Engineering 53 (2025) 139–154

Contents lists available at ScienceDirect

Engineering

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

Research
AI in Chemical Engineering—Review

Progress of Machine Learning in Molecular Crystal Design and Crystallization Development

Shengzhe Jia [a], Yiming Ma [a,b], Yuechao Cao [a], Zhenguo Gao [a,*], Sohrab Rohani [c], Junbo Gong [a], Jingkang Wang [a]

[a] School of Chemical Engineering and Technology, State Key Laboratory of Chemical Engineering & Collaborative Innovation Center of Chemical Science and Chemical Engineering, Tianjin University, Tianjin 300072, China
[b] Department of Chemical Engineering, Loughborough University, Leicestershire LE11 3TU, UK
[c] Department of Chemical and Biochemical Engineering, University of Western Ontario, London, ON N6A 5B9, Canada

ARTICLE INFO

Article history:
Received 27 August 2022
Revised 19 January 2025
Accepted 6 March 2025
Available online 4 June 2025

Keywords:
Machine learning
Artificial intelligence
Molecular crystal design
Process optimization
Crystallization control

ABSTRACT

Machine learning (ML) can optimize the research paradigm and shorten the time from discovery to application of novel functional materials, pharmaceuticals, and fine chemicals. Besides supporting material and drug design, ML is a potentially valuable tool for predictive modeling and process optimization. Herein, we first review the recent progress in data-driven ML for molecular crystal design, including property and structure predictions. ML can accelerate the development of the solvates, co-crystals, and colloidal nanocrystals, and improve the efficiency of crystal design. Next, this review summarizes ML algorithms for crystallization behavior prediction and process regulation. ML models support drug solubility prediction, particle agglomeration prediction, and spherical crystal design. ML-based *in situ* image processing can extract particle information and recognize crystal products. The application scenarios of ML algorithms utilized in crystallization processes and two control strategies based on supersaturation regulation and image processing are also presented. Finally, emerging techniques and the outlook of ML in drug molecular design and industrial crystallization processes are outlined.

© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Modeling has played an important role in the design, control, and optimization of processes over the past 130 years [1]. Levenspiel [2] pointed out that modeling is a major tool in chemical engineering. However, before establishing a model, a deep comprehension of the process mechanisms is usually required, which substantially extends the time of model development [3–5]. Recently, engineers have combined modeling with simulations and experiments, using massive data as a toolkit to support predictions derived from experiments. Among the various models, artificial intelligence (AI) algorithms, which simulate human intelligence or behaviors, can efficiently cope with the massive data [6]. Machine learning (ML) is an AI-derived concept through which a machine acquires knowledge from the user and inputs data to make automatic judgments and responses in the actual production. It can assist humans with problem-solving, reduce errors, improve efficiency, and involve the multidisciplinary specialty, covering probability theory, statistics, approximation theory, and complex algorithms (Fig. 1(a)). This AI-derived ML algorithm has emerged in various fields, including neuroscience, computer science, statistics, social science, chemistry, robotics, image analysis, and transportation (Fig. 1(b)) [7–15]. ML can be divided into several categories: supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning (Fig. 1(c)) [16–19]. Proposed in 2006, deep learning refers to analytic learning and data interpretation by neural networks that simulate the human brain [20]. ML algorithms have continuously evolved from linear regression models to convolutional neural networks (CNNs). The characteristics of ML algorithms in the evolution are presented in Fig. 1(d).

Studies on data-driven ML models for process optimization or property prediction in chemistry, materials, and engineering have increased obviously [21]. Dobbelaere et al. [1] concluded the advances of ML in chemical engineering and emphasized the weaknesses, threats, opportunities, and strengths of ML models. They

* Corresponding author.
E-mail address: zhenguogao@tju.edu.cn (Z. Gao).

https://doi.org/10.1016/j.eng.2025.03.036
2095-8099/© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

