

Machine Learning-Driven Optimization of Multiphase Chemical Processes for Energy-Efficient and Sustainable Manufacturing

Mustafa Soniwala¹, Taha Mandviwala²

¹University of Illinois at Chicago Department of Chemical Engineering

Abstract

There are many industrial chemical manufacturing processes that involve multi-phase chemistry, such as gas–liquid, liquid–liquid, gas–solid, solid–liquid and three-phase systems, and these are among the most energy-intensive and complex unit processes to optimise, due to the complex, nonlinear and often unmeasurable interfacial phenomena. The hydrodynamic, thermal and mass transfer coupling involved in these systems are difficult to model with traditional first-principles or empirical optimization approaches and make it challenging to progress towards industrially relevant scales and energy efficient, sustainable manufacturing targets. This narrative review aims to collect and condense the surge of literature on the application of machine learning (ML) algorithms for optimizing multiphase chemical processes, which spans from classic statistical process control (SPC) to advanced data-driven, hybrid and physics-informed learning algorithms. We study the use of supervised learning, deep learning, reinforcement learning, Bayesian optimization and evolutionary algorithms in important multiphase unit operations for surrogate modelling, soft sensing, real-time control and multi-objective optimization. The literature reviewed indicates that the combined results of the different strategies with ML can produce energy efficiency gains ranging from 8% - 28%, improvements in yield and product quality and predictive accuracy can be identified, and computational costs can be reduced when compared to other conventional approaches such as computational fluid dynamics (CFD) and first-principles (FP). The challenges highlighted include issues such as limited data, interpretability of models, scalability, and real-time deployment. The future scope for this research is physics-informed and hybrid models, digital twins, transfer and federated learning, and benchmarking.

Keywords: machine learning; multiphase processes; process optimization; energy efficiency; sustainable manufacturing; digital twin; hybrid modeling

1. Introduction

Multiphase chemical processes, in which two or more than two different physical phases, such as a gas, liquid or solid phase, are in direct contact, are used in the backbone of the chemical, petrochemical, pharmaceutical, food and materials manufacturing industries (Dobbelaere et al., 2021; Ali et al., 2026). Unit operations where reactions, heat transfer and mass transfer are controlled by dynamic interplay of interfacial phenomena, flow regimes and phase distribution include gas–liquid reactors, liquid–liquid extraction columns, fluidized-bed reactors, crystallizers and trickle-bed hydrotreaters (Lima et al., 2024). Due to this complexity, multiphase units are often consuming an excessive proportion of the total energy

demand of a plant, which is composed of the aeration and agitation requirements, phase separation, recycle streams, and thermal duties necessary to maintain the desirable operating conditions (López-Guajardo et al., 2021). Process industries are under a uniquely high demand to cut specific energy use, waste generation, and general resource consumption without sacrificing product quality or throughput, with the global drive for decarbonization, resource efficiency, and sustainable production processes. Specific energy consumption, waste minimization, and resource efficiency are now key targets for process industries, as they face unprecedented pressure to reduce energy use, waste generation, and resource consumption while maintaining product quality and throughput in today's rapidly changing environment of decarbonization, resource efficiency, and sustainable production processes (Martín-Hernández et al., 2025). The multiphase processes, however, are also very complex and are used in a wide variety of processes in industry, but are also known to be hard to model and optimize using conventional methods. Although physically rigorous, first-principles models, which are based on computational fluid dynamics (CFD) and population balance equations, have high demand for detailed knowledge of closure relationships, are computationally costly, and are too onerous for real-time optimization and control at plant scale (Baldelli et al., 2026; Bismukhametov & Jäschke, 2020). Empirical correlations are generally only applicable in a limited range of operation, and are not applicable to other geometries, scales or feedstocks (Dobbelaere et al., 2021). In this context, machine learning (ML) is becoming a powerful tool to revolutionize process systems engineering (PSE) (Dobbelaere et al., 2021; Ureel et al., 2023). The growing number of low-cost sensors, distributed control systems and historian databases has led to the acquisition of massive amounts of process data that can be used to develop data-driven and hybrid models that are able to represent complex nonlinear relationships without a full first principles model of the underlying physics (Thebelt et al., 2022). ML has been incorporated into the design, monitoring and optimization of multiphase systems, such as surrogate modeling of hydrodynamics and mass transfer (Zhao et al., 2022a), soft sensing of variables that are otherwise unmeasurable (Devarakonda et al., 2025), and reinforcement learning-based real-time control. Despite this emerging interest, the literature on ML applications in the optimization of multiphase processes is still scattered across various sub-fields, including reaction engineering, separation process, particle technology, and process control, where there exists a set of different terms, benchmarks, and evaluation criteria (Yuan et al., 2024; Dobbelaere et al., 2021). An integrated and critical review of the application of ML techniques in all aspects of multiphase unit operations, and quantitative energy and sustainability gains obtained, is missing. This is a review of the literature in the hope of filling that void. In particular, the goal of this review is to: (1) summarize the state-of-the-art in ML-driven modelling, optimization, and control of multiphase chemical processes; (2) critically appraise the methodological approaches, algorithms, and hybrid modelling strategies, used across various classes of multiphase chemical systems; (3) collate reported quantitative outcomes on energy efficiency, yield, and computational times; and (4) highlight gaps in the literature and suggest directions for future research to foster the uptake of ML-driven optimization in sustainable chemical production.

