

Review

A Review of Artificial Intelligence Applications for Biorefineries and Bioprocessing: From Data-Driven Processes to Optimization Strategies and Real-Time Control

Alex Butean ^{1,2,*} , Iulia Cutean ¹, Ruben Barbero ³ , Juan Enriquez ⁴ and Alexandru Matei ^{1,2} 

¹ Wiz Development and Services, 4 Samuel Von Brukenthal, 550178 Sibiu, Romania

² Computer Science and Electrical Engineering Department, Lucian Blaga University of Sibiu, Bulevardul Victoriei 10, 550024 Sibiu, Romania

³ Department of Energy Engineering, Universidad Nacional de Educación a Distancia (UNED), C/Juan del Rosal 12, 28040 Madrid, Spain

⁴ Dynamic & Security Computations SL, 54S1 C Fermín Caballero, 28034 Madrid, Spain

* Correspondence: alex@butean.com

Abstract

This paper reviews the integration of artificial intelligence (AI) and machine learning in biorefineries and bioprocessing, with applications in biocatalysis, enzyme optimization, real-time monitoring, and quality assurance. AI contributes to predictive modeling and allows the precise forecasting of process outcomes, resource management, and energy utilization. AI models, including supervised, unsupervised, and reinforcement learning, support improvements in important bioprocess stages, such as fermentation, purification, and microbial biosynthesis. Digital twins and soft-sensing technologies enable real-time control and increase operational precision in complex bioprocess environments. Hybrid modeling integrates data-driven AI techniques with common scientific principles, improving scalability and adaptability under dynamic operational conditions. This review addresses challenges in AI implementation, such as data standardization, model transparency, and the need for interdisciplinary collaboration. The discussion concludes with future directions and sustainable AI strategies, highlighting the potential of AI to strengthen scalable, efficient, and environmentally sustainable biorefinery operations. These findings highlight how AI-driven methodologies improve operational efficiency, reduce resource waste, and facilitate sustainable innovation in bioprocesses, thereby strengthening sustainability within the bioeconomy.

Keywords: artificial intelligence; machine learning; bioprocessing; biorefineries; predictive modeling; process optimization



Academic Editors: Giannis Penoglou, Alexandros Pavlou and Xiong Luo

Received: 25 June 2025

Revised: 9 July 2025

Accepted: 11 July 2025

Published: 12 August 2025

Citation: Butean, A.; Cutean, I.; Barbero, R.; Enriquez, J.; Matei, A. A Review of Artificial Intelligence Applications for Biorefineries and Bioprocessing: From Data-Driven Processes to Optimization Strategies and Real-Time Control. *Processes* **2025**, *13*, 2544. <https://doi.org/10.3390/pr13082544>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Growing concerns regarding resource consumption, environmental degradation, and climate change increasingly pressure the global economy to shift from finite fossil-based resources toward sustainable alternatives. In this context, biorefineries and bioprocessing are central to advancing the transition toward a more sustainable and climate-neutral economy, aligning well with global environmental initiatives like the European Green Deal [1]. These systems facilitate decarbonization across multiple sectors and promote circular resource utilization by converting waste into valuable materials. Moreover, biorefineries reduce fossil resource dependency and drive sustainable technological advancements.

A biorefinery is an integrated facility or system that fractionates and transforms renewable biomass—food waste, plants, and organic matter—into energy and new products. Employing diverse conversion technologies—chemical, biochemical, or thermochemical—biorefineries produce biofuels, bio-based chemicals, biomaterials, biofuels, or heat and electricity as co-products. These outputs support a variety of applications, including medical devices, pharmaceuticals, tissue engineering, drug delivery systems, and radiological technologies [2].

Bioprocessing specifically refers to processes that utilize living cells—microorganisms, plant cells, or animal cells—or their components to produce a specific useful product [2]. Key examples include fermentation, enzymatic hydrolysis, anaerobic digestion, and cell culture.

Despite their potential, biorefineries are characterized by operational complexity. These systems involve highly interconnected, dynamic, and data-intensive processes, spanning from biomass sourcing to conversion and product separation. Efficient operations require considering factors like the geographic availability and seasonality of biomass, optimizing multivariate processes for multiple products, and managing waste streams effectively. Successful industrial-scale implementations of biorefineries require careful consideration as well as balancing the economic, environmental, and social sustainability dimensions. Conventional biomass processing methods, like incineration, often waste valuable resources. Moreover, many current biorefinery initiatives suffer from a lack of holistic integration approaches and limited data integration, both of which reduce their effectiveness [3].

Considering the integrated and data-intensive nature of biorefineries and bioprocessing facilities, along with their need for sustainability and economic competitiveness, optimization is crucial across all stages. Traditional optimization approaches, such as mechanistic modeling and statistically designed experiments, often prove insufficient for addressing the nonlinear dynamics, high-dimensional data, and intricate interactions inherent in biorefinery systems. This limitation necessitates the adoption of more powerful, data-driven approaches. The convergence of pressing sustainability mandates and the technological advancements offered by Industry 4.0—encompassing digitalization, automation, Big Data analytics, and artificial intelligence (AI)—creates a strong impetus for leveraging these new tools.

This review critically evaluates AI and machine learning (ML) applications in biorefineries, synthesizes challenges and opportunities, and identifies unresolved questions to guide future research. AI, and particularly its subfield ML, offers a powerful suite of data-driven techniques well-suited to addressing the optimization challenges in biorefineries and bioprocessing [2]. ML algorithms are designed to learn patterns and relationships directly from data, enabling them to model, predict, and optimize complex systems without requiring complete mechanistic understanding or explicit programming for every task. AI and ML excel at analyzing the large and often heterogeneous datasets generated throughout the biorefinery value chain, uncovering hidden patterns, generating actionable insights, and advancing autonomous or semi-autonomous decision-making [4]. With the validation of AI and ML, the design of biomaterials has improved and become more efficient [5]. The biorefineries and bioprocesses that use AI and ML achieve results more quickly and at lower costs.

This review adopts principles derived from the PRISMA guidelines. The relevant literature was systematically searched across multiple databases, specifically Web of Science, IEEE Xplore, and Scopus. Using logical operators, the following specific keywords were applied to filter the search results: artificial intelligence (AI), machine learning (ML), biorefinery, bioprocessing, optimization, control, monitoring, digital twin (DT), applications, and biofuel. Selected articles are peer-reviewed and were published no earlier than 2020.

Articles that did not directly address the topic of this review, or lacked sufficient technical or methodological details, were excluded from the final selection.

To reinforce the integration of artificial intelligence (AI) within bioprocessing and biorefinery systems, it is essential to distinguish between three fundamental modeling approaches: mechanistic, data-driven, and hybrid models.

Mechanistic models are derived from fundamental physical and chemical laws, such as mass and energy balances, thermodynamics, and reaction kinetics. These models are valued for their transparency and reliability in well-defined systems, but may be less effective when applied to processes with high complexity or incomplete characterization [6].

Data-driven models, including those based on machine learning techniques, identify predictive relationships directly from experimental or operational data. These models are well-suited to nonlinear, multivariable processes and can adapt to changing conditions. However, their reliance on large datasets and the often limited interpretability of their internal structure can restrict their applicability in sensitive or regulated environments [4].

Hybrid models combine mechanistic models, grounded in fundamental scientific laws, with empirical data-driven methods, enhancing generalizability, adaptability, and stability under dynamic or variable operating conditions. These models are particularly relevant in modern bioprocess applications, where partial knowledge of underlying mechanisms coexists with extensive process data.

Throughout this review, these modeling frameworks are referenced to contextualize AI-based strategies for process optimization, monitoring, and control, with a focus on enabling sustainable and scalable biomanufacturing.

2. AI Applications in Biorefinery Optimization

Biorefineries convert biomass into biofuels, chemicals, and other value-added products, driving progress toward a sustainable bioeconomy. These operations are built upon multi-phase workflows, such as fermentation, extraction, and purification, and require careful control due to their sensitivity to feedstock variability and environmental shifts. Traditional methods often lack the flexibility to manage these dynamic conditions, pointing out the need for innovative solutions [6]. AI promotes these goals as it provides correct, real-time adjustments that coordinate production processes with targets for sustainability and effectiveness. Achieving these goals requires defining specific objectives that can be translated into mathematical expressions. These optimization objectives are invariably subject to a set of constraints, which are mathematical inequalities or equalities that represent the physical and technical limitations of the system. These constraints can include limits on reactor temperature and pressure, material balances that must be respected, or maximum flow rates. Algorithms such as reinforcement learning (RL) contribute to decision-making in complex, multi-stage biorefinery processes. These models interpret historical and real-time data to provide insights that identify optimal conditions for efficiency and yield. RL agents adapt continuously to changes in feedstock quality and environmental conditions, proving essential in biorefinery environments marked by variability. Optimization of process flow benefits from AI's ability to detect and address bottlenecks within multi-stage operations. In biorefineries, bottlenecks in a single phase can impact the entire production chain. AI-driven models identify inefficiencies and recommend adjustments to support balanced, continuous flows across each stage. Recent studies present improvements in process flows as AI optimizes reaction sequences and allocates inputs effectively [7], contributing to overall process stability. This section explores AI's contributions to biorefinery optimization, focusing on applications in fermentation control, yield improvement, resource allocation, and energy efficiency, while also addressing broader challenges related to scaling and consistency.

An important application of ML in the biomaterials field is the development of models that predict the behavior or characteristics of biomaterials. These models use algorithms trained on extensive datasets to make predictions about factors such as biocompatibility and degradation rate. By using predictive models, we gain numerous advantages, such as better and more accurate predictions, reduced resource usage, and less time consumption. Additionally, ML facilitates the analysis and interpretation of high-dimensional imaging data in the domain of biomaterials. These images are complex, making their analysis challenging. By using ML algorithms, image analysis becomes feasible, facilitating the diagnosis, monitoring, and treatment of diseases. The integration of AI and ML in biorefining and bioprocessing has resulted in multiple operational advantages, as illustrated in Figure 1 and detailed below:

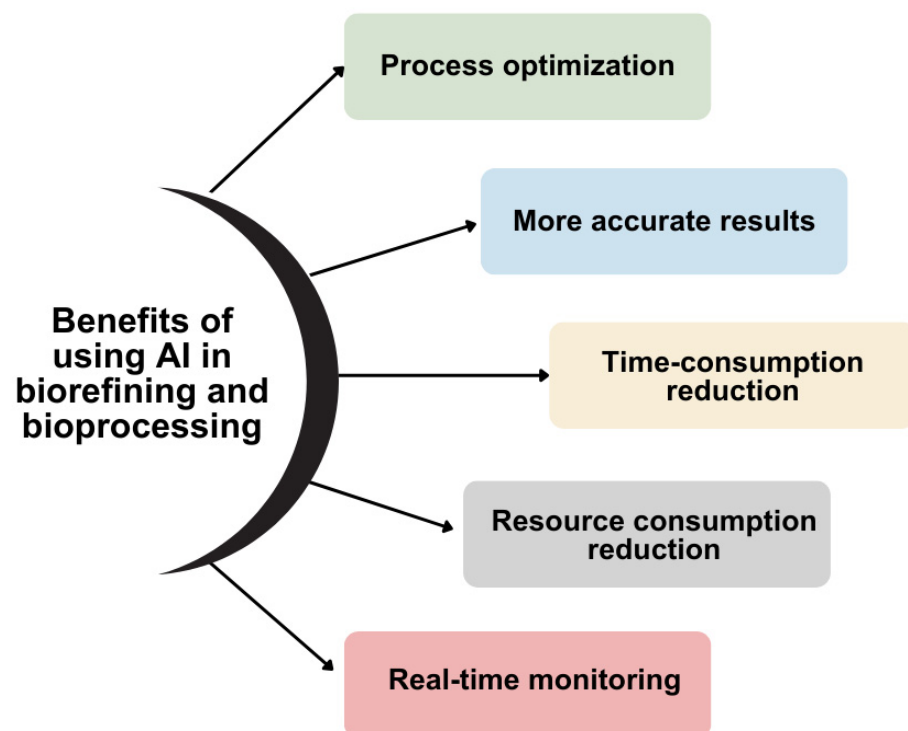


Figure 1. Advantages of using AI and ML in biorefining and bioprocessing.