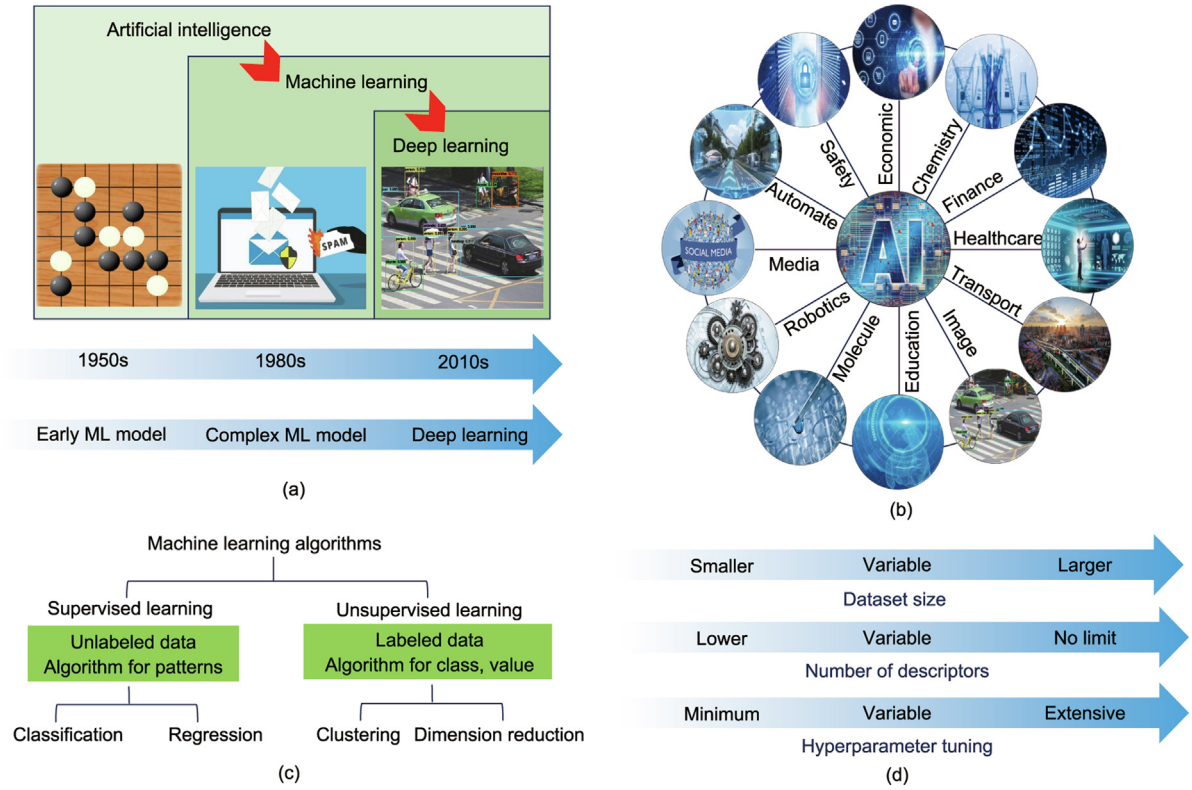

Fig. 1. (a) Evolution, (b) application, (c) algorithms, and (d) characteristics of ML models.

suggested that although ML possesses higher flexibility, accuracy, and execution speed than traditional modeling, its development is impeded by several limitations, such as the insufficient understanding of the underlying phenomenological principles. So, accurate real-time optimization is an important avenue for ML algorithms. Venkatasubramanian [22] reviewed data-driven ML models and highlighted their insufficiencies in providing explanations of mechanisms, developing conceptual frameworks, and discovering domain-specific knowledge. Xiouras et al. [23] summarized the applications of ML algorithms in the prediction of crystalline materials and simulation of the complex crystallization processes. They mentioned the remarkable contributions of ML to high-throughput automation, mathematics, and chemistry. Previously, we reviewed the advancements of AI-based image processing techniques in multiphase flow systems, covering fluidization, crystallization, dissolution, and emulsification [24]. Data-driven analytical technology has delivered outstanding performance in image segmentation and visualization of crystallization processes.

ML modeling integrated with crystallization techniques can substantially accelerate the development of new products. More than 50% of the new materials, food, drugs, agricultural chemicals, and dyes produced by chemical industries are crystalline [25]. Crystallization is a multi-objective and nonlinear processing technique that influences the purity, particle size distribution, and morphology of the product. Moreover, as crystallization involves molecules, particles, crystals, processes, and equipment on different spatial scales, it cannot be developed through theoretical investigation alone [25]. ML strategies can predict crystallization behavior and guide processing control. In particular, ML models can promote green, automated, and intelligent manufacturing. Combining the results of previous reports with our current understanding, this work reviews the recent ML-based developments in molecular crystal design and crystallization control (Fig. 2). Incorporating ML models can improve prediction accuracy and reduce the computational time of these endeavors.

2. Molecular crystal design

Molecular crystal drug design has been rigorously pursued in the pharmaceutical field. Drug polymorphisms have become increasingly important in fundamental research and intellectual property rights. Polymorphism threatens the generic safety of drugs and polymorphic tests increase dramatically with the expiry of numerous original drug patents [26]. In most software for predicting polymorphisms based on molecular mechanisms, the molecular energy is calculated by the Polymorph module, often based on quantum mechanics or molecular mechanics [27,28]. Particularly, the recent emergence of ML has accelerated molecular crystal design. Liu et al. [29] suggested that ML models can simulate crystal growth in multiple fields, resolving the difficulty of synthesizing high-quality single crystals. Here, the recent advances in ML models for molecular crystal design are presented, covering solvate screening and prediction, crystallographic structure prediction, crystal property prediction, and synthesis of co-crystals and colloidal nanocrystals.