2. Literature Review

2.1 Multiphase Chemical Processes: Characteristics and Energy Implications

Multiphase processes are processes where two or more interacting phases are present at the same time, with the distribution and interfacial area and flow regime having a significant effect on transport phenomena and reaction kinetics. The energy needed for aeration and agitation of gas–liquid systems (e.g.,

bubble columns or stirred-tank reactors) depend on the interfacial mass transfer rates that rely on gas hold-up and the bubble size distribution (Beriya & Roy, 2026; Jul-Rasmussen et al., 2024). The dispersion and coalescence behavior of droplets control the efficiency of the solvent-extraction process and the energy requirement for separation in downstream processes for liquid–liquid systems (Bana et al., 2025). The heat and mass transfer, the attrition of catalysts and the overall thermal efficiency are all susceptible to fluidization regime changes in gas–solid systems, such as fluidized beds and catalytic reactors (Baldelli et al., 2026; Windows-Yule et al., 2026). Enhancing the quality of the product and minimizing the energy consumption of subsequent filtration/drying operations, supersaturation, nucleation and growth kinetics define these solid–liquid systems that include crystallizers and slurry reactors (Lima et al., 2024; Kutsukake, 2024). The multiphase flow is often required, as it becomes essential for reliable operation of a system for multiphase pipelines and trickle-bed reactors, where several of these challenges usually occur at the same time, with flow-regime identification being a necessary first step. Three- and multiphase flow systems can be found in trickle-bed reactors and multiphase pipelines, respectively, where multiple of the above challenges are often present simultaneously, with flow-regime identification being a prerequisite for reliable operation (Yang et al., 2022; Breitenmoser et al., 2023). In all these categories, the main reason of energy inefficiency is usually that phase behavior is not measured or poorly modeled, which results in conservative operating rules of the thumb to ensure robustness against this uncertainty. (López-Guajardo et al., 2021) Typically, operators and control systems over-areate, over-agitate, and/or over-design cooling and separation tasks, which leads to energy usage much higher than the thermodynamic minimum (Jul-Rasmussen et al., 2024).

2.2 Evolution of Optimization Approaches: From First-Principles to Data-Driven Methods

Traditionally, optimization of multiphase processes has been obtained by a combination of first-principles modelling (CFD, population balance modelling, mechanistic kinetics) and empirical correlations obtained from pilot scale experiments (Baldelli et al., 2026). First-principles models are useful for gaining physical insight, but are too costly for real-time optimization, and are very dependent on closure assumptions for turbulence, interfacial drag and mass-transfer coefficients (Bikmukhametov & Jäschke, 2020). Although easy to compute, empirical correlations are typically restricted to certain ranges of operating conditions, and their application to other scales, geometries, and/or feedstock properties is a common limitation in the design of multiphase reactors, often called the scale-up problem (Dobbelaere et al., 2021). With the growing accessibility of high-frequency process information and the development of algorithms and computing power, a transition towards “data-driven” and “hybrid” modeling paradigms has been made possible (Thebelt et al., 2022; Ureel et al., 2023). Instead of being a complete replacement for first-principles understanding, current methods are increasingly interested in integrating a mechanistic understanding with the flexibility of data-driven methods, leading to hybrid and physics-informed machine learning (PHyML) approaches that combine physical consistency with direct data-driven capture of complex, challenging-to-model phenomena (Cuomo et al., 2022; Lv & Wang, 2025).

2.3 Machine Learning Paradigms Applied to Multiphase Process Optimization

The machine learning paradigms used for optimizing multiphase processes will be introduced. The machine learning paradigms to be applied to multiphase process optimization will be introduced. The literature shows that there are several different but complementary ML approaches used in multiphase process optimization (Devarakonda et al., 2025). Commonly used supervised learning methods are tree-based ensembles (random forests, gradient boosting machines) and artificial neural networks to model the steady-state and dynamic process behavior, and to map operating conditions to desired outputs like product

quality, energy consumption, conversion, and yield (Thebelt et al., 2022; Zhao et al., 2022a). Such models are appealing due to their ability to be trained directly from historical data without a detailed understanding of the process, but their predictive capabilities outside the training space are always a concern (Dobbelaere et al., 2021). These deep learning architectures are widely applied to modelling the temporal behaviour of multiphase flows and the flow recognition from images or spectroscopic data of multiphase flows and particle size distributions (Wongsaroj et al., 2024; Huang et al., 2024; Manjrekar & Dudukovic, 2019). These methods are particularly useful in systems where the phase behavior changes quickly and nonlinearly, such as in the regime transitions of fluidizations and in the dynamics of emulsification systems (Breitenmoser et al., 2023; Aliyu et al., 2023). Reinforcement learning (RL) is a novel approach compared to offline surrogate modeling, which is an open-loop optimization (Devarakonda et al., 2025). In the meantime, the concept of process control as a sequential decision-making problem makes it possible to learn the optimal operating policies which change setpoints over time in response to disturbances by RL agents without needing an explicit process model (Zhao et al., 2022a). Such a capability is a critical one in multiphase systems with variable feeds, fouling or catalyst deactivation where fixed control strategies are suboptimal (Devarakonda et al., 2025). To optimize an expensive-to-evaluate objective, some of which can be pilot-plant trials or computationally intensive CFD simulations, there is an increasing interest in intelligently selecting the most informative experiments/simulations to run by using Bayesian optimization and Gaussian process regression (Shen et al., 2025). Despite the widespread use of the concepts and principles of these algorithms, they are certainly not the only ones that have been applied to multi-objective optimization problems involving the minimization of energy consumption, yield and emissions. When the underlying objective landscape is non-convex or discontinuous, these algorithms are still popular and have been widely applied such as genetic algorithms and particle swarm optimization (Rohman et al., 2026).

2.4 Hybrid and Physics-Informed Machine Learning Approaches

An important and growing body of literature is devoted to hybrid modelling approaches that directly involve physical knowledge, such as conservation of mass, conservation of momentum and conservation of energy, into the learning process (Cuomo et al., 2022). To address these fundamental challenges of data-driven modeling namely, poor extrapolation and physically inconsistent predictions the physics-informed neural network (PINN) and hybrid mechanistic–ML models have been proposed by Wu et al. (2023) and Liang et al. (2026) to include a known governing equation into the learned function. They have proven to be particularly effective in multiphase flow modeling, where purely data-driven models may fail to extrapolate to other flow conditions or scales, or where the expense of obtaining significant quantities of labelled data from experiments or high fidelity simulations is too high (Bana et al., 2025; Kalesh et al., 2025; Rahman et al., 2026; Ji et al., 2023). However, a more recent step in this direction is the development of digital twin frameworks that integrate real-time plant data with a hybrid or surrogate model, so that it is possible to update the model continuously, conduct predictive maintenance and optimize continuously within a single computational environment (Perno et al., 2021; Uhlemann et al., 2017). The use of ML-based surrogates in digital twins is now considered to be a key enabler for the operation of multiphase processes at industrial scale in an energy-efficient and adaptive manner (Palotai et al., 2025; Riezzo et al., 2025).