- **Process optimization:** The integration of AI and ML in biorefinery and bioprocessing results in the design of superior processes, contributing to increased productivity. These AI-driven technologies enable process automation, improve results accuracy, and provide support throughout the process by enabling real-time monitoring.
- **More accurate results:** The machines used in biorefineries and bioprocessing that utilize AI and ML eliminate errors and succeed in performing a more comprehensive data analysis. This leads to clearer results.
- **Time-consumption reduction:** Another advantage of using AI and ML in biorefineries and bioprocesses is faster production through process automation. Thus, these systems are capable of optimizing parameters, predicting optimal conditions, detecting various errors, etc.
- **Resource consumption reduction:** AI and ML models can predict process conditions, raw material characteristics, biomaterial behavior, and other factors that contribute to reducing resource consumption. Article [8] explores how using ML models to predict the properties and uncertainties of the process lowers resource consumption.

- **Real-time monitoring:** Real-time monitoring facilitates the continuous analysis and adjustment of data, resulting in more efficient processes and improved outcomes. Moreover, automated monitoring can promptly detect and correct deviations and errors.

2.1. Applications in Fermentation Control and Yield Improvement

A primary optimization objective in biorefineries is maximizing product yield and quality from fermentation, a core bioprocess that converts raw biomass into products such as biofuels, enzymes, and specialty chemicals. AI improves fermentation control by managing factors like temperature, pH, and nutrient supply. These factors directly affect product yield and quality. Model predictive control (MPC) and neural networks forecast the effects of adjustments to fermentation parameters and guide operators toward settings that foster high-quality fermentation results.

ML models contribute to yield improvement by identifying ideal conditions for biomass conversion and tracking variables that affect productivity. Recent studies indicate that AI predictions of yield and conversion rates enable timely interventions to maximize output [9]. Adaptive control strategies based on AI outputs ensure consistent yields, even as fermentation conditions change [10]. Soft sensors, or virtual sensors, increase fermentation by estimating variables that are difficult to measure directly, such as microbial growth rates and substrate use. Digital twins (DTs), virtual representations of physical systems, replicate fermentation processes in real-time and allow operators to test adjustments without disrupting the physical system. This approach increases control over fermentation parameters and ensures optimal yield and product quality [11].

2.2. Resource Allocation and Waste Reduction

Resource allocation and waste reduction remain the main objectives in biorefineries, where variability in input materials often complicates consistent production. AI models, particularly supervised learning and predictive algorithms, allow for the data-driven management of resources like raw materials, energy, and time. Neural networks reveal correlations between input variables and waste, supporting strategies for more efficient production cycles and sustainable waste management.

The predictive capabilities of AI also contribute to waste reduction by identifying potential reuse options for by-products. For example, AI models predict waste volumes and composition, allowing operators to repurpose by-products as feedstock in other production stages. This closed-loop approach aligns with the circular economy principles, converting waste into valuable resources and reducing environmental impact. Recent articles demonstrate how AI identifies valuable by-products, such as essential oils and biogas [12], from agricultural waste, like orange peels [13]. These by-products serve as feedstock in other production stages, which advances both economic and environmental sustainability within a circular economy framework [14].

2.3. Optimization of Energy Consumption

Minimizing energy consumption is a crucial optimization objective that directly impacts both the economic viability and environmental footprint of a biorefinery. Sustainability in biorefineries critically depends on optimizing resource utilization throughout the value chain, with energy consumption being a key factor influencing environmental performance and economic success. High energy demands during biomass processing, conversion, and product separation can significantly offset the intended environmental benefits [15].

AI-assisted energy management systems analyze usage patterns to identify optimal settings that reduce energy demand while maintaining process quality. AI models predict energy requirements under different process conditions and enable operators to adjust

energy-intensive tasks, such as heating and agitation, to line up with production needs and reduce peak energy loads.

Building upon the analysis of past data, AI can predict future energy requirements. Predictive models—ranging from time-series forecasting techniques (like ARIMA, used for demand estimation) to more advanced ML algorithms (e.g., random forest, XGBoost, and AdaBoost used for bioethanol demand forecasting), as well as artificial neural networks (ANNs, used for waste heat recovery prediction)—can forecast energy demand based on a multitude of influencing factors. These factors include planned production schedules, the type and characteristics of biomass feedstock being processed, anticipated process conditions, and external variables, like ambient temperature and weather patterns, which impact heating and cooling loads [16].

MPC and fuzzy logic systems, which use AI for the adaptive control of process parameters, help manage energy use in dynamic environments. AI-assisted systems predict and adjust energy input to match production demand, thereby reducing the environmental impact.

2.4. Challenges in Scalability and Maintaining Process Consistency

Scaling AI models from pilot to industrial levels in biorefineries presents technical and operational challenges, particularly in maintaining process consistency under varying conditions. As operations transition from controlled pilot settings to industrial-scale production, variability in biomass feedstocks, temperature, pH, and other factors can disrupt processes and lead to inconsistencies in product quality and yield. Addressing these challenges requires targeted adaptations that allow AI models to handle real-world variability without compromising accuracy or reliability.

Table 1 describes the main challenges in scaling AI within biorefineries and the corresponding AI-driven solutions that provide targeted benefits. By employing predictive modeling through RL, adaptive control systems, and hybrid models, biorefineries can mitigate these challenges. These methods lead to improved efficiency, consistent product quality, and reliable scalability.

Table 1. Challenges and solutions for scalability and consistency.

Challenge	AI Solution	Benefits
Variability in feedstock	Predictive Modeling	Consistent product quality
Process flow inefficiencies	RL	Enhanced throughput
Energy and resource fluctuations	Adaptive MPC, Fuzzy Logic Systems	Reduced energy costs, efficiency
Maintaining consistency at scale	Hybrid Models	Reliable, scalable processes

One practical approach to achieving initial scalability involves hybrid models that combine data-driven predictions with established scientific knowledge. These models integrate kinetic modeling with ML to align processing conditions and maintain consistent results across production scales. By bridging empirical data with theoretical frameworks, hybrid models provide greater resilience to fluctuations in biorefinery inputs and conditions, facilitating the transition from pilot to industrial settings.

Studies such as [17] highlight the importance of collaboration between AI specialists, engineers, and scientists to overcome the early scalability challenges of AI in biorefineries. This interdisciplinary approach ensures that AI-driven models reflect both practical and scientific insights, strengthening the adjustments needed to maintain consistency at larger scales.

3. Machine Learning for Bioprocess Modeling and Control

In bioprocessing, maintaining precise control over interconnected, multi-stage operations is critical to achieving reproducible and high-quality results. Minor fluctuations in conditions, such as temperature, pH, and feed rates, often impact outcomes, making stability, efficiency, and adaptability challenging to maintain. Traditional control methods struggle with these variabilities, especially in dynamic environments, like fermentation and purification, where conditions change rapidly.

ML meets these challenges through data-driven modeling, enabling real-time decision-making and precise control over each stage of the bioprocess [18]. ML models analyze patterns in both historical and real-time data, allowing operators to predict outcomes, optimize settings, and respond to fluctuations. This section explores how different ML approaches—including supervised, unsupervised, and RL—enhance control across bioprocess stages, supporting more reliable, efficient, and adaptive operations.

3.1. Machine Learning Applications in Bioprocessing

In bioprocessing, ML contributes to predictive modeling, process optimization, and the development of adaptive systems capable of responding to changing conditions across multiple stages. ML methods analyze vast datasets and reveal insights that are often inaccessible to conventional methods. As shown in Figure 2, the primary ML methods used in bioprocessing include supervised learning, unsupervised learning, and RL. Each method offers distinct capabilities for extracting actionable knowledge from process data, depending on the structure and availability of labeled or unlabeled datasets.

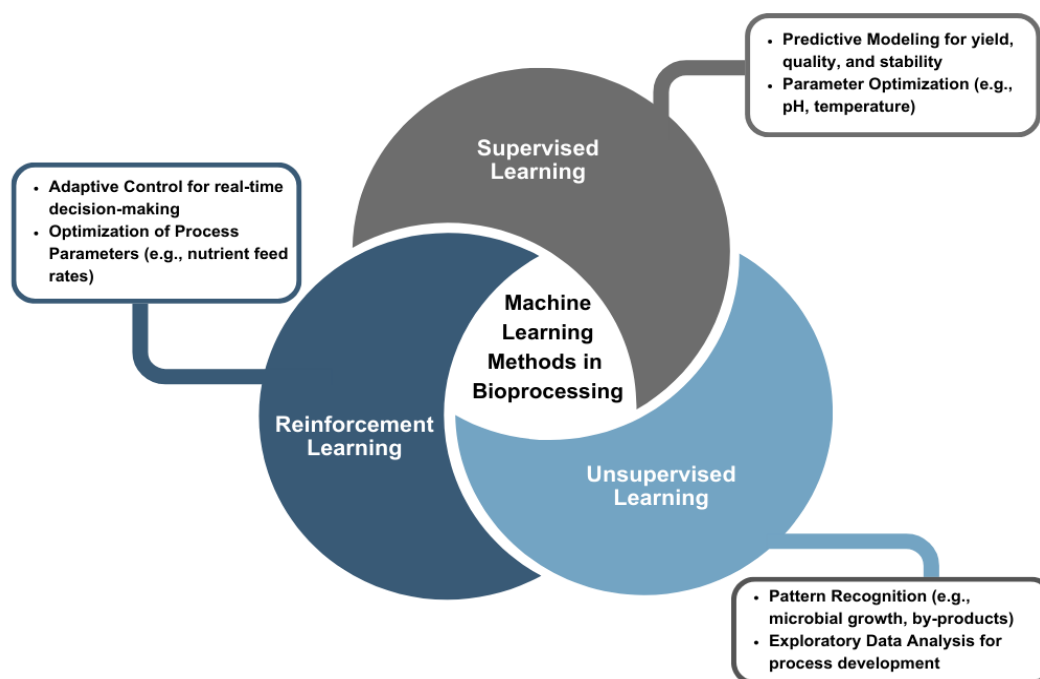


Figure 2. ML methods in bioprocessing.

Supervised learning is one of the most widely used ML methods in bioprocess modeling. It relies on labeled datasets with clearly defined input–output relationships to predict outcomes based on historical patterns. This enables operators to anticipate process fluctuations and make proactive adjustments to avoid potential disturbances. Supervised learning algorithms, such as regression models and neural networks, are used to generate accurate predictions for performance indicators, including yield, product quality, and process stability [18]. These models reveal the effects of specific parameters on outcomes, which

improves the accuracy and reliability of production results. For example, in therapeutic protein production, supervised learning anticipates how changes in pH, temperature [19], or nutrient concentration affect product quality [20]. This enables operators to make proactive adjustments to process parameters, helping maintain consistent quality standards even as conditions fluctuate. The insights generated by supervised learning drive a stable bioprocessing environment, which is particularly important in regulated sectors, such as pharmaceuticals and biomanufacturing. One challenge is that bioprocess datasets are frequently characterized by heterogeneity, different time scales, and multiple sources that, coupled with high data dimensionality, substantial cost, and time associated with experimental data acquisition, often result in small or limited training datasets [21]. Successful supervised learning applications include the optimization of cell culture conditions with a titer increase of up to 48% [22], anomaly detection [23], or soft sensor development [24].