2.1. Solvate screening and prediction

As the solvate influences the solubility, stability, morphology, powder flowability, and compressibility of a drug, accidental solvate formation can unpredictably change the properties of the drugs [30]. ML can effectively screen different solvents and predict their solvate effects. Among numerous algorithms, random forest and support vector machines are commonly adopted in multivariate data analysis. Random forest is often applied in the classification or regression problems in chemistry, manufacturing, and biology

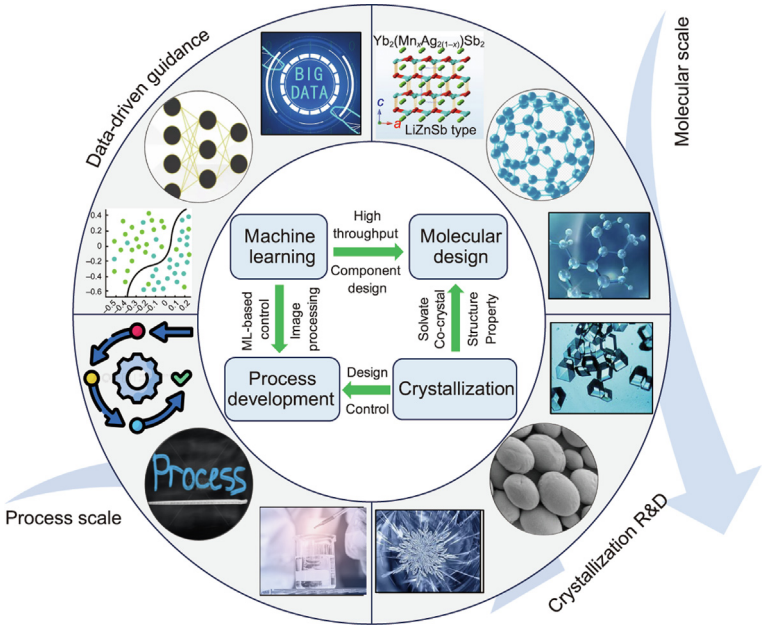

Fig. 2. Outline of ML for molecular crystal design and crystallization development.

[31–33]. Takieddin et al. [34] defined a knowledge-based model that forecasts hydrates and solvates from molecular formulae alone. They collected more than 19 000 datasets of organic, non-ionic, and nonpolymeric molecules from the Cambridge Structural Database and verified the accuracy of model predictions in water, chloroform, dichloromethane, methanol, and ethanol. Takieddin et al. [34] fitted logistic regression models using the corresponding molecular descriptors of the molecule, to evaluate the solvate probability, with more than 80% accuracy rate. Xin et al. [31] predicted the solvate formation probabilities of drug molecules with an accuracy of 86% through the random forest model. ML algorithms can also predict inclusion complexes. Ma et al. [35] predicted the formation probabilities of cyclodextrin inclusion complexes in aqueous solutions using support vector machines, logistic regression, and artificial neural network models. All three models delivered good prediction performance.

Although the existing ML models can effectively predict solvate formation, their prediction accuracy and computational time should be further improved. Unclear correlations between the experimental results and crystal forms restrict the further development of ML algorithms. Therefore, ML models should not be established based on existing databases alone. Experimental results must be combined to excavate the intrinsic correlations between the technical variables and crystal products. The descriptor design should additionally consider the molecular structure and crystal properties of the product.

2.2. Crystallographic structure prediction

Crystallographic structures are predicted from chemical diagrams based on statistics and computational science [36]. Traditional first-principle prediction requires highly accurate energy calculations. Solid-state density functional theory (DFT) is a high-precision computing tool that incurs high computational costs. Molecular dynamics simulations effectively track the trajectories of particles in phase space to explain the physical phenomena [37]. Pèpe et al. [38] forecasted the crystallographic structure of molecules in different solvents using experimentally assisted molecular modeling. After obtaining the chemical formulae of the given molecules, they analyzed the crystallographic structures under the minimum-energy principle. Previously, we analyzed the intermolecular interactions and charge distributions using the Hirshfeld surface and the molecular electrostatic potential surface, and simulated solute–solvent interactions with the solvation-free energy model [39,40]. The radial distribution function can model the hydrogen-bond interactions among molecules. However, these calculation tools are restricted by high computational costs and long runtimes.

ML-based algorithms simplify complex calculations and shorten the time from discovery to application of a product. When used in conjunction with experiments, ML replaces numerous calculations with probabilistic predictions. ML algorithms can predict the crystallographic structure and explore the chemical properties of a given molecule. Egorova et al. [36] replaced the expensive hybrid functional DFT with a statistical ML model and successfully predicted the crystal properties of oxalic acid, maleic hydrazide, and urazole. Bhardwaj et al. [32] predicted the three-dimensional (3D) crystal fillers and crystal forms of olanzapine using a random forest model. Doi et al. [41] used an ML scheme to extract the effective order parameters for classifying the crystallographic structure. Wengert et al. [42] optimized traditional ML algorithms and designed a data-efficient ML routine to forecast molecular structures, which accelerated the screening of crystals by considering the subtle interplay among the molecules.

ML can potentially improve the accuracy of targeted parameters, especially those derived solely from human insight. While, when predicting crystallographic structures, ML algorithms are less successful due to the requirement of massive training sets and specialized knowledge for encoding and algorithm deployment [27]. As no ML algorithm can predict the crystallographic structures of all molecules, an exhaustive search of the parameter combinations is necessary.