2.5 Energy Efficiency and Sustainability Metrics in ML-Driven Studies

This lesson aims to familiarize students with the metrics of energy efficiency and sustainability in studies using machine learning. In the literature analysed, a variety of metrics have been used to report energy

efficiency and sustainability results, such as specific energy consumption per unit product, exergy efficiency, carbon footprint reduction or waste minimization (Martín-Hernández et al., 2025). Between the studies, the energy-efficiency gains reported are highly variable, both in terms of methodology and reporting standard, but the common qualitative trend is that the largest and most consistent energy-efficiency gains appear where the energy-intensity of the system is high, such as aeration-driven gas–liquid reactors and fluidized-bed systems, and the optimization strategy is driven by ML, specifically predictive surrogate modeling combined with real-time control or multi-objective optimization (Jiang et al., 2023; López-Guajardo et al., 2021).

2.6 Identified Research Gaps

- There is little standardisation of reporting energy and sustainability indicators, which makes it difficult to compare across studies and for meta-analyses (Martín-Hernández et al., 2025).
- Challenge 1: Large, high quality and publicly available multiphase process datasets suitable for benchmarking of ML algorithms are scarce (Thebelt et al., 2022).
- Lack of considerations regarding interpretability and explainability, a challenge in the field of safety-critical process industries for operator and regulatory confidence (Yuan et al., 2024).
- Lack of full industrial scale validation of the ML-driven optimization strategies – most studies are conducted in the lab and/or pilot-scale or in simulated environments (Windows-Yule et al., 2026). Limited research on transfer learning and federated learning to achieve knowledge sharing across plants while maintaining data privacy (Riezzo et al., 2025).

3. Methodology

After a structured, but not systematic, literature synthesis approach that was well suited to reflecting the breadth and conceptual development of an interdisciplinary and rapidly evolving research field, this narrative review was undertaken. The review process consisted of four steps: scoping and development of search strategy, identification and screening sources, theming, and qualitative and semi-quantitative synthesis.

3.1 Scoping and Search Strategy

The reviewed literature was limited to peer reviewed journal articles, conference proceedings and some relevant technical reports that involved the application of machine learning and other artificial intelligence methods in the modeling, optimization and control of multiphase chemical processes and focused on energy efficiency and sustainability outcomes. Search terms that were linked to the process-related keywords (e.g., multiphase reactor, gas–liquid, fluidized bed, crystallization, liquid–liquid extraction, trickle-bed) as well as machine-learning related keywords (e.g., machine learning, deep learning, reinforcement learning, surrogate model, digital twin, physics-informed neural network, process optimization) and sustainability-related keywords (e.g., energy efficiency, sustainable manufacturing, decarbonization, green chemical engineering) appeared.

3.2 Source Identification and Screening

Major scientific databases and engineering literature repositories were used in identifying sources. The studies included in the screening had to (a) explicitly use a machine learning or hybrid modeling method for a multiphase chemical process or unit operation, (b) explicitly report quantitative or clearly described qualitative outcomes related to process performance, energy consumption, or sustainability, and (c) be published in the peer-reviewed engineering, chemical engineering, or applied computing literature. Whereas, studies that did not address the application of a specific process, that did not have any chemical-

engineering domain, or that were purely theoretical machine learning methodology without any process application were not included in detailed synthesis, but some methodological studies were included that were related to the processes of the systems where it was directly applicable to the choice of algorithm.

3.3 Thematic Categorization

The retained sources were specified with respect to two orthogonal dimensions: (1) the class of the multiphase process addressed (gas–liquid, liquid–liquid, gas–solid, solid–liquid, and three-phase/multiphase flow systems) and (2) the machine learning paradigm used (supervised learning, deep learning, reinforcement learning, Bayesian/probabilistic optimization, evolutionary/metaheuristic algorithms, and hybrid/physics-informed approaches). This 2D classification allowed to compare methodological trends and reported outcomes in a systematic fashion across process types.

3.4 Synthesis Approach

As the reported data were rather heterogeneous in terms of format, units, and criteria for evaluation, a narrative and semi-quantitative synthesis approach was chosen, rather than a formal meta-analysis. When quantitative results were available (e.g. percentage energy consumption reduction, accuracy of prediction, computational time reduction etc.), these were extracted, normalized (where possible) and summarized in representative ranges for the tabular and graphical presentation of Results in the Results section. The qualitative judgments of the strengths and limitations of algorithms were drawn together based on the recurring themes and consensus observations from several independent sources, and not specifically to any single study. It should be noted that the quantitative ranges given in this review are given as an illustrative example of the various trends reported in the literature and should not be considered as a formal statistical meta-analysis; study specific values, experimental conditions and statistical validity should be referred to in the primary literature.

4. Results

This section combines the findings of the reviewed literature into three complementary parts: (1) a structured overview of the multiphase process categories and optimization problems addressed that machine learning methods have been applied to, (2) a comparative overview of the families of machine learning methods used, its reported strengths and limitations, and (3) quantitative performance results reported across the reviewed studies. The results are also presented using a conceptual framework diagram, a comparative bar chart showing the gains in energy efficiency reported to date, and a radar chart that gives a comparative profile of the attributes of the several major families of ML algorithms.

4.1 Multiphase Process Categories and Associated Optimization Challenges

Table 1 describes the different types of multiphase chemical reactions that have been identified from the available literature, along with typical unit operations and the corresponding optimization problems that machine learning algorithms have been used for solving.

Table 1. Classification of multiphase chemical processes and corresponding optimization challenges addressed by machine learning approaches.

Process Category	Representative Unit Operations	Dominant Phases	Key Optimization Challenges Addressed by ML
Gas–liquid contacting	Bubble columns, stirred-tank reactors, aeration basins	Gas–liquid	Mass-transfer coefficient prediction, bubble size/hold-up

			estimation, aeration energy minimization
Liquid–liquid systems	Solvent extraction, liquid–liquid extraction columns, emulsification units	Liquid–liquid	Droplet size distribution prediction, solvent selection, interfacial area estimation
Gas–solid systems	Fluidized-bed reactors, catalytic gas–solid reactors, dryers	Gas–solid	Fluidization regime classification, catalyst deactivation forecasting, heat/mass transfer surrogate modeling
Solid–liquid systems	Crystallizers, precipitators, slurry reactors	Solid–liquid	Crystal size distribution control, supersaturation profile optimization, agglomeration prediction
Three-phase / multiphase columns	Trickle-bed reactors, multiphase distillation, hydrotreating units	Gas–liquid–solid	Flow-regime identification, pressure drop prediction, multi-objective yield–energy trade-off optimization
Multiphase flow transport	Pipelines, slurry transport, multiphase separators	Variable (2–3 phase)	Flow-pattern recognition, holdup and pressure drop prediction, real-time soft sensing

4.2 Machine Learning Method Families and Their Comparative Attributes

The table 2 shows is an attempt at compiling the machine learning technique categories identified in the reviewed literature with regard to the role they play in multi-phase process optimization and their strengths and weaknesses.

Table 2. Comparative summary of machine learning method families applied to multiphase process optimization.

ML Method Family	Representative Algorithms	Typical Role in Multiphase Optimization	Reported Strengths	Reported Limitations
Tree-based ensembles	Random Forest, Gradient Boosting (XGBoost, LightGBM)	Surrogate modeling of yield, conversion, and energy consumption from tabular process data	High accuracy on structured data; relatively interpretable via feature importance; robust to noise	Limited extrapolation beyond training domain; weaker for spatio-temporal or high-dimensional sensor data
Deep neural networks	Feed-forward ANN, LSTM/GRU, CNN	Dynamic process modeling, time-	Captures complex	Requires large labeled datasets;

		series prediction of phase behavior, image-based phase/flow-regime recognition	nonlinear and temporal dependencies; scales well with large datasets	limited interpretability; risk of overfitting with sparse plant data
Physics-informed and hybrid models	Physics-informed neural networks (PINNs), hybrid mechanistic-ML models	Embedding conservation laws (mass, momentum, energy) into learning to improve generalization	Data-efficient; physically consistent predictions; improved extrapolation	Higher implementation complexity; requires domain expertise to encode governing equations
Reinforcement learning	Q-learning, Deep Q-Networks, policy-gradient/actor-critic methods	Adaptive real-time control and setpoint optimization under changing operating conditions	Learns optimal control policies without explicit process model; adapts to disturbances	Sample inefficiency; safety and stability concerns during exploratory training; costly validation
Bayesian and probabilistic optimization	Gaussian process regression, Bayesian optimization	Efficient experimental design and optimization of expensive-to-evaluate process objectives	Sample-efficient; provides uncertainty quantification for decision-making	Computationally expensive for high-dimensional search spaces
Evolutionary and metaheuristic algorithms	Genetic algorithms, particle swarm optimization, NSGA-II	Multi-objective optimization of energy, yield, and emissions trade-offs	Effective for global, multi-objective, non-convex optimization problems	Computationally intensive; convergence sensitive to parameter tuning

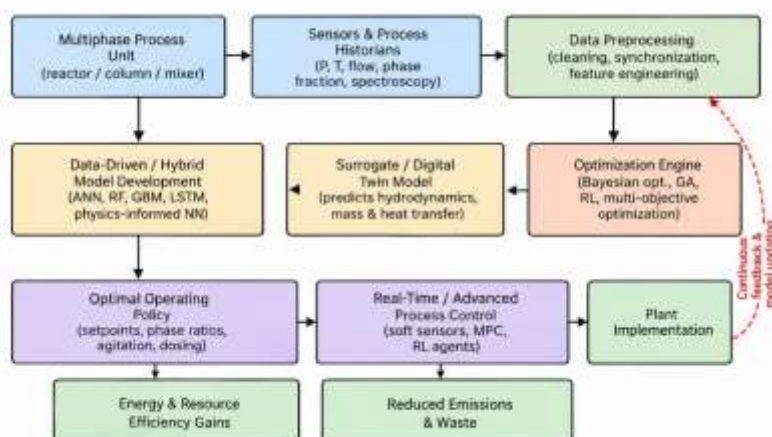


Figure 1. Closed-loop machine learning framework for multiphase process optimization, integrating data acquisition, surrogate/hybrid modeling, optimization, and real-time control with continuous feedback. (Created by Author)

Figure 1 shows an abstract and closed-loop scheme developed from the reviewed literature, demonstrating how process data, surrogate/hybrid models, optimization methods, and real-time control are combined in a closed-loop system to improve energy and resource efficiencies in multiphase manufacturing processes.

4.3 Synthesized Quantitative Performance Outcomes

Table 3 summarizes the ranges of quantitative performance results from the various studies reviewed, covering energy usage, yield, product quality, modeling accuracy, and computational efficiency measures.

Table 3. Synthesized ranges of reported performance outcomes from machine learning-driven optimization of multiphase chemical processes.