While supervised learning provides predictive insights, unsupervised learning shifts the focus toward exploring and identifying hidden patterns within datasets without labeled data. This method proves invaluable during the exploratory stages of process development, where the objective is to uncover relationships and groupings within large amounts of raw data [18]. In bioprocessing, unsupervised learning models identify distinct microbial growth patterns, fermentation by-products, or processing trends that might go unobserved with traditional methods. These insights support the adjustment of process parameters and refinement of experimental design to achieve improved outcomes. Main applications include anomaly or contamination detection [23,25] using support vector machines (SVMs) and autoencoders, process monitoring using long short-term memory autoencoders [26], or deep learning [27]. Beyond anomaly detection and batch monitoring, unsupervised learning is also used for phenotype analysis [28] and cell morphology analysis [29].

In contrast, RL focuses on decision-making in dynamic environments. RL algorithms learn optimal strategies through trial-and-error interactions, receiving feedback in the form of rewards and continuously improving their performance. This approach proves beneficial in bioprocesses where conditions frequently fluctuate, where RL agents adjust parameters in real-time [18]. For instance, in microbial fermentation [30], RL algorithms regulate factors like nutrient feed rates and oxygen levels to optimize biomass production, adapting as conditions within the bioreactor evolve. RL models foster resilient, adaptive systems that improve through each learning cycle. Most notably RL applications include microbial co-culture population control [31], batch-to-batch bioprocess optimization [32], and general process control, where classical PID control strategies are enhanced by RL methods [33].

3.2. Real-Time Control and Adaptive Feedback in Bioprocessing

Adaptive feedback mechanisms are a key component of modern bioprocess control. These systems measure critical conditions, such as temperature, pH, nutrient concentrations, and oxygen levels, and automatically adjust these parameters to sustain optimal process conditions. By responding continuously to real-time data, adaptive feedback loops help maintain stability across stages, like fermentation and purification, ensuring the process meets targeted standards. This adaptability proves valuable in bioprocessing, where input variability and shifting conditions often impact outputs, like yield or product quality. According to [10], such mechanisms enhance operational stability and optimize resource utilization, reducing the unnecessary consumption of costly materials.

Among adaptive methods, MPC is recognized as one of the most effective. MPC predicts future shifts in process variables and adjusts operational parameters in advance, in contrast to traditional methods that react after changes occur. This proactive approach supports stability in complex, nonlinear bioprocesses. As described in [34], MPC uses predictive models to assess potential changes and optimizes parameters such as temperature

or feed rates, which enhances growth conditions in microbial fermentation and minimizes resource use and waste.

In microbial fermentation, for example, MPC is applied to control nutrient delivery, oxygen levels, and temperature to ensure optimal microbial growth and productivity. These adjustments contribute to higher product yield and lower waste, meeting the goals of sustainable biomanufacturing and ensuring consistent quality in high-demand sectors like biopharmaceuticals. This predictive control helps maintain consistent quality standards, even under fluctuating conditions. In purification, for example, MPC balances flow rates, pH, and buffer conditions, maintaining product purity and preventing contamination. Advanced MPC models integrate data from multiple sensors and create a comprehensive view of the process, allowing synchronized adjustments across variables. This holistic control strategy reduces variability in product quality. As noted in [35], these adaptive systems are particularly valuable for scaling up from laboratory to industrial settings, where increased system complexity and variability present additional control challenges.

In addition to MPC, other technologies, like soft sensing, support adaptive feedback and real-time control in bioprocessing. While MPC directly predicts and adjusts operational variables to stabilize processes, soft sensing enables the indirect, real-time measurement of critical variables.

3.3. Hybrid Models for Improved Adaptability and Control

Hybrid models integrate ML techniques with traditional mechanistic frameworks, uniting data-driven insights with established principles from biology and chemistry. In bioprocessing, this integration enables the development of control strategies that are both responsive to real-time data and anchored in fundamental scientific laws. ML identifies patterns in high-dimensional datasets, while mechanistic models provide predictive stability. Their integration enhances robustness and adaptability. This synergy allows hybrid models to address the challenges of bioprocess environments more effectively than either modeling approach alone.

Table 2 presents the main differences between mechanistic and hybrid models and illustrates their respective strengths and ideal use cases.

Table 2. Comparison of mechanistic and hybrid models for bioprocess control.

Aspect	Mechanistic Models	Hybrid Models
Foundation	Based on established scientific principles and equations	Combines scientific principles with empirical data
Adaptability	Limited adaptability to new conditions or variabilities	High adaptability, especially under variable conditions
Data Requirements	Lower; primarily relies on theoretical knowledge	High; requires real-time and historical data
Application Complexity	Works well for well-understood processes	Better suited for complex, dynamic processes
Accuracy	High in controlled, stable environments	High in dynamic, variable environments

In practical terms, hybrid models excel in predicting and controlling variables that fluctuate due to nonlinear dynamics or unexpected perturbations. For example, in large-scale fermentation, these models help forecast microbial growth, optimize nutrient delivery, and regulate temperature by utilizing both empirical data and scientific frameworks that capture cellular metabolism and reaction kinetics. This procedure increases control over

microbial behavior and assures that processes operate efficiently and consistently over different production scales, from research to industry. Therefore, hybrid models support both quality and yield optimizations, which are essential for sectors like pharmaceuticals and biomanufacturing.

These models are valuable in bioreactor environments, where the precise management of resources and conditions is critical. For example, in bioreactors, hybrid models combine data-driven algorithms with chemical modeling to predict reaction outputs and make adjustments to parameters like pH or oxygen levels, which directly impact product consistency. This combination reduces waste and energy demands by minimizing the need for ongoing manual recalibrations, offering significant advantages for sustainable production practices [10].

Hybrid models also achieve an effective balance between adaptability and precision. Considering that pure ML models may excel in flexibility but lack interpretability in complex environments, mechanistic models offer transparent and reproducible predictions but may be constrained by their assumptions when applied to highly variable environments. Hybrid models provide the strengths of each and allow bioprocesses to respond to evolving conditions without compromising accuracy or stability. This characteristic helps in applications such as therapeutic protein production, where variations in conditions could lead to expensive deviations from quality standards.

Another advantage of hybrid models lies in their ability to manage real-time monitoring and feedback loops. In bioprocess environments, rapid adjustments directly influence production outcomes. By combining empirical data with theoretical models, hybrid frameworks can respond dynamically, making adjustments based on immediate feedback. For instance, in the production of biologics, these models continuously optimize conditions such as nutrient flow and environmental factors, enhancing cell growth and protein yield while minimizing impurities. This integration of real-time adaptability with mechanistic consistency contributes to the delivery of high-quality outputs in advanced processes [10].

Hybrid models also facilitate the transition from small-scale research to large-scale production without losing control accuracy. As processes scale, hybrid models can incorporate feedback from extensive datasets, adjusting seamlessly to the added complexity and variability of industrial systems. This adaptability ensures that quality standards and operational efficiency remain intact, even in high-volume production environments, which is especially important in biomanufacturing where compliance with regulatory standards is critical.

Overall, hybrid models provide an adaptable, precise, and resilient approach to bioprocess control. By merging empirical data with established scientific knowledge, they enable industries to maintain consistency, efficiency, and sustainability in their operations, meeting the evolving demands of biomanufacturing and bioprocessing and positioning these sectors to leverage future technological advancements.

ML-based approaches face substantial implementation challenges, notably regarding data quality and model transparency. The 'black-box' nature of models, such as deep neural networks, poses ethical and regulatory concerns, particularly in regulated environments, such as pharmaceuticals, where clear interpretability and traceability of model decisions are mandatory for compliance. This is a significant issue especially for pharmaceutical production where ethical and regulatory considerations should be taken into account [36]. While these opaque AI models offer unprecedented potential to optimize these manufacturing processes, enhance quality control, and increase efficiency, their lack of transparency introduces risks in the form of accountability gaps. These gaps make it close to impossible to conduct a proper root cause analysis due to the inability to understand an AI's decision-making process after a failure, an important requirement of Good Manufacturing

Practice (GMP) [37]. Another risk is that these frameworks can learn and amplify hidden biases from historical data, leading to discriminatory or suboptimal outcomes, and pose direct threats to patient safety through unpredictable, systemic errors. To counter this, regulatory bodies, like the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA), are establishing risk-based frameworks that, while not banning black-box models, demand robust governance, stringent data integrity, comprehensive validation, and meaningful human oversight. The primary technical solution to this challenge is explainable AI [38], a set of methods, such as LIME and SHAP, designed to make AI decisions transparent and interpretable for regulators, allowing them to understand why an AI device made a specific prediction.

In order to address these challenges, the development of more accurate data collection methods, improved sensor technologies, and greater collaboration between biologists, engineers, and AI specialists are required. As these tools and collaborations evolve, ML has the potential to become a major component of bioprocessing, transforming the field with greater precision, adaptability, and efficiency.

4. Digital Twins and Soft-Sensing for Real-Time Monitoring

DTs are advanced virtual models that simulate the real-time behavior and dynamics of physical systems, which allow researchers and engineers to test control strategies, predict outcomes, and adjust parameters without direct interference in ongoing processes. In bioprocessing, DTs replicate environments like bioreactors, where conditions such as temperature, pH, and nutrient levels require precise control to support biological activities. This virtual representation allows for the safe evaluation of operational changes, facilitating system adjustments without disrupting ongoing production. As a result, DTs contribute to reduced development time, improved resource efficiency, and minimized risk in managing advanced bioprocesses [11].

The cycle presented below in Figure 3 presents the steps involved in a DT-based method for real-time bioprocess monitoring and optimization.

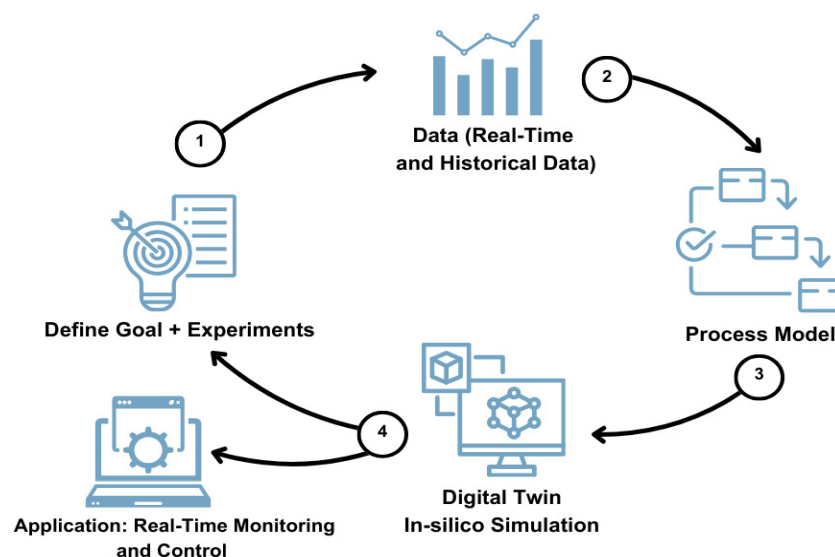


Figure 3. DT cycle for real-time bioprocess monitoring.

Step-by-Step Breakdown of the DT Cycle:

- **Define Goal and Experiments:** The initial step involves defining specific objectives, such as achieving a target yield, optimizing reaction conditions, or improving resource efficiency. Experiments are designed to gather critical data for the system, setting a

foundational understanding of process dynamics and requirements. This phase establishes the desired outcomes, quality standards, and any specific process parameters that need optimization [39].