2.3. Crystal property prediction

In crystal property analyses, DFT calculations provide the electronic band structures, but their further applicability is limited by poor accuracy and high computational costs [43–45]. ML can

achieve high-throughput prediction; especially, it can map correlations between data when the physical mechanisms are unclear [46]. Increasingly, the design of crystalline materials has relied on the integration of ML with prior knowledge [47]. The entire ML workflow, from establishing the mathematical descriptors to selecting specific ML algorithms, has been already developed [48]. High-throughput, automated synthesis, and testing techniques can now generate large material databases for algorithm training, allowing ML models to predict the properties of crystalline materials and design novel materials [49,50]. As shown in Fig. 3 [51], property prediction has already evolved through three generations. First-generation prediction calculated the physical properties using the Schrödinger equation or by optimizing the locations of atomic forces. In the second generation, the chemical-component inputs were mapped to the outputs of structural predictions containing the adopted combinations of elements after global optimization. However, the first and second generations relied on the physical insight of human experts [52]. The third generation predicts crystallographic structures or properties using ML models and optimizes the prediction process with training data. Pilania et al. [53] discovered the properties of crystalline materials using the kernel ridge regression model. ML models can forecast the crystal plasticities, thermal conductivities, and configurational energies [50,54–56]. Artificial neural networks are used for predicting chemical properties at the atomic level, such as formation energy [57], band gap, Fermi energy [58], and optimal conditions [59].

However, property prediction is a difficult task within the complex space of molecular structures. Mastering the features of crystalline materials and improving the prediction accuracy are especially challenging. To optimize complex chemical parameters, ML algorithms must be combined with design theories of crystalline materials. The rapid development of quantum chemicals and calculations has accelerated material discovery and provided abundant data for ML models.

2.4. Co-crystal formation

Co-crystal screening is a critical and challenging task in drug development. Traditional crystal-design routes have adopted solvent-mediated or solid-based approaches. Solvent-mediated approaches include antisolvent, reaction, and evaporation crystallization, whereas the solid-based approach refers to solvent-assisted grinding [60,61]. Previous prediction strategies mainly depended on the melting point difference between the co-crystal and its pure components, which requires numerous and accurate potential calculations. The internal correlations between the co-crystal and co-formers have been explored through various statistical analyses and algorithms, but no complete prediction model has appeared to date until the existence of the ML algorithms. Wicker et al. [62] reported that ML algorithms can guide the selection of co-formers with simple descriptors, boosting the prediction probability. Mswahili et al. [63] extracted the molecular descriptors of the starting compounds and compared the prediction performance of different ML models. They reported a higher accuracy and sensitivity of the artificial neural network than of other models. Wang et al. [64] also accurately predicted the synthesis of captopril–proline and captopril–sarcosine using ML. Chabalenge et al. [65] synthesized 54 co-crystals through the hot-melt extrusion technique and integrated them into an open-sourced decision tree model for investigating the key parameters of co-crystal formation and enriching the database.

ML models can accelerate the progress and lower the experimental effort of designing molecular crystals. Similarly, co-crystal prediction requires the selection of suitable descriptors. Moreover, statistical analyses coupled with data mining can sufficiently reveal the properties of co-crystal compositions and simplify complex calculations. Data-driven ML techniques are widely prospective for molecular crystal design, predicting the crystallographic structures, properties, and formations of solvates and co-crystals [66,67]. Margraf [68] reviewed data-driven ML algorithms for

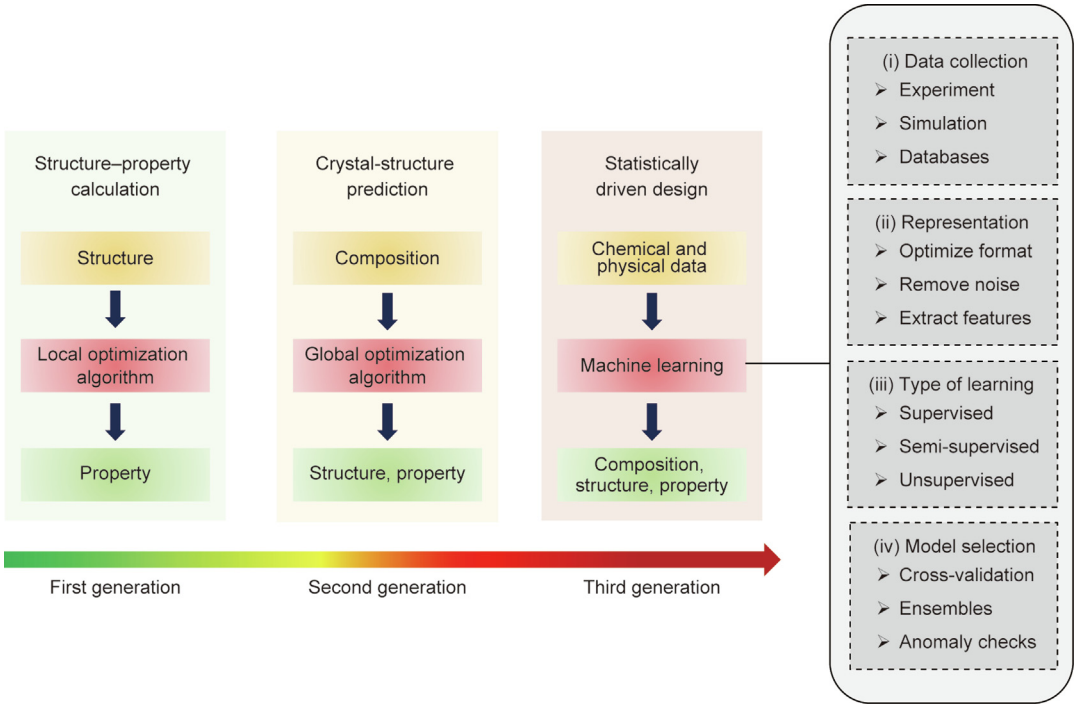

Fig. 3. Three-generation evolution of crystal properties prediction. Reproduced from Ref. [51] with permission.

novel material design and emphasized the importance of model selection on property prediction. Ashraf et al. [69] determined an important role for AI in the discovery of new crystal molecules. Li et al. [70] suggested that ML can be combined with quantum mechanics computation and experimental datasets to forecast molecular energies, crystallographic structure, and properties.

2.5. Synthesis of colloidal nanocrystals

AI strategies have been embraced not only by the crystal engineering field but also in the design and synthesis of colloidal nanocrystals. The unique physical properties, ability to assemble into large structures and size-dependent optical properties of colloidal nanocrystals, have been exploited in electrochemical, biomedical, and optical applications [71,72], such as light-emitting diodes [73], electronic devices [74], and luminescent labels [75]. Chen et al. [76] summarized the synthesis of colloidal nanocrystals, prospecting the emerging applications and prospects of these fields.

However, the complex growth process raises the difficulty of regulating the properties or structures of colloidal nanocrystals and correlating them with the operational parameters. Recently, AI integration with robotic synthesis techniques has substantially reduced labor requirements [77,78]. Zhao et al. [79] designed a robotic platform that controls the morphologies of colloidal nanocrystals. The first step of this platform excavates the data and parameters of materials from the literature. The following step automatically and controllably synthesizes the desired morphologies. The final step imports ML algorithms that identify the relations between the morphologies and structure-guiding agents after training on an experimental database. This new data-driven robotic synthesis framework effectively reduces the trial-and-error costs, promising the digital synthesis of nanocrystals. Chan et al. [80] established an automated platform that synthesizes colloidal nanocrystals with high-throughput optimization. The robotic platform can precisely control the reaction conditions and complete the complex synthesis process. The accuracy and utility of the platform were verified by the small coefficient of variation of the mean diameters (0.2%) of the nanocrystals. The size, crystal phase, and polydispersity of the nanocrystals were precisely controlled, substantially improving the unconverted luminescence compared with the traditional experimental methods. This automated synthesis platform greatly accelerates the design of colloidal nanomaterials. Overall, combined robotic and AI techniques promise the synthesis of colloidal nanocrystals with controllable morphology and size while reducing time and labor costs.

3. Crystallization development

Besides molecular crystal design, ML algorithms have demonstrated potential in process design and robust control. Herein, the application scenarios of ML models in the design, optimization, and control of crystallization are presented.

3.1. Crystallization behavior

3.1.1. Crystallization propensities

Industrial crystallization urgently requires cheaper and cleaner production technologies providing safer and more efficient solutions. Crystallization is widely applied in high-purity chemical production and target crystal product design. However, selecting the appropriate crystallographic structure, thermodynamic, and kinetic parameters for crystallization incurs considerable time costs [81–83]. ML models provide guidance for the prediction of crystallization propensity and behavior.

Venkatram et al. [84] developed a data-driven ML model that predicts the crystallinity of polymers based on experimental and theoretical group-contribution strategies. Adopting a multi-fidelity information fusion approach, they accurately predicted the crystallization tendency at a low cost (Fig. 4(a)) [84]. Ghosh et al. [85] evaluated the crystallization propensities of small organic compounds using support vector machine regression, neural networks, and random forest regression, integrating molecular descriptors, training sets, and experimental conditions to optimize the prediction accuracy. On a small training set, the three algorithms achieved similar accuracies but on a training set with more than 150 members, random forest regression outperformed the other algorithms. Ghosh et al. [85] then compared the prediction performances of an active pharmaceutical ingredient (API)-only model (Fig. 4(b)) and an API + solvent model (Fig. 4(c)). The prediction performance of both models improved after excluding chemical degradations and impurities, and the API + solvent model is more outstanding on larger training sets. Applying ML algorithms, Pereira [86] predicted the crystallization propensity from the fingerprints of two-dimensional (2D) and 3D chemical descriptors. ML algorithms have proven their worth in crystallization propensity prediction, especially in pharmaceutical projects.

3.1.2. Solubility and the metastable zone

Solubility is an important parameter in drug development as it determines the synthesis, process design, and formulation of a crystal product [87,88]. Drug solubility determines the formation of co-crystals, solvates, and the efficacy of drug delivery [40,89]. Therefore, approaches that forecast drug solubility in different solvents can save appreciable time and cost [90–93].

Previously, we trained random decision forests and artificial neural networks on the solubility data of 120 APIs [94]. The trained ML algorithms were then applied to dataset analysis and prediction model optimization. The ML algorithms outperformed traditional models in terms of accuracy and prediction ability. Multiple linear regression and stepwise regression were then integrated to analyze the key elements of solubility. Ge and Ji [95] combined molecular thermodynamic simulations and predicted drug solubilities using a single-hidden-layer neural network. They described drug–drug and drug–solvent interactions using 16 molecular descriptors input to five ML algorithms, which accurately predicted drug solubility. Boobier et al. [87] combined ML models with computational chemistry for solubility predictions in organic solvents and determined the physicochemical correlations between solubility and molecular properties.

The metastable zone width (MSZW) is another important parameter. The MSZW depends on the experimental conditions and raw material properties, which directly determine the operational regimes [96–99]. We previously predicted the MSZW of a reaction crystallization process using an artificial neural network [100]. Among several parameters (concentration, reaction volume, stirring speed, and temperature), temperature and concentration were strongly positively correlated with the MSZW.

3.1.3. Crystalline particle agglomeration prediction

Agglomeration directly impacts the particle size distribution, crystal shape, and purity of crystal products, thus influencing the downstream processing. For example, agglomeration is desired in the wet granulation process but undesired in the filter or drying process. Complex interactions between the solvent and solute increase the difficulty of quantifying or controlling the agglomeration degree [101,102], but image processing is useful for detecting and quantifying agglomerates [103,104]. Ochsenbein et al. [105] identified agglomeration behaviors using a nonlinear support vector machine and image analysis strategy, defining different dimensions, such as one-dimensional (1D) particle size distribution of

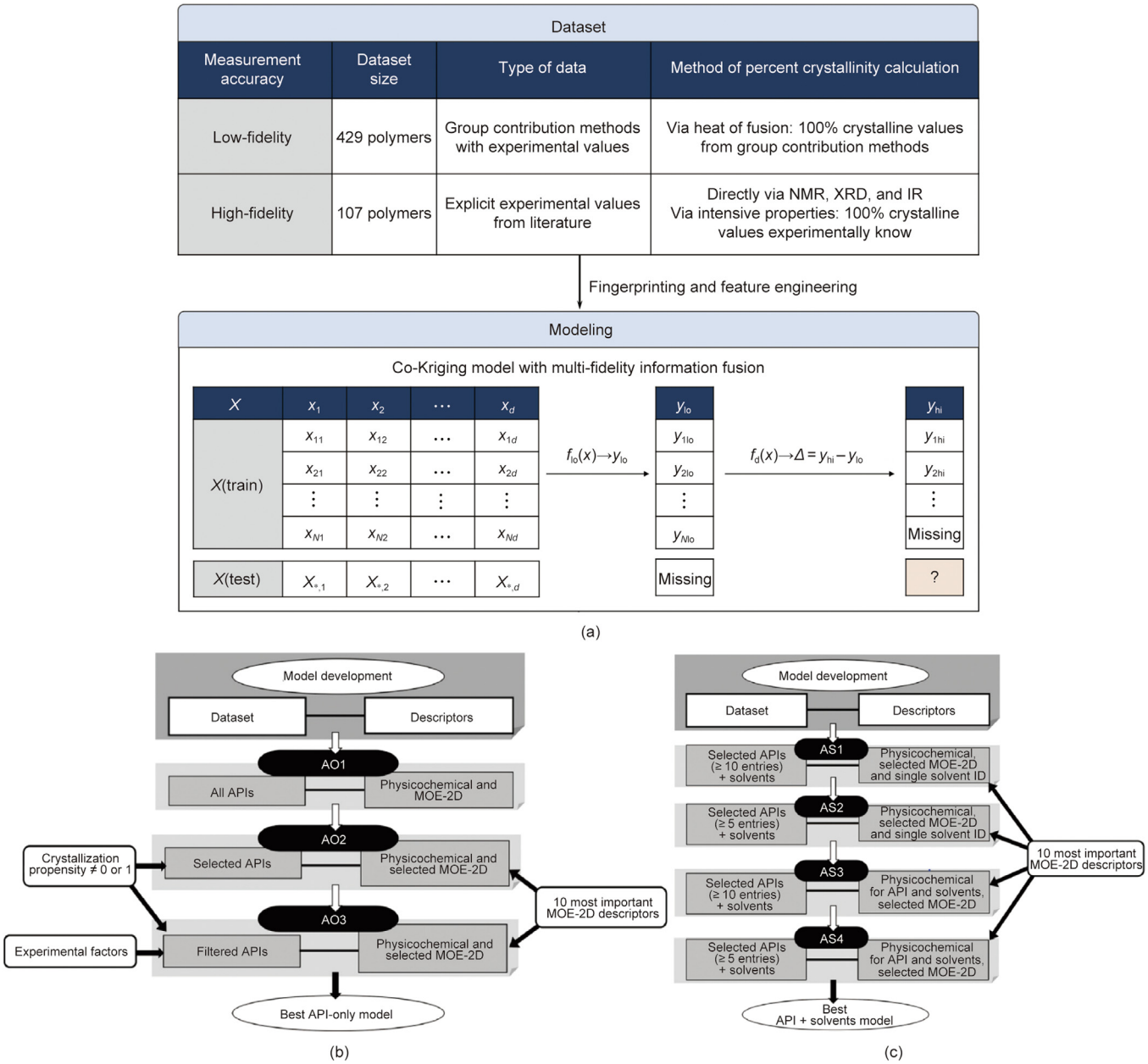

Fig. 4. Workflows of (a) crystallization tendency prediction, (b) API-only model, and (c) API + solvent model. X(train): training samples; X(test): test samples; $f_{lo}(x)$: low-fidelity function; y_{lo}: output from the low-fidelity function; $f_{hi}(x)$: high-fidelity function; y_{hi}: output from the high-fidelity function; N: the number of properties; d: data point in the test or training samples; x: training data point; X_*: test data point; Δ: the results difference between the high-fidelity function and the low-fidelity function; NMR: nuclear magnetic resonance; XRD: X-ray diffraction; IR: infrared; AO: API only; MOE: Molecular Operating Environment; 2D: two-dimensional; AS: API + solvent; ID: identification. (a) Reproduced from Ref. [84] with permission; (b, c) reproduced from Ref. [85] with permission.

agglomerates and 2D particle size and shape distributions of primary crystals, to evaluate the particle properties and crystal growth behavior. Lins et al. [106] applied ML to the evaluation of crystalline particle properties. Deep learning-based image processing can effectively describe particle agglomeration and detailed descriptions of the agglomeration level. Sinha et al. [107] proposed a sticky zone-based agglomeration model (SZAM) (Fig. 5(a)) that predicts the degree of particle agglomeration by coupling the agglomeration risk zone and sticky zone with the discrete element method. The bed is classified into various 3D Cartesian bins or voxels, and each is supposed in a microenvironment. Combined with ML models, the agglomeration degree can be described as a functional of input variables, such as the material properties and processing conditions.

Particle agglomeration leads to spherical crystal formation. Commonly, spherical crystals possess larger specific surface areas, higher bulk densities, and higher fluidities than needle- and plate-shaped forms [40,89]. Previously, a spherical crystallization technique was designed based on liquid–liquid phase separation theory and ML models [108]. The model involves three steps (Fig. 5(b)) [108]. First, 149 API structures and three solvent systems (water, ethanol, and water–ethanol) are selected as model compounds, and 21 descriptors are extracted from chemical-structure descriptions and solute–solvent interactions to evaluate the particle behaviors. Next, the liquid–liquid phase separation processes are predicted by three ML algorithms: logistic regression, artificial neural network, and support vector machine. The algorithms are evaluated in terms of the accuracy and recall metrics. Finally, the

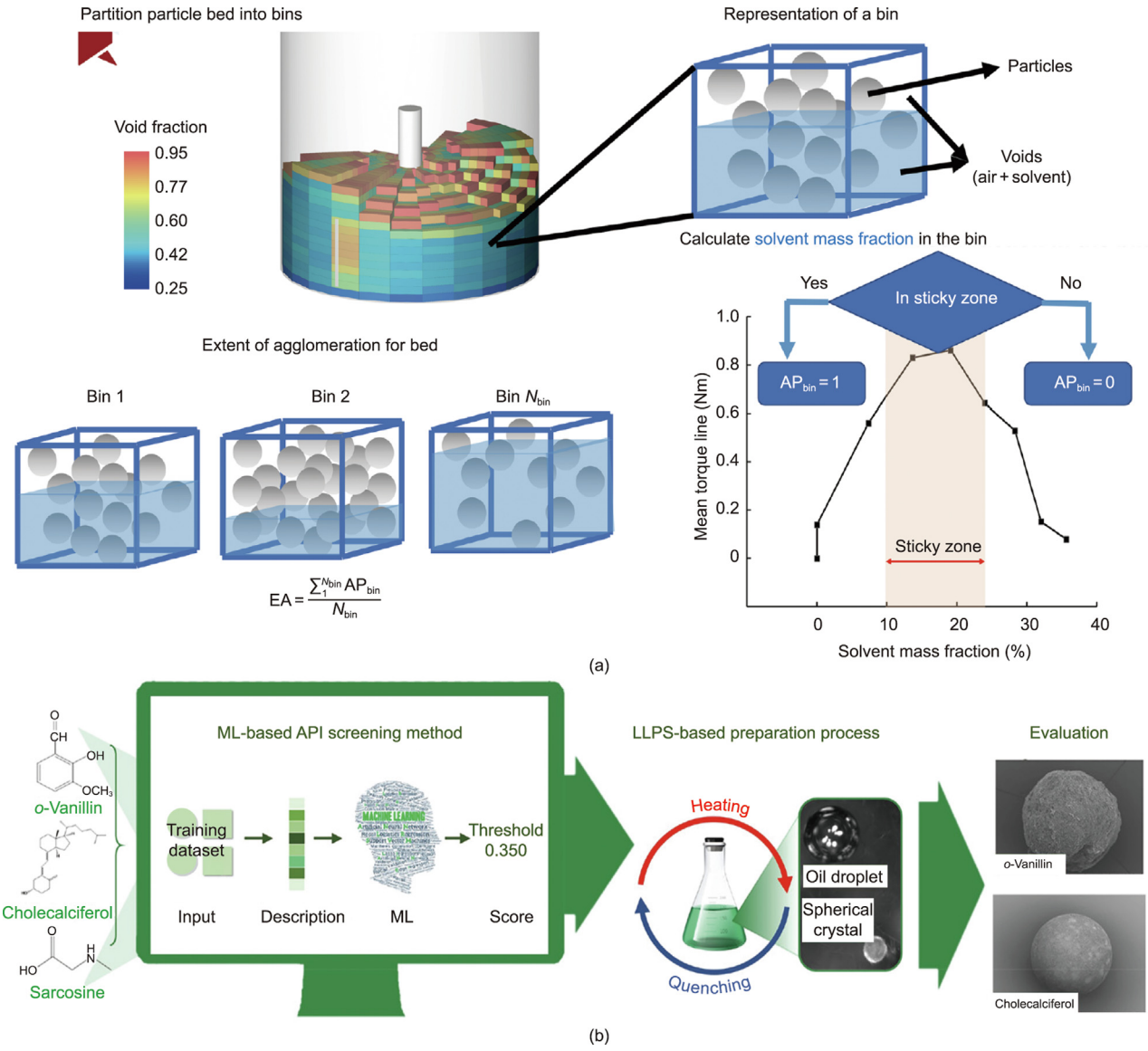

Fig. 5. (a) Flowchart of the SZAM and (b) design of ML-based spherical crystallization. N_{bin}: the total number of bins; EA: the extent of agglomeration; AP_{bin}: the agglomeration propensity of the bin; LLPS: liquid–liquid phase separation. (a) Reproduced from Ref. [107] with permission; (b) reproduced from Ref. [108] with permission.

ML algorithms for designing spherical crystallization are successfully validated on vanillin, cholecalciferol, and sarcosine.

In summary, relevant parameters and descriptors should be selected based on experimental results, to improve the comprehensibility and generalization performances of ML models [109–111]. However, it is challenging to optimize various parameters of precision, recall, and accuracy in the ML model for a single algorithm. Additionally, it may introduce errors in the sample classification when lying near the decision boundary [64,86]. Hence, ML-based design strategies must consider the requirements and application scenarios of the targeted crystallization process.

3.2. ML-based online crystal-image processing

In the pharmaceutical and fine chemical industries, online monitoring of particle shape and size distribution is challenged by complex processes and the lack of adequate *in situ* sensors. The emergence of real-time imaging hardware and process analysis technology (PAT) has sparked the development of advanced online control techniques [112]. As highlighted in several studies, image processing techniques are necessary for online control of crystallization, which required timely extraction of the crystal information.

Large images usually increase the difficulty of analysis and processing in real-time crystallization monitoring. ML algorithms can feasibly accelerate image recognition. Doan et al. [113] combined multivariate image analysis and AI algorithms for analyzing solution crystallization. Image processing techniques suitable for object detection include image enhancement, edge detection, morphology operations, and feature extraction techniques. Previously, we proposed a robust nucleation tracking technology for monitoring the nucleation of *L*-glutamic acid [114]. Combined with image analysis tools and ML models, we successfully tracked *in situ* crystallization and detected the product information of crystal habit, polymorphism, and particle size distribution [115]. Compared with traditional analysis techniques, such as Raman and focus beam reflection measurement, deep learning-based image analysis techniques can enhance the accuracy of segmentation and classification, and monitor individual particles or crystals. Table 1 summarizes the main differences between image analysis and

Table 1
Comparisons between image analysis and instrumental measurements.

Item	Image analysis	Instrument measurements
Crystal size	Yes	Yes
Crystal count	Yes	Yes
Crystal shape	Yes	Limited
Surface area and texture	Yes	No
Agglomeration	Yes	No
Working environment	Image-based	Assisted signals
In-line and off-line	Yes	Yes
Slurry density	Limited	Wide range
Detection limit	1–2 μm	0.5 μm

instrumental measurements using lasers, ultrasonic irradiation, and other sources. