122495 https://doi.org/10.1016/j.biortech.2019.122495.

Xing, J., Kurose, R., Luo, K., Fan, J., 2022. Chemistry-Informed Neural Networks modelling of lignocellulosic biomass pyrolysis. Bioresour. Technol. 355, 127275 https://doi.org/10.1016/j.biortech.2022.127275.

Xu, R.-Z., Cao, J.-S., Luo, J.-Y., Feng, Q., Ni, B.-J., Fang, F., 2022. Integrating mechanistic and deep learning models for accurately predicting the enrichment of polyhydroxyalkanoates accumulating bacteria in mixed microbial cultures. Bioresour. Technol. 344, 126276 https://doi.org/10.1016/j.biortech.2021.126276.

Xu, Y., Wu, Y., Lv, X., Sun, G., Zhang, H., Chen, T., Du, G., Li, J., Liu, L., 2021. Design and construction of novel biocatalyst for bioprocessing: Recent advances and future outlook. Bioresour. Technol. 332, 125071 https://doi.org/10.1016/j.biortech.2021.125071.

Xue, W., Hu, X., Wei, Z., Mei, X., Chen, X., Xu, Y., 2019. A fast and easy method for predicting agricultural waste compost maturity by image-based deep learning. Bioresour. Technol. 290, 121761 https://doi.org/10.1016/j.biortech.2019.121761.

Yang, G., Li, Y., Zhen, F., Xu, Y., Liu, J., Li, N., Sun, Y., Luo, L., Wang, M., Zhang, L., 2021. Biochemical methane potential prediction for mixed feedstocks of straw and manure in anaerobic co-digestion. Bioresour. Technol. 326, 124745 https://doi.org/10.1016/j.biortech.2021.124745.

Yang, H., Liu, Y., Li, J., Liu, L., Du, G., Chen, J., 2020. Systems metabolic engineering of *Bacillus subtilis* for efficient biosynthesis of 5-methyltetrahydrofolate. Biotechnol. Bioeng. 117, 2116–2130. https://doi.org/10.1002/bit.27332.

Zafar, M., Kumar, S., Kumar, S., Dhiman, A.K., 2012. Artificial intelligence based modeling and optimization of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) production process by using *Azohydromonas lata* MTCC 2311 from cane molasses supplemented with volatile fatty acids: A genetic algorithm paradigm. Bioresour. Technol. 104, 631–641. https://doi.org/10.1016/j.biortech.2011.10.024.

Zhang, L., Chao, B., Zhang, X., 2020. Modeling and optimization of microbial lipid fermentation from cellulosic ethanol wastewater by *Rhodotorula glutinis* based on the support vector machine. Bioresour. Technol. 301, 122781 https://doi.org/10.1016/j.biortech.2020.122781.

Zhang, D., del Rio-Chanona, E.A., Petsagkourakis, P., Wagner, J., 2019. Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. Biotechnol. Bioeng. 116, 2919–2930. https://doi.org/10.1002/bit.27120.

Zhang, W., Li, J., Liu, T., Leng, S., Yang, L., Peng, H., Jiang, S., Zhou, W., Leng, L., Li, H., 2021. Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae. Bioresour. Technol. 342, 126011 https://doi.org/10.1016/j.biortech.2021.126011.

Zhou, X., Zhao, J., Chen, M., Wu, S., Zhao, G., Xu, S., 2022. Effects of hydration parameters on chemical properties of biocrudes based on machine learning and experiments. Bioresour. Technol. 350, 126923 https://doi.org/10.1016/j.biortech.2022.126923.