- **Data Collection of Real-Time and Historical Data:** Once the goals are set, the system collects both historical and real-time data from various sensors embedded within the bioprocess. These data include critical parameters, like temperature, pH, oxygen levels, and nutrient concentrations. By combining historical data (providing baseline insights and trends) with real-time information, the DT has a comprehensive view of current conditions, enabling accurate analysis and predictions [40,41].
- **Process Model Development:** This phase involves creating a detailed process model that integrates both mechanistic (physics-based) and data-driven components. Mechanistic models rely on scientific principles to describe the bioprocess, while data-driven models (e.g., ML algorithms) leverage empirical data to identify patterns and adjust predictions. Together, these models form a hybrid approach that balances accuracy and adaptability, allowing the system to capture complex dynamics that may arise in the bioprocess [42].
- **DT In Silico Simulation:** In this stage, the DT simulates various scenarios based on real-time data inputs and the established process model. This simulation environment allows operators to test adjustments, such as modifying temperature or nutrient levels, and observe the projected outcomes without impacting the actual process. Using these ‘what-if’ scenarios, the DT identifies the best strategies to optimize process conditions [43,44].
- **Application: Real-Time Monitoring and Control:** The insights generated by the DT’s simulations are applied to the live bioprocess. Through real-time monitoring and feedback loops, the DT communicates with the physical system to make precise adjustments, maintaining optimal conditions for yield and quality. This real-time control capability, enhanced by AI, ensures continuous alignment with target parameters and adapts swiftly to any fluctuations or disturbances in the process [45,46].

This iterative cycle of data collection, modeling, simulation, and real-time control forms a solid framework to optimize bioprocess. For example, a DT model of a bioreactor simulates the result of alterations in temperature or oxygen levels on microbial growth [47]. Engineers use the model to test and adjust process parameters virtually, reducing expenses by avoiding direct trial and error in live systems. DTs usually rely on a synergistic combination of mechanistic modeling and machine learning, often integrating real-time data and historical insights to continuously adapt to and predict system behavior. These features make DTs valuable in dynamic and variable bioprocess environments, like fermentation and cell culture.

When combined with soft sensors and AI-based inference, DTs offer substantial benefits for applications requiring tight control, such as cell culture and microbial fermentation. The ability of AI to detect subtle patterns and forecast necessary adjustments enables a more reliable and efficient process, with fewer interruptions for manual adjustments. In bioprocessing, the integration of AI with DTs and soft sensors forms a robust framework that ensures product quality, optimizes resource usage, and improves adaptability across a range of operations.

4.1. Soft-Sensing: Virtual Sensors for Real-Time Estimation

Soft sensing, also known as virtual sensing, refers to the estimation of critical process variables that are difficult, expensive, or time-intensive to measure directly. Unlike traditional sensors that capture physical data, soft sensors rely on mathematical models and algorithms to infer values from accessible data. In bioprocessing, where precise conditions

are necessary to maintain product quality, soft sensors enable the continuous monitoring of key variables that may otherwise require invasive sampling or laboratory analysis, such as biomass concentration and product levels [34].

Soft-sensing frameworks provide reliable, real-time estimates of hard-to-measure variables by using more accessible indicators, like dissolved oxygen, pH, and temperature. For example, in fermentation processes, soft sensors estimate biomass concentration by analyzing readily available parameters, such as temperature, pH, and dissolved oxygen. This capability allows operators to maintain bioprocesses within ideal ranges without frequent interruptions.

In a bioprocessing environment, it is important to maintain optimal conditions for consistent quality and yield. Soft-sensing frameworks represent an adaptable and efficient solution for real-time monitoring, especially in scenarios where a direct measurement is impractical or expensive. Figure 4 illustrates how these systems function by integrating a combination of hardware sensors and computational estimators [34].

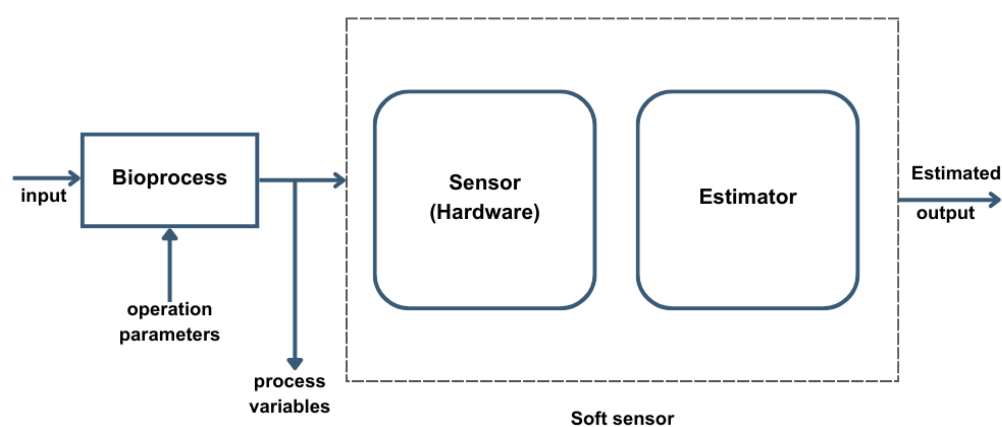


Figure 4. Soft-sensing framework for real-time bioprocess monitoring.

Components of the Soft-Sensing Framework:

- **Input to the Bioprocess:** The initial input represents operational parameters, such as feed rates or environmental adjustments (e.g., temperature, pH), which are required to sustain optimal process conditions. These parameters are set based on desired output and quality standards and can be adjusted dynamically based on feedback from the system [48].
- **Bioprocess Module:** The bioprocess module receives inputs and operates under controlled conditions that influence critical process variables, such as microbial growth or enzyme activity. These variables, which usually are difficult to measure directly, can have a major impact on the final product quality. The bioprocess module captures both the operational parameters and output variables, passing them to the soft-sensing system for further analysis.
- **Soft Sensor:**
 - **Sensor (Hardware):** Traditional hardware sensors measure accessible variables, like temperature, dissolved oxygen, or pH, which serve as indicators for other critical but less accessible process variables.
 - **Estimator:** The estimator combines mathematical models and algorithms to infer values for complex process variables (e.g., biomass concentration, product concentration). The estimator can predict these hard-to-measure variables accurately as it analyzes accessible sensor data.

- **Estimated Output:** The soft sensor's output provides real-time estimates of critical process variables. These estimates allow operators to maintain the bioprocess within target ranges, without needing invasive sampling or laboratory analysis [49].

Soft sensing enables the continuous monitoring of process conditions, which is necessary for adaptive control. By converting accessible measurements into estimated values for the main variables, soft sensors reduce the need for manual intervention. For example, in fermentation, a soft sensor might predict biomass concentration by correlating dissolved oxygen levels and pH changes with microbial growth rates.

To achieve precise and consistent results, soft sensing follows a structured process with several key steps.

Figure 5 represents the process involved in developing a soft-sensing framework for bioprocess monitoring. It begins with data collection and pre-processing, where historical and real-time data are gathered and prepared through actions such as addressing missing values, reducing noise, and standardizing the dataset for quality. The next phase focuses on variable selection, identifying the most relevant parameters—like temperature or pH—that correlate closely with the target variable, ensuring the soft sensor's simplicity and precision. Following this, model training and selection are conducted, where engineers choose and refine a model, ranging from linear equations to advanced ML algorithms, based on the specific needs of the bioprocess. Finally, the trained model undergoes validation and evaluation, ensuring that its predictions are accurate and reliable during actual operations. This structured approach facilitates the creation of adaptive and efficient soft sensors, as visualized in the diagram [34].



Figure 5. Soft-sensing framework for real-time bioprocess monitoring.

Soft sensors generally fall into two categories: mechanistic models, which rely on scientific principles and equations, and data-driven models, which learn from historical data. Mechanistic models are stable but may not adapt easily to new conditions, while data-driven models adapt more flexibly but depend on large, high-quality datasets. Hybrid models blend these approaches, creating reliable and adaptive real-time estimations for dynamic conditions.

4.2. AI Integration with Digital Twins and Soft Sensors in Bioprocesses

AI extends the capabilities of DTs and soft sensors, which improves their ability to monitor, predict, and adjust bioprocesses in real time. AI algorithms process large datasets and make real-time adjustments, which enables bioprocessing systems to respond rapidly to fluctuations.

An AI-powered DT uses both past and real-time data to predict when changes in oxygen levels are necessary to sustain ideal growth conditions. Article [11] explains how AI algorithms, such as RL, help DTs learn from past data, which improves predictions and control strategies as conditions change. AI also enhances soft sensors, which allows them to interpret large datasets and recognize subtle patterns, increasing their accuracy.

AI-powered DTs and soft sensors provide multiple benefits in applications like cell culture and microbial fermentation, where stable conditions facilitate reliable, high-quality results. With the ability to forecast changes and provide predictive insights, these tools help maintain consistent process control. In bioprocessing, the integration of DTs, soft

sensors, and AI forms a strong foundation to ensure product quality and efficiency across complex operations.

4.3. *Applications of Digital Twins and Soft Sensors in Bioprocess Optimization*

4.3.1. Bioreactors

Bioreactors grow cells or microorganisms to produce pharmaceuticals, biofuels, and other valuable products, illustrating how DTs and soft sensors work together. Operators control environmental factors inside a bioreactor to maximize cell growth and product yield. A DT mirrors the actual bioreactor, allowing operators to test operational adjustments virtually—such as nutrient adjustments, aeration, or pH—in a virtual environment. This approach enables operators to observe the effects of adjustments before applying them to the real bioreactor, minimizing risks.

Soft sensors estimate challenging variables, such as biomass and product levels, by analyzing accessible data, like dissolved oxygen and pH. For example, a soft sensor uses oxygen and pH levels to estimate biomass and sends this datum to the DT, which then refines its predictions. When the soft sensor detects a drop in oxygen, the DT simulates different aeration levels and guides operators toward the best corrective action [50].

This configuration establishes a closed feedback loop: soft sensors provide real-time data, while the DT continuously evaluates system dynamics and proposes optimized adjustments. This system gives operators a clear view of the bioprocess and enables them to maintain optimal conditions. This article highlights that this combination is particularly beneficial in biopharmaceuticals, where product quality and yield directly affect consistency and effectiveness.

4.3.2. Cell Culture in Biopharmaceutical Production

In the production of therapeutic proteins, mammalian and microbial cell cultures are widely used, and their growth conditions must be stringently controlled. Even minor deviations in nutrient supply or pH can significantly impact product yield and bioactivity. DTs, powered by AI algorithms, help monitor and predict when adjustments are needed to maintain ideal conditions, such as nutrient replenishment or pH correction.

Soft sensors advance the estimation of parameters that are difficult to measure directly, such as cell density and nutrient consumption, from accessible data, like dissolved carbon dioxide and oxygen levels. For example, when the soft sensor detects a decline in nutrient levels, the DT simulates different replenishment rates and suggests the most effective approach to maintain cell health and productivity. This setup forms a feedback loop, where the DT suggests actions and the soft sensor provides live data on cell culture status, ensuring consistent and high-quality protein yields [34].

4.3.3. Anaerobic Digestion for Biofuel Production

Anaerobic digestion is a widely adopted method for producing biogas from organic waste streams, including agricultural residues and food waste. Maintaining optimal conditions, such as temperature and pH, is essential for maximizing biogas yield and efficiency. An AI-enhanced DT models the anaerobic digestion process, predicting microbial responses to changes in feed rate or environmental factors and suggesting adjustments to maximize biogas production.

Soft sensors estimate difficult-to-measure factors, such as volatile fatty acids (VFAs) and ammonia concentrations, which are critical for maintaining microbial health but challenging to measure in real time. When the soft sensor detects an increase in VFAs, the DT simulates the effects of different pH adjustments and feed rates, guiding operators to the optimal action. This feedback loop allows operators to dynamically control anaerobic digestion conditions, improving biogas yield and process stability [34].

5. AI in Biocatalysis and Enzyme Optimization

ML introduces major advancements to enzyme screening and optimization, providing precise predictions for enzyme properties, such as stability, activity, and compatibility, with specific reactions. Traditional enzyme optimization often requires extensive experimental trials to identify enzymes with desired properties, a process that can be both time and financially intensive. ML models accelerate this process, making enzyme selection and optimization faster and more targeted.