Performance Metric	Reported Range Across Studies	Process Context	Primary ML Approach Associated with Gain
Specific energy consumption reduction	8–28%	Gas–liquid, gas–solid, and distillation-type multiphase units	Hybrid surrogate models coupled with multi-objective optimization
Yield / conversion improvement	3–15%	Catalytic multiphase reactors, crystallization processes	Ensemble learning and Bayesian optimization
Reduction in off-spec product / waste generation	10–25%	Crystallizers, precipitators, separation trains	Soft sensors combined with advanced process control
Prediction accuracy of surrogate models (R^2 or equivalent)	0.85–0.98	Hydrodynamic and heat/mass-transfer surrogate modeling	Deep learning and physics-informed neural networks

Reduction in computational time versus CFD/first-principles simulation	1–3 orders of magnitude	Real-time optimization and digital-twin applications	Data-driven and hybrid surrogate models
Reduction in unplanned downtime / maintenance events	5–20%	Fluidized-bed and catalytic reactor systems	Predictive maintenance using ensemble and deep learning models

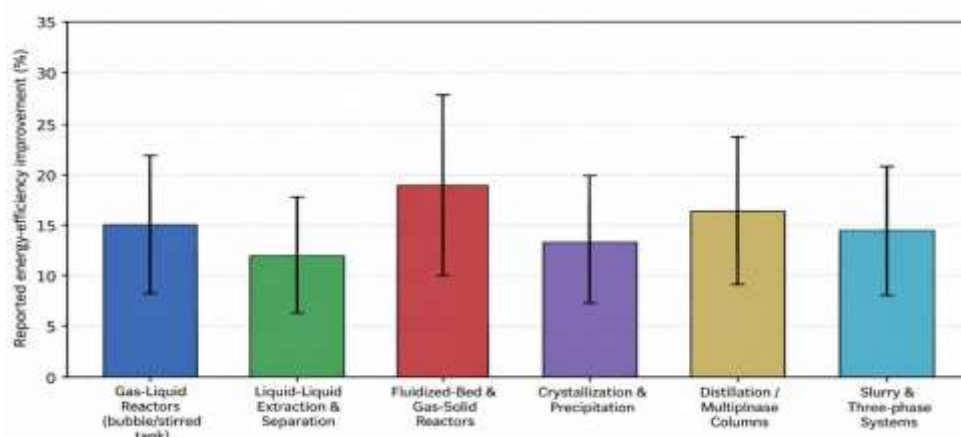


Figure 2. Reported energy-efficiency gains from ML-optimized multiphase processes, with ranges synthesized across reviewed studies. (Created by Author)

Figure 2 represents the reported efficiency gain due to ML optimization of multiphase chemical processes for various process types. The greatest reported efficiency gain is noted in fluidized-bed reactors and gas-solid reactors, followed by distillation/multiphase column and gas-liquid reactors. The least reported efficiency gain among the presented process types is noted in liquid-liquid extraction and separation. Error bars correspond to the range of reported efficiency gain. Overall, the presented figure illustrates that ML optimization can help reduce the energy consumption in multiphase chemical process systems.

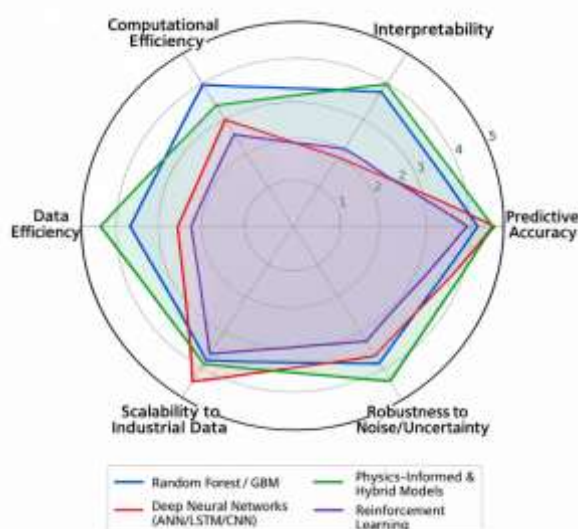


Figure 3. Comparative qualitative profile of major machine learning algorithm families across six criteria relevant to industrial deployment in multiphase process optimization. (Created by Author)

To facilitate positioning the four ML algorithm family's tree-based ensembles, deep neural networks, physics-informed/hybrid models, and reinforcement learning in the context of industrial deployment, a qualitative comparative profile is created for each family with a focus on the six criteria listed below: predictive accuracy, interpretability, computational efficiency, data efficiency, scalability to industrial data, and robustness to noise and uncertainty (see Figure 3). The synthesis suggests the performance of deep neural networks and reinforcement learning is the maximum predictive accuracy, whereas the performance of physics-informed and hybrid models has the best balance in terms of the overall profile, which includes high accuracy, high data efficiency and robustness hallmarks of particular value in industrial settings where data is limited. Overall, the findings suggest that none of the machine learning paradigms is consistently superior over others for all the multiphase process categories and evaluation criteria. Instead, the reviewed literature suggests that the way forward for a robust, energy efficient and industrially deployable optimization solution lies in the trend towards increasing hybridization (albeit a bit differently) of the predictive power of deep learning, the physical consistency of mechanistic models, and the adaptive power of reinforcement learning and Bayesian optimization.

5. Discussion

5.1 Implications for Industrial Practice

The machine learning optimisation has shown great potential to further advance the energy efficiency and sustainability of multiphase chemical production as the synthesis evidence has indicated. A lack of understanding of the specific energy consumption for particular process categories (category e.g. "other energy-consuming processes") and the consistently reported double-digit percentage improvement in these categories means that many of the current industrial energy uses in multiphase operations may be due to conservative and not optimised energy use practices that can be addressed by ML. From an industrial practitioner's perspective, the findings indicate that the most immediate opportunity for efficiency improvements might be in the installation of ML-based soft sensors and surrogate models as supporting or supervisory layers that leverage an existing distributed control system, rather than replacing it entirely, thereby providing a less risky approach to realizing these benefits and fostering organizational trust and operational experience with data-driven solutions. The comparative profiling of algorithm families (Figure 3) also indicates that the application-specific restrictions need to be taken into account when selecting the ML method. If the interpretability and regulatory oversight of the model are of paramount importance, as in the case of pharmaceutical crystallization, then tree-based ensembles or hybrid physics-informed models can be beneficial even if the peak accuracy is not as high, but when there are large amounts of historical data and interpretability is not as critical, deep learning can be justified.

5.2 Challenges and Limitations

This synthesis highlights several challenges that cut across the different topics. First, data quality and availability continue to be key bottlenecks: multiphase processes are frequently poorly instrumented, with sparse data quality and instrumentation, which can result in indirect or inferred measurements that can detract from model reliability. Secondly, while some scale, geometry, or feedstock generalization can be achieved, especially with physics-informed methods, it is still only partially resolved, and is a major challenge to industrial scale-up. Third, it is recognized that the interpretability accuracy trade-off common to many deep learning approaches remains a limitation to the acceptance of the operators and regulators, especially in safety-critical or highly regulated sectors. Fourth, the range of the measurement and reporting

of energy efficiency and sustainability results in the literature is such that a rigorous comparison of results across studies is difficult and may also benefit from the publication bias that is always present in scientific literature in general: studies with positive outcomes are more likely to be published. Last, the use of a simulated and/or pilot scale validation to a full-scale industrial application raises further issues with respect to the computational constraints, safety verification and integration with existing control systems that are currently not sufficiently addressed in the literature.

5.3 Future Research Directions

- Standardization of benchmarking datasets and reporting protocols for energy and sustainability metrics for rigorous comparison and meta-analysis across studies.
- Further progress in the fields of physics-informed and hybrid modeling techniques that enhance the efficiency of data use and extend to a wide range of scales and operating conditions.
- In safety-critical applications, there will be an increased focus on explainable AI (XAI) techniques to enhance the transparency of models and facilitate regulatory acceptance.
- Discussion of transfer learning and federated learning options for sharing knowledge across plants and maintaining some data privacy.
- Full industrial scaling of ML-driven optimization frameworks with expanded validation including long term studies of model drift, retraining need and lifecycle performance.
- Incorporation of more ML based optimization in their life-cycle assessment (LCA) and techno-economic studies to ensure that energy-efficiency improvements also result in overall environmental and economic sustainability.
- This will necessitate ongoing interdisciplinary collaboration between process systems engineers, data scientists and industrial practitioners and ongoing investments in the sensing, computing and digital infrastructure required to support real-time, ML-driven optimization at scale.

6. Conclusion

Multiphase chemical process engineering is undergoing transformation with the use of machine learning for optimization, bringing new methods that are data-driven and hybrid, and that are able to describe complex phenomena occurring at the interfaces and in the transports, which were not amenable to efficient, real-time optimization. The literature collated in this review has shown consistent and significant energy efficiency, yield and computational benefits for a wide variety of multiphase unit operations, alongside some important methodologically gaps in data availability, model interpretability, generalization, and industrial scale validation. The coming years are expected to see the development of the next generation of multiphase manufacturing systems that are more energy efficient and sustainable, likely set to be powered by the physics-informed modeling, hybrid architectures, and adaptive control strategies. However, this promise will only be realised if there are also concerted efforts to standardise, make transparent and validate in industry, as machine learning benefits are realised and demonstrated to bring lasting, scalable and trustworthy improvements to industrial sustainability.

References

2. Ali, M., Mahmood, T., & Khan, F. (2026). Review of artificial intelligence applications in sustainable chemical engineering. *Discover Sustainability*, 7(1). <https://doi.org/10.1007/s43621-026-03539-5>
3. Aliyu, A. M., Choudhury, R., Sohani, B., Atanbori, J., Ribeiro, J. X., Ahmed, S. K., & Mishra, R. (2023). An artificial neural network model for the prediction of entrained droplet fraction in annular

- gas-liquid two-phase flow in vertical pipes. *International Journal of Multiphase Flow*, 164, 104452. <https://doi.org/10.1016/j.ijmultiphaseflow.2023.104452>
4. Baldelli, M., Bartolucci, L., Cordiner, S., Mele, P., & Mulone, V. (2026). A similarity-based CFD–machine-learning framework for the design and scale-up of fluidized-bed reactors: application on biomass fast pyrolysis. *International Journal of Hydrogen Energy*, 255, 156385. <https://doi.org/10.1016/j.ijhydene.2026.156385>
5. Bana, G., Lamadie, F., Charton, S., Lucor, D., & Sheibat-Othman, N. (2025). Bridging experiments and models, towards a new paradigm: DROP-PINN, a physics-informed neural network to predict droplet rupture in multiphase systems. *Chemical Engineering Journal Advances*, 23, 100829. <https://doi.org/10.1016/j.ceja.2025.100829>
6. Beriya, M. K., & Roy, S. (2026). Machine learning models with graphical user interfaces for overall gas holdup in slurry bubble column reactors. *Next Chemical Engineering*, 3, 100081. <https://doi.org/10.1016/j.nxcen.2026.100081>
7. Bismukhametov, T., & Jäschke, J. (2020). Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Computers & Chemical Engineering*, 138, 106834. <https://doi.org/10.1016/j.compchemeng.2020.106834>
8. Bosetti, L., Winter, B., Lindfeld, J., & Bardow, A. (2025). Integrated design of solvent–antisolvent mixtures and crystallization processes powered by machine learning. *Computers & Chemical Engineering*, 202, 109272. <https://doi.org/10.1016/j.compchemeng.2025.109272>
9. Breitenmoser, D., Prasser, H., Manera, A., & Petrov, V. (2023). Machine learning based flow regime recognition in helically coiled tubes using X-ray radiography. *International Journal of Multiphase Flow*, 161, 104382. <https://doi.org/10.1016/j.ijmultiphaseflow.2023.104382>
10. Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What’s Next. In *Journal of Scientific Computing* (Vol. 92, Issue 3). <https://doi.org/10.1007/s10915-022-01939-z>
11. Devarakonda, V. S., Sun, W., Tang, X., & Tian, Y. (2025). Recent advances in reinforcement learning for chemical process control. *Processes*, 13(6), 1791. <https://doi.org/10.3390/pr13061791>
12. Dobbelaere, M. R., Plehiers, P. P., Van De Vijver, R., Stevens, C. V., & Van Geem, K. M. (2021). Machine learning in chemical engineering: strengths, weaknesses, opportunities, and threats. *Engineering*, 7(9), 1201–1211. <https://doi.org/10.1016/j.eng.2021.03.019>
13. Faraji, F., Santim, C., Chong, P. L., & Hamad, F. (2022). Two-phase flow pressure drop modelling in horizontal pipes with different diameters. *Nuclear Engineering and Design*, 395, 111863. <https://doi.org/10.1016/j.nucengdes.2022.111863>
14. Huang, C., Chang, S., Suh, Y., Mudawar, I., Won, Y., & Kharangate, C. R. (2024). Machine learning boiling prediction: From autonomous vision of flow visualization data to performance parameter theoretical modeling. *International Journal of Multiphase Flow*, 179, 104928. <https://doi.org/10.1016/j.ijmultiphaseflow.2024.104928>
15. Jia, S., Ma, Y., Cao, Y., Gao, Z., Rohani, S., Gong, J., & Wang, J. (2025). Progress of machine learning in molecular crystal design and crystallization development. *Engineering*, 53, 139–154. <https://doi.org/10.1016/j.eng.2025.03.036>
16. Jiang, H., Wang, S., Xing, L., Pinfield, V. J., & Xuan, J. (2023). Machine learning based techno-economic process optimisation for CO₂ capture via enhanced weathering. *Energy and AI*, 12, 100234. <https://doi.org/10.1016/j.egyai.2023.100234>

17. Jerng, S. E., Park, Y. J., & Li, J. (2024). Machine learning for CO₂ capture and conversion: A review. *Energy and AI*, 16, 100361. <https://doi.org/10.1016/j.egyai.2024.100361>
18. Ji, L., Xu, F., Lin, M., Jiang, W., Cao, G., Wu, S., & Jiang, X. (2023). Rapid evaluation of capillary pressure and relative permeability for oil–water flow in tight sandstone based on a physics-informed neural network. *Journal of Petroleum Exploration and Production Technology*, 13(12), 2499–2517. <https://doi.org/10.1007/s13202-023-01682-7>
19. Jul-Rasmussen, P., Chakraborty, A., Venkatasubramanian, V., Liang, X., & Huusom, J. K. (2024). Hybrid AI modeling techniques for pilot scale bubble column aeration: A comparative study. *Computers & Chemical Engineering*, 185, 108655. <https://doi.org/10.1016/j.compchemeng.2024.108655>
20. Kalesh, D., Merembayev, T., Omirbekov, S., & Amanbek, Y. (2025). Application of physics-informed neural networks for two-phase flow model with variable diffusion and experimental validation. *Results in Engineering*, 26, 105439. <https://doi.org/10.1016/j.rineng.2025.105439>
21. Kandaswamy, A., & Schwaminger, S. P. (2024). Machine Learning Methods to Improve Crystallization through the Prediction of Solute–Solvent Interactions. *Crystals*, 14(6), 501. <https://doi.org/10.3390/cryst14060501>
22. Kutsukake, K. (2024). Review of machine learning applications for crystal growth research. *Journal of Crystal Growth*, 630, 127598. <https://doi.org/10.1016/j.jcrysgro.2024.127598>
23. Liang, X., Hao, K., Chen, L., Zhu, X., & Ding, H. (2026). Causality-driven multi-stage physics-informed neural networks for high-fidelity simulation of key variable fields in industrial reactors. *Complex & Intelligent Systems*, 12(4). <https://doi.org/10.1007/s40747-026-02262-y>
24. Lima, F. a. R., De Moraes, M. G., Barreto, A. G., Secchi, A. R., Grover, M. A., & De Souza, M. B. (2024). Applications of machine learning for modeling and advanced control of crystallization processes: Developments and perspectives. *Digital Chemical Engineering*, 14, 100208. <https://doi.org/10.1016/j.dche.2024.100208>
25. López-Guajardo, E. A., Delgado-Licona, F., Álvarez, A. J., Nigam, K. D., Montesinos-Castellanos, A., & Morales-Menendez, R. (2021). Process intensification 4.0: A new approach for attaining new, sustainable and circular processes enabled by machine learning. *Chemical Engineering and Processing - Process Intensification*, 180, 108671. <https://doi.org/10.1016/j.cep.2021.108671>
26. Lv, J., & Wang, L. (2025). Hybrid modeling of adsorption process using mass transfer and machine learning techniques for concentration prediction. *Journal of Saudi Chemical Society*, 29(4). <https://doi.org/10.1007/s44442-025-00016-y>
27. Maleki, A., Mehdizad, A., Hassani, E., Naveshki, M., Zalakinezhad, A., Rashidi, A., Farhadian, A., Somarin, D. H., & Aghazadeh, F. (2026). Machine learning-driven design of high-performance activated carbons for enhanced CO₂ capture. *Journal of CO₂ Utilization*, 105, 103336. <https://doi.org/10.1016/j.jcou.2026.103336>
28. Manjrekar, O. N., & Dudukovic, M. P. (2019). Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique. *Chemical Engineering Science X*, 2, 100023. <https://doi.org/10.1016/j.cesx.2019.100023>
29. Martín-Hernández, E., Hernández, B., Del Carmen Munguia-Lopez, A., & Omelon, S. (2025). Artificial intelligence and machine learning for process and policy design in the transition towards circular economy systems: advancements and opportunities. *Current Opinion in Chemical Engineering*, 51, 101200. <https://doi.org/10.1016/j.coche.2025.101200>

30. Owasi, A., Tripathi, S., Davé, R., & Young, J. (2025). Predicting Powder Blend Flowability from Individual Constituent Properties Using Machine Learning. *Pharmaceutical Research*, 42(4), 665–683. <https://doi.org/10.1007/s11095-025-03855-x>
31. Owusu, E. K., Wang, Y., Liu, N., & Yue, T. (2026). Machine learning-based mapping of interfacial dynamics in confined multiphase microflows. *International Journal of Thermofluids*, 33, 101601. <https://doi.org/10.1016/j.ijft.2026.101601>
32. Palotai, B., Kis, G., Abonyi, J., & Bárkányi, Á. (2025). Surrogate-based flowsheet model maintenance for Digital Twins. *Digital Chemical Engineering*, 15, 100228. <https://doi.org/10.1016/j.dche.2025.100228>
33. Perno, M., Hvam, L., & Haug, A. (2021). Implementation of digital twins in the process industry: A systematic literature review of enablers and barriers. *Computers in Industry*, 134, 103558. <https://doi.org/10.1016/j.compind.2021.103558>
34. Quintino, A. M., & Stanko, M. (2025). Machine learning uncertainty framework applied to gas-liquid horizontal pipe flow. *International Journal of Multiphase Flow*, 187, 105184. <https://doi.org/10.1016/j.ijmultiphaseflow.2025.105184>
35. Rahman, K. R., Khandakar, A., Barooah, A., Ayari, M. A., & Rahman, M. A. (2026). Multiphase system leak detection, localization, and dimension estimation using multi-head Physics-informed neural network. *Process Safety and Environmental Protection*, 211, 108532. <https://doi.org/10.1016/j.psep.2026.108532>
36. Riezzo, L., Kay, H., Feng, Y., Jing, K., & Zhang, D. (2025). Accelerating bioprocess digital twin development by integrating hybrid modelling with transfer learning. *Chemical Engineering Journal*, 511, 162018. <https://doi.org/10.1016/j.ccej.2025.162018>
37. Rohman, F. S., Muhammad, D., Azudin, N. Y., Murat, M. N., Lee, C. S., Termizi, S. N. a. A., Azmi, A., & Shukor, S. R. A. (2026). Machine Learning Assisted-Barnacles Mating Optimization of non-enzymatic isoamyl acetate in a solvent-free system using miniaturized intensified reactor. *Discover Chemical Engineering*, 6(1). <https://doi.org/10.1007/s43938-026-00123-1>
38. Schmid, D., Verlaet, B., Petagna, P., Revellin, R., & Schiffmann, J. (2021). Flow pattern observations and flow pattern map for adiabatic two-phase flow of carbon dioxide in vertical upward and downward direction. *Experimental Thermal and Fluid Science*, 131, 110526. <https://doi.org/10.1016/j.expthermflusci.2021.110526>
39. Shah, P., Liu, J., & Kwon, J. S. (2025). Two-timescale-based hybrid modeling framework for crystallization processes. *Process Safety and Environmental Protection*, 225, 376–386. <https://doi.org/10.1016/j.cherd.2025.12.020>
40. Shen, R., Luo, G., & Su, A. (2025). Bayesian Optimization for Chemical Synthesis in the era of Artificial Intelligence: Advances and applications. *Processes*, 13(9), 2687. <https://doi.org/10.3390/pr13092687>
41. Thebelt, A., Wiebe, J., Kronqvist, J., Tsay, C., & Misener, R. (2022). Maximizing information from chemical engineering data sets: Applications to machine learning. *Chemical Engineering Science*, 252, 117469. <https://doi.org/10.1016/j.ces.2022.117469>
42. Uhlemann, T. H., Lehmann, C., & Steinhilper, R. (2017). The Digital Twin: Realizing the Cyber-Physical production system for industry 4.0. *Procedia CIRP*, 61, 335–340. <https://doi.org/10.1016/j.procir.2016.11.152>

43. Ureel, Y., Dobbelaere, M. R., Ouyang, Y., De Ras, K., Sabbe, M. K., Marin, G. B., & Van Geem, K. M. (2023). Active machine learning for chemical engineers: A bright future lies ahead! *Engineering*, 27, 23–30. <https://doi.org/10.1016/j.eng.2023.02.019>
44. Von Stosch, M., Portela, R. M., & Varsakelis, C. (2021). A roadmap to AI-driven in silico process development: bioprocessing 4.0 in practice. *Current Opinion in Chemical Engineering*, 33, 100692. <https://doi.org/10.1016/j.coche.2021.100692>
45. Windows-Yule, C. R., Brayson, C., & Roy, A. (2026). Artificial intelligence in fluidisation technology: promise, progress and pitfalls. *Current Opinion in Chemical Engineering*, 53, 101293. <https://doi.org/10.1016/j.coche.2026.101293>
46. Wongsaroj, W., Thong-Un, N., Hansot, J., Shoji, N., Treenuson, W., & Kikura, H. (2024). The application of ultrasonic measurement and machine learning technique to identify flow regime in a bubble column reactor. *Sensors International*, 6, 100294. <https://doi.org/10.1016/j.sintl.2024.100294>
47. Wu, G., Yion, W. T. G., Dang, K. L. N. Q., & Wu, Z. (2023). Physics-informed machine learning for MPC: Application to a batch crystallization process. *Process Safety and Environmental Protection*, 192, 556–569. <https://doi.org/10.1016/j.cherd.2023.02.048>
48. Yang, Z., He, Y., Xiang, Q., Zio, E., He, L., Luo, X., Su, H., Wang, J., & Zhang, J. (2022). Gas liquid cylindrical cyclone flow regime identification using machine learning combined with experimental mechanism explanation. *Petroleum Science*, 20(1), 540–558. <https://doi.org/10.1016/j.petsci.2022.09.005>
49. Yuan, Y., Chaffart, D., Wu, T., & Zhu, J. (2024). Transparency: the missing link to boosting AI transformations in chemical engineering. *Engineering*, 39, 45–60. <https://doi.org/10.1016/j.eng.2023.11.024>
50. Zhao, T., Zheng, Y., Gong, J., & Wu, Z. (2022a). Machine learning-based reduced-order modeling and predictive control of nonlinear processes. *Process Safety and Environmental Protection*, 179, 435–451. <https://doi.org/10.1016/j.cherd.2022.02.005>
51. Zhao, T., Zheng, Y., & Wu, Z. (2022). Improving computational efficiency of machine learning modeling of nonlinear processes using sensitivity analysis and active learning. *Digital Chemical Engineering*, 3, 100027. <https://doi.org/10.1016/j.dche.2022.100027>