A range of ML models demonstrates the potential of data-driven methods in enzyme screening. Supervised ML models, such as regression algorithms and neural networks, apply datasets of enzyme structures and experimentally measured properties to predict characteristics like thermostability, pH tolerance, and substrate specificity. In biofuel production, for instance, ML models select enzymes that perform consistently at high temperatures and varying pH levels, offering durability in industrial environments where enzyme performance directly affects productivity and cost-effectiveness [51].

In enzyme optimization, predictive ML models—including decision trees and random forests—propose structural modifications that enhance enzyme function. These models analyze enzyme sequence data and known properties to identify specific adjustments that improve reaction speed, stability, or shelf life. Article [51] explains how AI-driven predictive models streamline enzyme optimization, suggesting amino acid substitutions that increase catalytic efficiency and bioconversion rates, critical for large-scale biopharmaceutical and industrial applications.

Additionally, hybrid models combining ML with established scientific principles offer further advantages in enzyme optimization. This integrated approach merges ML-driven predictions with foundational biochemical knowledge to maintain accuracy across diverse conditions. In applications, such as food processing, hybrid models support enzyme stability and adaptability under varying conditions, ensuring consistent results even when environmental factors fluctuate.

5.1. AI's Role in Optimizing Biosynthetic Pathways

AI contributes not only to individual enzyme optimization but also to the design and improvement of entire biosynthetic pathways, improving overall efficiency and yield. These pathways, which consist of sequences of enzyme-catalyzed reactions, convert substrates into valuable biochemical products. Using insights gathered with AI, biorefineries boost the efficiency and yield of biosynthetic pathways, which supports high-output and more affordable production.

5.1.1. Pathway Prediction and Optimization in Metabolic Engineering

AI-based pathway prediction tools analyze large datasets of metabolic reactions to find smooth routes for producing target compounds. Study [51] explains how AI models simulate pathways and identify routes that improve productivity while limiting by-products. This process is valuable in pharmaceutical production, where high purity and careful resource use are priorities. With these models, researchers can discover the best pathways to synthesize high-value chemicals, such as antibiotics, improving yield and reducing the need for extensive purification.

5.1.2. Optimization of Flux in Biosynthetic Pathways Using ML

ML models assist in identifying optimal control strategies for metabolite distribution in biosynthetic pathways. Algorithms, such as flux balance analysis, simulate metabolite flow within these pathways, identify bottlenecks, and reveal areas of resource waste. With these insights, operators adjust pathway parameters to increase yield and reduce resource use.

The research on biofuel production demonstrates this, where balanced substrate availability and controlled by-product formation are essential for streamlined production [52].

5.1.3. Hybrid Models for Managing Complex Pathways

For pathways that involve complex regulatory mechanisms and feedback loops, hybrid models—combining ML predictions with biochemical principles—assist in addressing these complexities. These models integrate ML insights with foundational principles, such as mass balance and reaction kinetics, ensuring a stable performance across varying conditions. Studies show that these hybrid models enable pathways to operate consistently, even in fluctuating environments. For example, in amino acid production, pathway stability is essential to ensure product quality, underscoring the role of hybrid models in managing intricate and condition-sensitive processes [34].

6. AI for Quality by Design Optimization

Quality by Design (QbD) principles in biomanufacturing provide a systematic scheme that ensures product quality through a proactive approach that controls the attributes of the product from the initial stages. The integration of ML into QbD increases the ability to identify and control critical process parameters (CPPs) that impact critical quality attributes (CQAs) directly. ML's capacity to process large, high-dimensional datasets provides deeper insights into biomanufacturing variables. This reduces dependency on manual testing and it also increases precision. ML not only improves the reliability of QbD methods, but also allows manufacturers to address instability in real time by isolating influential parameters [53].

The application of AI within QbD frameworks in biomanufacturing enhances control over CQAs by enabling adaptive, data-driven adjustments throughout the production process. Figure 6 offers an overview of an AI-driven QbD workflow and illustrates how AI technologies support continuous monitoring, predictive modeling, and optimization to ensure consistent product quality.

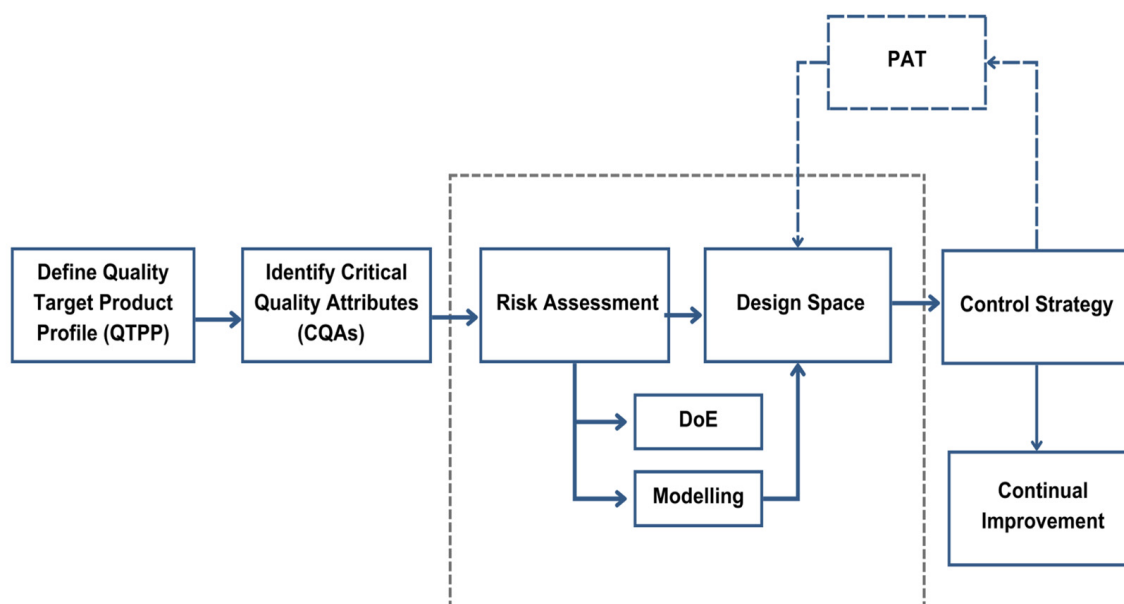


Figure 6. AI-driven QbD workflow for consistent quality in bioprocessing.

This framework serves as the foundation for the upcoming content, where we explore specific applications of AI in identifying critical parameters, hybrid modeling for process control, and real-time quality adjustments in bioprocessing.

6.1. ML for Critical Parameter Detection

ML identifies CPPs in scenarios with multiple interacting variables, which impact the quality of the outcomes. Traditional QbD methods use statistical tools to analyze changes in parameters, such as pH, temperature, and nutrient levels, and their effects on CQAs. However, these methods usually miss important patterns in complex datasets, specifically when interactions are nonlinear or when parameters show dependencies.

Study [53] demonstrates that ML improves QbD by uncovering patterns and dependencies within high-dimensional datasets. Supervised ML models, like regression and neural networks, analyze historical and real-time production data to determine which CPP combinations influence CQAs most directly. In therapeutic protein production, for instance, ML models identify how specific temperature and pH ranges affect protein folding, stability, and purity factors necessary for meeting industry standards. ML enables operators to set precise parameter ranges that reduce batch variability as it isolates these CPPs.

Unsupervised learning methods, such as clustering algorithms, reveal data relationships and do not require predefined categories. During early development, clustering identifies optimal parameter combinations that consistently produce high-quality outputs. This observation allows manufacturers to refine initial process designs and minimize errors in later production stages. For example, clustering across production runs reveals parameter combinations associated with CQAs, enabling early-stage control of product quality.

6.2. Hybrid QbD Models

In complex bioprocessing, where factors like temperature and pH actively interact, ML alone may struggle to maintain precise control over variable conditions. To address this, hybrid QbD models integrate ML insights with the existing scientific principles, combining the flexibility of data-driven models with the stability of mechanistic frameworks. Design of Experiments (DoE) contributes to this setup, as it allows for the systematic exploration of variable interactions, helping to define and optimize the design space for consistent process outcomes. As a result, hybrid models provide a structured way to manage variability by utilizing both empirical data and fundamental principles to monitor CPPs.

Study [53] outlines how hybrid QbD models apply both ML-driven predictions and scientific rules, ensuring that the resulting model remains consistent across different conditions. In biocatalytic processes, hybrid models use ML insights into real-time data on enzyme activity while referring to biochemical principles to predict enzyme–substrate interactions. This dual-layered modeling, supported by DoE, refines control over parameters affecting enzyme function, a crucial element in applications with high sensitivity to environmental changes, such as therapeutic protein production [49].

In therapeutic protein production, hybrid QbD models link ML predictions on variables like growth rate and nutrient utilization with kinetic equations governing cellular metabolism. This approach, enhanced by DoE, improves the model's robustness, allowing it to adjust nutrient feed rates accurately based on feedback from real-time data and established biological principles. Through this hybrid framework, operators achieve more consistent yields and quality in therapeutic protein manufacturing by adjusting critical inputs dynamically, keeping the process within defined quality ranges.

6.3. Adaptive Control with Hybrid Models for Real-Time Quality Adjustments

Hybrid QbD models enable adaptive control systems that react to bioprocess fluctuations in real-time. Analyzing the current data and forecasting parameter shifts, these models allow operators to make adjustments that keep processes within optimal ranges, reducing reliance on manual recalibration. Process Analytical Technology (PAT) integrates with adaptive systems by providing real-time insights into critical process parameters

(CPPs), facilitating continuous monitoring and timely intervention. Integrated with AI, PAT enhances the accuracy and responsiveness of hybrid QbD models, allowing for a proactive approach in maintaining CQAs.

The following diagram presented in Figure 7 illustrates the concept of AI-driven process control within a QbD framework. In this setup, AI continuously monitors and adjusts CPPs to maintain stability in CQAs, ensuring consistent product quality and quantity.



Figure 7. AI-driven quality control in bioprocessing.

MPC, widely used within hybrid QbD frameworks, predicts parameter shifts and adjusts settings preemptively to maintain process stability. MPC uses predictive models to guide decision-making based on anticipated changes, rather than waiting to react. In microbial fermentation, for instance, MPC monitors shifts in nutrient and oxygen availability, adjusting feed rates to balance these factors. A recent study [53] illustrates how MPC models stabilize growth rates and biomass yield, supporting consistent quality.

Integrating AI with QbD models enhances automated, scalable quality control, enabling continuous, data-based adjustments. In biofuel production, adaptive feedback from hybrid models manages pH and oxygen levels in anaerobic digestion, promoting stable methane production. This approach minimizes manual interventions and improves reliability in process outcomes.

Hybrid QbD models adapt biomanufacturing processes from laboratory to industrial scales, handling variations in inputs and conditions. Article [54] highlights how hybrid models manage these shifts by integrating real-time data with scientific principles, essential in biopharmaceuticals, where regulatory compliance depends on tight quality control.

Scalability challenges—such as temperature shifts, nutrient gradients, and changes in reactor size—can disrupt quality, but hybrid models mitigate these effects by combining empirical data and scientific foundations. This dual approach enables a controlled scale-up, keeping quality consistent as production volume increases. By reducing recalibration and manual adjustments, hybrid QbD models decrease downtime, which supports efficient, sustainable production.

7. Sustainable and Scalable AI Strategies

Bioprocesses are increasingly being used, with applications in different fields, such as the pharmaceutical industry, agriculture, and energy. As their utilization continues to grow, the challenge of scaling them from the laboratory to the industrial levels arises. AI plays an important role in solving this problem. With the help of this technology, various strategies have been developed that manage to overcome these obstacles [55].

One of the main strategies involves combining different modeling approaches—mechanistic, data-driven, and hybrid models—to optimize and scale bioprocesses. Table 3 below provides a detailed comparison of these approaches, and it highlights their founda-

tions, strengths, challenges, and typical applications. As seen in Table 3, hybrid models stand out for their ability to balance precision and adaptability, making them particularly effective for dynamic bioprocess scaling [35].

Table 3. Comparison of mechanistic, data-driven, and hybrid models.

Aspect	Mechanistic Models	Data-Driven Models	Hybrid Models
Foundation	Based on scientific laws, such as kinetics	Derived from empirical data using ML	Combines scientific laws with empirical data
Strengths	High precision in describing stable systems	Identifies patterns in large datasets	Balances precision and adaptability
Challenges	Limited adaptability to dynamic conditions	Requires high-quality, large datasets	More complex to design and implement
Applications	Fermentation optimization, reaction kinetics	Pattern recognition, error detection	Scaling processes, dynamic control

In addition to modeling, real-time monitoring is another essential AI-driven strategy for scaling bioprocesses. This is achieved by using AI-based sensors and control systems that continuously analyze the process, maintaining the necessary conditions. The use of sensors leads to increased process efficiency, as continuous analysis helps eliminate errors and adjust parameters according to optimal conditions. Article [50] presents the advantages of using real-time monitoring to improve the efficiency of bioprocesses (making adjustments, obtaining information about optimization scenarios, which reduces resource consumption and time-consumption). Thus, bioprocesses can be scaled more easily, with problems being detected and resolved on time.

Forward reaction prediction is another efficient strategy for scaling up bioprocesses. It helps predict the resources needed and the optimal conditions. This aids in reducing time and resource consumption by minimizing the number of experimental tests. Article [56] describes the advantages of using these models and their applications. This strategy, through process optimization, resource consumption reduction, and adaptability to different conditions, overcomes the obstacles related to scaling processes.

The use of DTs significantly improves bioprocesses, as they offer advantages such as error reduction and achieving optimal results. According to this article [54], DTs are detailed, virtual representations of production systems. They provide feedback between the virtual model and physical systems. By using them, bioprocesses become more efficient, enabling the simulation and prediction of results. This facilitates the scaling of bioprocesses by effectively solving problems that could obstruct it.

The use of AI in bioprocesses will become common, because it helps optimize processes and increase precision, leading to a higher demand for bioprocesses that have a higher level of performance. While the use of AI in laboratory-scale bioprocesses can deliver optimal results, scaling up to industrial levels requires increased attention to maintain stability and optimize processes. Industrial bioprocesses rely on data from dynamic environments, so we must find solutions to overcome these challenges.

One of these solutions is the use of MPC, which predicts the behavior of a system to provide optimal and precise results while reducing time and resource consumption. MPC is based on three components: a model that predicts the system's outcomes, an objective function that defines the optimization goals, and a control law that guides system

adjustments. This study [49] highlights MPC's ability to optimize calculations while considering constraints, making it invaluable in dynamic environments.

Another solution is the use of fuzzy logic controllers. Fuzzy logic is capable of managing uncertainties, improving performance, and detecting errors. Because of the advantages it offers, fuzzy logic is applied in various fields, such as medicine, economics, pharmaceuticals, etc. When it comes to scaling bioprocesses, fuzzy logic is used to maintain stability and achieve optimal results. Its strength lies in handling nonlinear and variable operations in dynamic environments through simplified rule-based decision-making. As noted in [10], fuzzy logic was applied to processes such as fermentation, where conventional control methods may fall short. By addressing uncertainty and imprecision in process data, fuzzy logic contributes to more stable and accurate control outcomes.

Together, these AI-driven strategies, real-time monitoring, forward reaction prediction, DTs, MPC, and fuzzy logic, form a cohesive framework for scaling bioprocesses sustainably and efficiently. By optimizing resource use, minimizing experiments, and maintaining stability, these approaches enable bioprocesses to meet the growing demands of industrial applications with consistent quality and adaptability.

8. Analysis, Challenges, and Future Prospects

8.1. Analysis

A comparative analysis of the referenced literature detailed in Table 4 reveals that ANNs are, by a significant margin, the most prevalent AI method used in biorefineries and bioprocessing. Their large number of applications display their versatility and success across a wide spectrum of use cases, from modeling and optimization to anomaly detection or phenotyping. This makes them a foundational and multipurpose tool for enhancing bioprocess efficiency and understanding. Following ANNs, several other ML techniques have been used in the literature. SVM, fuzzy logic, and other evolutive methods are frequently used for modeling and optimization.

Table 4. Comparative analysis of the referenced literature.

AI Method	Application Domain	References
ANN	Modeling and optimization, Spectroscopy, Soft sensors, Estimation, Bioprocess control, Anomaly detection, Phenotyping, Forecasting	[4,5,10,12,13,17–19,21,22,25,26,28,29,31,34,35,41,48,49,51–54,56]
SVM	Modeling and optimization, Anomaly detection, Soft sensors	[4,23,25,34,52,54]
Fuzzy Logic	Modeling and optimization, Bioprocess control, Soft sensors	[4,10,34]
GA	Modeling and optimization, Soft sensors	[4,34]
PSO	Modeling and optimization	[4]
Hybrid Models	Computational fluid dynamics, Control strategies, Fault detection, Process design	[9,10,26,27,33,39,42,44,54]
MPC	Control strategies	[10,35]
Partial Least Squares	Soft sensor	[24]
RL	Process control	[30,32,33,52,54]

A growing trend in the field is the application of hybrid models and advanced control-oriented strategies to tackle more complex, integrated problems. Hybrid models, which combine different AI techniques or integrate them with mechanistic models, are being used for specialized tasks, such as computational fluid dynamics simulations. At the same time, methods like MPC or RL are being implemented specifically for advanced process control. The emergence of RL, in particular, points toward a future of more autonomous and adaptive bioprocesses capable of optimizing performance in dynamic and uncertain environments.

AI simplifies operations in bioprocesses and brings improvements. To ensure these goals are achieved, certain conditions must be followed when implementing AI-based models. One of these conditions is creating interpretable AI models. Interpretability provides clarity to the model. When models are clear, detecting and fixing errors becomes easier. In contrast, black-box models are difficult to improve or diagnose when problems arise.

The true competitive advantage of RL in this domain lies not merely in its ability to optimize a well-defined system, but in its inherent capacity to learn robust control policies in the face of the defining challenges of bioprocessing: uncertainty, model–plant mismatch, and sparse data. A traditional PID controller is reactive and performs best with frequent, clean data, which are often not available. MPC, a more advanced technique, is proactive, but its performance is critically dependent on the accuracy of an underlying process model, which is often difficult to obtain for complex biological systems. This suggests that RL is not just another algorithm in the process control toolkit; it is a framework uniquely suited to the noisy, uncertain, and often unpredictable reality of biomanufacturing. Compared to supervised learning and unsupervised learning, for bioprocess control, RL represents a paradigm shift from passive prediction to active, goal-directed decision-making that enables true process autonomy. This represents a higher level of abstraction, moving from the question “What will happen next?” to the more powerful question, “What should I do next?”. While its industrial adoption is still nascent due to significant implementation challenges, the analyzed articles prove its unique potential. The future development of structured sim-to-real transfer frameworks is a critical innovation that provides a pragmatic and safe pathway for deploying these advanced controllers in real-world manufacturing settings. The ongoing research should focus on safety, robustness, and data efficiency.

Furthermore, interpretability is crucial, especially when the model is used in pharmaceutical manufacturing. Models must follow strict rules to gain approval from regulatory agencies, such as the FDA or the EMA, for transitioning to commercial production [54]. Only clear models that explain how decisions are made can receive these approvals.

Another benefit of interpretable AI models is that they can be easily understood by the scientists who are using them. If the model includes unclear operations, users might not be able to fully understand how the model works, which could lead to incorrect use. These benefits highlight the importance of developing interpretable AI models, ensuring significant advantages in their application.

As mentioned above, AI is increasingly being used in bioprocesses. These are complex and dynamic processes that require high-quality data to successfully achieve real-time predictions, process optimizations, and other improvements. Thus, as a solution to this need for high-quality data, interdisciplinary collaboration between bioprocess engineers, biologists, and AI experts is essential.

Through such a collaboration, the necessary data can be obtained and used in the developed models. Moreover, an interdisciplinary collaboration is essential for developing comprehensive systems, as each specialist contributes domain-specific expertise that ensures all aspects of the model are accurately represented. An issue solved through collaborative work is the efficient discovery of optimal solutions. When multiple specialists

from different domains work together, their knowledge and skills are combined, resulting in more effective outcomes.

Article [2] presents AI applications in fields such as medicine, agriculture, and bio-based industry. In medicine, AI is employed for drug development, disease detection, and diagnosis. In agriculture, AI models can predict water and fertilizer requirements, yield predictions, and optimal conditions. In the bio-based industry, AI is used to improve bioprocesses by finding optimal conditions, predicting outcomes, and optimizing resource usage. In each of these domains, the integration of AI was made possible through teamwork. Specialists from various fields collaborated to develop models that bring significant performance improvements in their respective industries.

8.2. Challenges

The use of AI in bioprocesses has brought substantial improvements, such as process optimization, result prediction, reduced consumption of time and resources, and others. However, to fully achieve the benefits AI offers, it is important to consider the potential obstacles that may arise. These obstacles are, in most cases, caused by the data we provide. AI-based models rely on data, and if the data are incorrect or insufficient, we will not be able to achieve the desired optimizations. Therefore, when integrating AI-based systems, it is essential to ensure that high-quality data are provided.

AI applications in bioprocesses are bottlenecked by a lack of data quality and insufficient data quantity. Bioprocesses are complex and dynamic, and for AI software to make reliable decisions, a considerable amount of quality data is essential. One solution to meet the need for data is using hybrid models instead of data-based models. By using these, the amount of data required is reduced, as the model also relies on mechanistic models that use scientific equations. The development of sensor technologies that improve the quality and quantity of data is another solution to this problem. Soft sensors—mathematical models that estimate unmeasured variables based on correlated process data—are used to make real-time predictions of a system. They can provide information about the variables and data used in the process [54]. As a solution to issues caused by incomplete or low-quality data, soft sensors have been implemented. They monitor the process variables that cannot be measured reliably or at all. This allows timely intervention to solve problems, ensuring that the final results are accurate.

To provide a clearer understanding of the consequences associated with insufficient or low-quality data, article [2] presents examples in which AI models failed to perform as intended due to data-related constraints, resulting in incorrect outcomes. The first example involves a model for analyzing dental radiographs. Due to a lack of clinical data, the model was unable to provide accurate analysis, and the predictions it made were incorrect. Another example presented in the article is an instrument for analyzing facial dysmorphism. It failed to identify people with Down syndrome who were of African heritage, while it successfully identified people who were of European heritage. After incorporating data specific to the category where the predictions were poor, the instrument was able to achieve its goal. These examples clearly show that a lack of data or poor data quality leads to model dysfunction.

8.3. Future Prospects for Industrial-Use Cases

From an applied perspective, the authors have taken this research to the next level. The methods and findings from this material are currently being applied for two industry-level initiatives:

- SUSHEAT [57,58]—heat supply for high-heat-demand industrial processes, like dairy products, fish oil, pulp and paper, beverages, petrochemical, textile and leather, and basic metals.
- BIODT—industrial system for commercial development—biodegradable-packaging materials—biodegradable plastics for food, drinks, cosmetics, and pharmaceuticals.

8.3.1. SUSHEAT—Smart Integration of Waste and Renewable Energy for Sustainable Heat Upgrade in the Industry

The SUSHEAT project [57,58] exemplifies emerging strategies for tackling critical energy demands. A significant future challenge for biorefineries lies in sustainably sourcing process heat, particularly in the 150–250 °C range required for operations like sterilization or drying, a range where conventional decarbonization technologies often face limitations. SUSHEAT directly confronts this challenge by developing an integrated system featuring a high-temperature heat pump [59], bio-inspired thermal storage [60], and crucially for future optimization strategies, an AI-driven Control & Integration DT (CIDT) [61,62]. This CIDT represents a key future prospect, utilizing AI algorithms and DT concepts to navigate the complexity of integrating variable inputs, like waste heat and renewable energy, with advanced storage and conversion technologies. The sophisticated AI-based optimization and control methodologies being validated within SUSHEAT offer a glimpse into future possibilities for AI applications in bioprocessing, not only for optimizing the energy supply infrastructure but potentially providing transferable models for enhancing the efficiency, flexibility, and overall sustainability of integrated biorefinery operations themselves.

8.3.2. BIODT—Industrial System for Commercial Development

The industrial system for commercial development aims to develop, optimize, and demonstrate safe, sustainable, and economically feasible (bio)conversion routes for low-cost food wastes and residual biomass toward the production of the next generation of completely biodegradable biopolymers and composite materials. The data-driven modeling focuses on three main activities: conversion of information (process, measurements, and outcomes) to AI data structures, mapping of data structures to algorithms, and visualizing and fast changing of properties in order to enhance solutions. This system is currently in the prototyping phase and will be further described in future studies.

8.3.3. Other Relevant Use-Cases

Besides the two industrial use cases detailed above as an ongoing activity from the authors, the tangible impact of AI in biorefineries and bioprocessing is very advanced and has already produced results [63–65] in the following use cases:

- Enhancing biofuel production (fine-tuning for maximum yield in biodiesel and biofuel from algae)—benchmarks reveal yield improvements of over 80%.
- Fermentation process control (real-time analysis of concentration to allow dynamic adjustments to fermentation conditions)—benchmarks reveal values for coefficient of determination in the range of 0.92–0.95—very close to the maximum (1).
- Microbial strain engineering (algorithms can analyze vast genomic datasets to identify gene-editing targets that could enhance the production of desired compounds in microorganisms)—benchmarks are still a work in progress.
- Raw material selection (AI models are built to analyze the chemical and physical properties of biomass feedstocks to predict potential biochemical yield. This enables biorefineries to select promising and cost-effective raw materials)—benchmarks reveal predictions with accuracy rates over 95%.

9. Conclusions

This review analyzed the role of artificial intelligence (AI) and ML in enhancing biorefinery and bioprocessing systems, with a particular focus on their ability to improve production efficiency and enable advanced system-level control. These technologies effectively manage complex workflows, optimize resource allocation, and maintain operational stability under fluctuating process conditions. AI-based models facilitate accurate forecasting and real-time process adjustments, ensuring consistency, reliability, and improved decision-making across the production chain.

Tools such as soft sensors and DTs facilitate the precise control of complex bioprocess operations. Soft-sensing techniques enable non-invasive, real-time estimations of process variables that are otherwise difficult or costly to measure directly. DTs create virtual representations of physical systems, allowing operators to simulate modifications and evaluate their impact before implementation. Additionally, hybrid modeling approaches—integrating empirical data with mechanistic principles—offer a balanced combination of accuracy and flexibility, ensuring reliable and stable system performance across diverse operational scenarios.

Despite these advances, multiple challenges persist—including data quality limitations, model interpretability, and the need for interdisciplinary collaboration. Addressing these barriers requires a combination of rigorous data handling, transparent modeling approaches, and collaboration across disciplines. Hybrid methods, which merge computational insights with scientific understanding, provide solutions that can accommodate the variability inherent in scaling processes from laboratory to industrial levels.

AI-based systems respond to the operational demands of contemporary bioprocessing by improving process control, minimizing variability, and supporting efficient use of resources. These developments correspond to the broader shift toward scalable production models that meet environmental and industrial performance targets, reinforcing the role of biomanufacturing in addressing the current economic and ecological priorities.

Author Contributions: Conceptualization, A.B., I.C., R.B., J.E. and A.M.; methodology, A.B., I.C., and R.B.; formal analysis, I.C.; investigation, A.B., I.C., R.B. and A.M.; data curation, A.B., J.E. and I.C.; writing—original draft preparation I.C., A.B. and A.M.; writing—review and editing, A.B., I.C., R.B. and A.M.; visualization, I.C.; supervision, A.B.; project administration, A.B. and R.B.; funding acquisition, A.B., J.E. and R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study received funding from the European Union—Grant Agreement No 101103552—SUSHEAT—Smart Integration of Waste and Renewable Energy for Sustainable Heat Upgrade in the Industry. The views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Alex Butean, Iulia Cutean and Alex Matei were contracted by the company Wiz Development and Services; Juan Enriquez was contracted by the company Dynamic & Security Computations SL. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. The European Green Deal—European Commission. Available online: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en (accessed on 15 April 2025).
2. Bhardwaj, A.; Kishore, S.; Pandey, D.K. Artificial Intelligence in Biological Sciences. *Life* **2022**, *12*, 1430. [CrossRef] [PubMed]
3. Schipfer, F.; Burli, P.; Fritsche, U.; Hennig, C.; Stricker, F.; Wirth, M.; Proskurina, S.; Serna-Loaiza, S. The Circular Bioeconomy: A Driver for System Integration. *Energ. Sustain. Soc.* **2024**, *14*, 34. [CrossRef]
4. Cheng, Y.; Bi, X.; Xu, Y.; Liu, Y.; Li, J.; Du, G.; Lv, X.; Liu, L. Artificial Intelligence Technologies in Bioprocess: Opportunities and Challenges. *Bioresour. Technol.* **2023**, *369*, 128451. [CrossRef] [PubMed]
5. Suwardi, A.; Wang, F.; Xue, K.; Han, M.; Teo, P.; Wang, P.; Wang, S.; Liu, Y.; Ye, E.; Li, Z.; et al. Machine Learning-Driven Biomaterials Evolution. *Adv. Mater.* **2022**, *34*, 2102703. [CrossRef]
6. Böhner, F.D.; Prado-Rubio, O.A.; Huusom, J.K. Challenges in Optimization and Control of Biobased Process Systems: An Industrial-Academic Perspective. *Ind. Eng. Chem. Res.* **2021**, *60*, 14985–15003. [CrossRef]
7. Arias, A.; Feijoo, G.; Moreira, M.T. How Could Artificial Intelligence Be Used to Increase the Potential of Biorefineries in the near Future? A Review. *Environ. Technol. Innov.* **2023**, *32*, 103277. [CrossRef]
8. Gokcekuyu, Y.; Ekinci, F.; Guzel, M.S.; Acici, K.; Aydin, S.; Asuroglu, T. Artificial Intelligence in Biomaterials: A Comprehensive Review. *Appl. Sci.* **2024**, *14*, 6590. [CrossRef]
9. Du, Y.-H.; Wang, M.-Y.; Yang, L.-H.; Tong, L.-L.; Guo, D.-S.; Ji, X.-J. Optimization and Scale-Up of Fermentation Processes Driven by Models. *Bioengineering* **2022**, *9*, 473. [CrossRef]
10. Rathore, A.S.; Mishra, S.; Nikita, S.; Priyanka, P. Bioprocess Control: Current Progress and Future Perspectives. *Life* **2021**, *11*, 557. [CrossRef]
11. Appl, C.; Moser, A.; Baganz, F.; Hass, V.C. Digital Twins for Bioprocess Control Strategy Development and Realisation. In *Digital Twins*; Herwig, C., Pörtner, R., Möller, J., Eds.; Advances in Biochemical Engineering/Biotechnology; Springer International Publishing: Cham, Switzerland, 2020; Volume 177, pp. 63–94. ISBN 978-3-030-71655-4.
12. Kumar Sharma, A.; Kumar Ghodke, P.; Goyal, N.; Nethaji, S.; Chen, W.-H. Machine Learning Technology in Biohydrogen Production from Agriculture Waste: Recent Advances and Future Perspectives. *Bioresour. Technol.* **2022**, *364*, 128076. [CrossRef] [PubMed]
13. Fajardo Muñoz, S.E.; Freire Castro, A.J.; Mejía Garzón, M.I.; Páez Fajardo, G.J.; Páez Gracia, G.J. Artificial Intelligence Models for Yield Efficiency Optimization, Prediction, and Production Scalability of Essential Oil Extraction Processes from Citrus Fruit Exocarps. *Front. Chem. Eng.* **2023**, *4*, 1055744. [CrossRef]
14. Arango-Manrique, S.; Agudelo Patiño, T.; Matallana Pérez, L.G.; Ortiz-Sanchez, M.; Cardona Alzate, C.A. Conceptual Design and Economic Optimization of Different Valorization Routes for Orange Peel Waste: The Application of the Biorefinery Concept for an Integral Use of Raw Material. *Processes* **2024**, *12*, 2298. [CrossRef]
15. Solarte-Toro, J.C.; Cardona Alzate, C.A. Sustainability of Biorefineries: Challenges and Perspectives. *Energies* **2023**, *16*, 3786. [CrossRef]
16. Rojek, I.; Mikołajewski, D.; Mroziński, A.; Macko, M. Green Energy Management in Manufacturing Based on Demand Prediction by Artificial Intelligence—A Review. *Electronics* **2024**, *13*, 3338. [CrossRef]
17. Agharafeie, R.; Ramos, J.R.C.; Mendes, J.M.; Oliveira, R. From Shallow to Deep Bioprocess Hybrid Modeling: Advances and Future Perspectives. *Fermentation* **2023**, *9*, 922. [CrossRef]
18. Duong-Trung, N.; Born, S.; Kim, J.W.; Schermeyer, M.-T.; Paulick, K.; Borisyak, M.; Cruz-Bournazou, M.N.; Werner, T.; Scholz, R.; Schmidt-Thieme, L.; et al. When Bioprocess Engineering Meets Machine Learning: A Survey from the Perspective of Automated Bioprocess Development. *Biochem. Eng. J.* **2023**, *190*, 108764. [CrossRef]
19. Notin, P.; Rollins, N.; Gal, Y.; Sander, C.; Marks, D. Machine Learning for Functional Protein Design. *Nat. Biotechnol.* **2024**, *42*, 216–228. [CrossRef] [PubMed]
20. Zhao, W.; Zhu, J.; Yang, S.; Liu, J.; Sun, Z.; Sun, H. Microalgal Metabolic Engineering Facilitates Precision Nutrition and Dietary Regulation. *Sci. Total Environ.* **2024**, *951*, 175460. [CrossRef]
21. Pham, T.D.; Manapragada, C.; Sun, Y.; Bassett, R.; Aickelin, U. A Scoping Review of Supervised Learning Modelling and Data-Driven Optimisation in Monoclonal Antibody Process Development. *Digit. Chem. Eng.* **2023**, *7*, 100080. [CrossRef]
22. Richter, J.; Wang, Q.; Lange, F.; Thiel, P.; Yilmaz, N.; Solle, D.; Zhuang, X.; Beutel, S. Machine Learning-Powered Optimization of a CHO Cell Cultivation Process. *Biotech Bioeng.* **2025**, *122*, 1153–1164. [CrossRef]
23. Pääkkönen, S.; Pölönen, I.; Salmi, P. Semi-Supervised Anomaly Detection from *Chlorella Vulgaris* Cultivations Using Hyperspectral Imaging. *Smart Agric. Technol.* **2025**, *12*, 101121. [CrossRef]
24. Medeiros Garcia Alcántara, J.; Iannacci, F.; Morbidelli, M.; Sponchioni, M. Soft Sensor Based on Raman Spectroscopy for the In-Line Monitoring of Metabolites and Polymer Quality in the Biomanufacturing of Polyhydroxyalkanoates. *J. Biotechnol.* **2023**, *377*, 23–33. [CrossRef]

25. Nguyen, X.D.J.; Liu, Y.A.; McDowell, C.C.; Dooley, L. Methodology for Contamination Detection and Reduction in Fermentation Processes Using Machine Learning. *Bioprocess Biosyst. Eng.* **2025**. [\[CrossRef\]](#)
26. Aghaee, M.; Krau, S.; Tamer, I.M.; Budman, H. Unsupervised Hybrid Models Integrating Deep Autoencoders and Process Controllers' Models for Enhanced Process Monitoring and Fault Detection. *Ind. Eng. Chem. Res.* **2024**, *63*, 14748–14760. [\[CrossRef\]](#)
27. Agarwal, P.; Aghaee, M.; Tamer, M.; Budman, H. A Novel Unsupervised Approach for Batch Process Monitoring Using Deep Learning. *Comput. Chem. Eng.* **2022**, *159*, 107694. [\[CrossRef\]](#)
28. Hong, Y.; Nguyen, T.; Arbter, P.; Utesch, T.; Zeng, A. Phenotype Analysis of Cultivation Processes via Unsupervised Machine Learning: Demonstration for *Clostridium pasteurianum*. *Eng. Life Sci.* **2022**, *22*, 85–99. [\[CrossRef\]](#)
29. Thite, N.G.; Yarnell, M.; Fry, T.J.; Seefeldt, M.; Calderon, C.P.; Randolph, T.W. Unsupervised Machine Learning-Based Process Analytical Tools for Near Real-Time Cell Morphology Analysis During CAR-T Cell Manufacturing. *Biotechnol. Bioeng.* **2025**, *122*, 2377–2388. [\[CrossRef\]](#)
30. Li, D.; Zhu, F.; Wang, X.; Jin, Q. Multi-Objective Reinforcement Learning for Fed-Batch Fermentation Process Control. *J. Process Control.* **2022**, *115*, 89–99. [\[CrossRef\]](#)
31. Treloar, N.J.; Fedorec, A.J.H.; Ingalls, B.; Barnes, C.P. Deep Reinforcement Learning for the Control of Microbial Co-Cultures in Bioreactors. *PLoS Comput. Biol.* **2020**, *16*, e1007783. [\[CrossRef\]](#)
32. Monteiro, M.; Kontoravdi, C. Bioprocess Control: A Shift in Methodology Towards Reinforcement Learning. In *Computer Aided Chemical Engineering*; Elsevier: Amsterdam, The Netherlands, 2024; Volume 53, pp. 2851–2856. ISBN 978-0-443-28824-1.
33. Bloor, M.; Ahmed, A.; Kotecha, N.; Mercangöz, M.; Tsay, C.; Del Río-Chanona, E.A. Control-Informed Reinforcement Learning for Chemical Processes. *Ind. Eng. Chem. Res.* **2025**, *64*, 4966–4978. [\[CrossRef\]](#)
34. Zhu, X.; Rehman, K.U.; Wang, B.; Shahzad, M. Modern Soft-Sensing Modeling Methods for Fermentation Processes. *Sensors* **2020**, *20*, 1771. [\[CrossRef\]](#)
35. Baako, T.-M.D.; Kulkarni, S.K.; McClendon, J.L.; Harcum, S.W.; Gilmore, J. Machine Learning and Deep Learning Strategies for Chinese Hamster Ovary Cell Bioprocess Optimization. *Fermentation* **2024**, *10*, 234. [\[CrossRef\]](#)
36. Dara, M.; Azarpira, N. Ethical Considerations Emerge from Artificial Intelligence (AI) in Biotechnology. *AJMB* **2025**, *17*, 80–81. [\[CrossRef\]](#)
37. Niazi, S.K. Regulatory Perspectives for AI/ML Implementation in Pharmaceutical GMP Environments. *Pharmaceuticals* **2025**, *18*, 901. [\[CrossRef\]](#)
38. Karim, M.R.; Islam, T.; Shajalal, M.; Beyan, O.; Lange, C.; Cochez, M.; Rebholz-Schuhmann, D.; Decker, S. Explainable AI for Bioinformatics: Methods, Tools and Applications. *Brief. Bioinform.* **2023**, *24*, bbad236. [\[CrossRef\]](#)
39. Bayer, B.; Dalmau Diaz, R.; Melcher, M.; Striedner, G.; Duerkop, M. Digital Twin Application for Model-Based DoE to Rapidly Identify Ideal Process Conditions for Space-Time Yield Optimization. *Processes* **2021**, *9*, 1109. [\[CrossRef\]](#)
40. Chen, Y.; Yang, O.; Sampat, C.; Bhalode, P.; Ramachandran, R.; Ierapetritou, M. Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review. *Processes* **2020**, *8*, 1088. [\[CrossRef\]](#)
41. Park, S.; Kim, S.; Park, C.; Kim, J.; Lee, D. Data-driven Prediction Models for Forecasting Multistep Ahead Profiles of Mammalian Cell Culture toward Bioprocess Digital Twins. *Biotech Bioeng.* **2023**, *120*, 2494–2508. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Riezzo, L.; Kay, H.; Feng, Y.; Jing, K.; Zhang, D. Accelerating Bioprocess Digital Twin Development by Integrating Hybrid Modelling with Transfer Learning. *Chem. Eng. J.* **2025**, *511*, 162018. [\[CrossRef\]](#)
43. Portela, R.M.C.; Varsakelis, C.; Richelle, A.; Giannelos, N.; Pence, J.; Dessoy, S.; Von Stosch, M. When Is an In Silico Representation a Digital Twin? A Biopharmaceutical Industry Approach to the Digital Twin Concept. In *Digital Twins*; Herwig, C., Pörtner, R., Möller, J., Eds.; Advances in Biochemical Engineering/Biotechnology; Springer International Publishing: Cham, Switzerland, 2020; Volume 176, pp. 35–55. ISBN 978-3-030-71659-2.
44. Cardillo, A.G.; Castellanos, M.M.; Desailly, B.; Dessoy, S.; Mariti, M.; Portela, R.M.C.; Scutella, B.; Von Stosch, M.; Tomba, E.; Varsakelis, C. Towards in Silico Process Modeling for Vaccines. *Trends Biotechnol.* **2021**, *39*, 1120–1130. [\[CrossRef\]](#)
45. Udugama, I.A.; Lopez, P.C.; Gargalo, C.L.; Li, X.; Bayer, C.; Gernaey, K.V. Digital Twin in Biomanufacturing: Challenges and Opportunities towards Its Implementation. *Syst. Microbiol. Biomanuf.* **2021**, *1*, 257–274. [\[CrossRef\]](#)
46. Helgers, H.; Schmidt, A.; Strube, J. Towards Autonomous Process Control—Digital Twin for CHO Cell-Based Antibody Manufacturing Using a Dynamic Metabolic Model. *Processes* **2022**, *10*, 316. [\[CrossRef\]](#)
47. Moser, A.; Appl, C.; Pörtner, R.; Baganz, F.; Hass, V.C. A New Concept for the Rapid Development of Digital Twin Core Models for Bioprocesses in Various Reactor Designs. *Fermentation* **2024**, *10*, 463. [\[CrossRef\]](#)
48. Brunner, V.; Siegl, M.; Geier, D.; Becker, T. Challenges in the Development of Soft Sensors for Bioprocesses: A Critical Review. *Front. Bioeng. Biotechnol.* **2021**, *9*, 722202. [\[CrossRef\]](#)
49. Sharma, N.; Liu, Y.A. A Hybrid Science-guided Machine Learning Approach for Modeling Chemical Processes: A Review. *AIChE J.* **2022**, *68*, e17609. [\[CrossRef\]](#)

50. Palladino, F.; Marcelino, P.R.F.; Schlogl, A.E.; José, Á.H.M.; Rodrigues, R.D.C.L.B.; Fabrino, D.L.; Santos, I.J.B.; Rosa, C.A. Bioreactors: Applications and Innovations for a Sustainable and Healthy Future—A Critical Review. *Appl. Sci.* **2024**, *14*, 9346. [CrossRef]
51. Jang, W.D.; Kim, G.B.; Kim, Y.; Lee, S.Y. Applications of Artificial Intelligence to Enzyme and Pathway Design for Metabolic Engineering. *Curr. Opin. Biotechnol.* **2022**, *73*, 101–107. [CrossRef]
52. Sampaio, P.S.; Fernandes, P. Machine Learning: A Suitable Method for Biocatalysis. *Catalysts* **2023**, *13*, 961. [CrossRef]
53. Walsh, I.; Myint, M.; Nguyen-Khuong, T.; Ho, Y.S.; Ng, S.K.; Lakshmanan, M. Harnessing the Potential of Machine Learning for Advancing “Quality by Design” in Biomanufacturing. *mAbs* **2022**, *14*, 2013593. [CrossRef]
54. Helleckes, L.M.; Hemmerich, J.; Wiechert, W.; Von Lieres, E.; Grünberger, A. Machine Learning in Bioprocess Development: From Promise to Practice. *Trends Biotechnol.* **2023**, *41*, 817–835. [CrossRef]
55. Pandey, K.; Pandey, M.; Kumar, V.; Aggarwal, U.; Singhal, B. Bioprocessing 4.0 in Biomanufacturing: Paving the Way for Sustainable Bioeconomy. *Syst. Microbiol. Biomanuf.* **2024**, *4*, 407–424. [CrossRef]
56. He, C.; Zhang, C.; Bian, T.; Jiao, K.; Su, W.; Wu, K.-J.; Su, A. A Review on Artificial Intelligence Enabled Design, Synthesis, and Process Optimization of Chemical Products for Industry 4.0. *Processes* **2023**, *11*, 330. [CrossRef]
57. Horizon EU SUSHEAT. Available online: <https://susheat.eu/> (accessed on 15 April 2025).
58. Rovira, A.; Guedez, R.; Trevisan, S.; Høeg, A.; Vérez, D.; Cabeza, L.F.; Butean, A.; Enríquez, J.; Law, R.; Muegge, M.; et al. Smart Integration of Waste and Renewable Energy for Sustainable Heat Upgrade in the Industry (SUSHEAT). In Proceedings of the International Seminar on ORC Power Systems 2023 (ORC2023), Seville, Spain, 4–6 September 2023. [CrossRef]
59. Høeg, A.; Løver, K.; Asphjell, T.-A.; Lümmen, N. Performance of a New Ultra-High Temperature Industrial Heat Pump. In Proceedings of the 14th IEA Heat Pump Conference, Chicago, IL, USA, 15–18 May 2023.
60. Cabeza, L.F.; Mani Kala, S.; Zsembinszki, G.; Vérez, D.; Risco Amigó, S.; Borri, E. Development of a Bio-Inspired TES Tank for Heat Transfer Enhancement in Latent Heat Thermal Energy Storage Systems. *Appl. Sci.* **2024**, *14*, 2940. [CrossRef]
61. Matei, A.; Butean, A.; Zamfirescu, B.C.; Marcos, J.D. Designing a Conceptual Digital Twin Architecture for High-Temperature Heat Upgrade Systems. *Appl. Sci.* **2025**, *15*, 2350. [CrossRef]
62. Butean, A.; Enriquez, J.; Matei, A.; Rovira, A.; Barbero, R.; Trevisan, S. A Digital Twin Concept for Optimizing the Use of High-Temperature Heat Pumps to Reduce Waste in Industrial Renewable Energy Systems. *Procedia Comput. Sci.* **2024**, *237*, 123–128. [CrossRef]
63. Shah, M.; Wever, M.; Espig, M. A Framework for Assessing the Potential of Artificial Intelligence in the Circular Bioeconomy. *Sustainability* **2025**, *17*, 3535. [CrossRef]
64. Razzaq, L.; Abbas, M.M.; Miran, S.; Asghar, S.; Nawaz, S.; Soudagar, M.E.M.; Shaukat, N.; Veza, I.; Khalil, S.; Abdelrahman, A.; et al. Response Surface Methodology and Artificial Neural Networks-Based Yield Optimization of Biodiesel Sourced from Mixture of Palm and Cotton Seed Oil. *Sustainability* **2022**, *14*, 6130. [CrossRef]
65. Ning, H.; Li, R.; Zhou, T. Machine Learning for Microalgae Detection and Utilization. *Front. Mar. Sci.* **2022**, *9*, 947394. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.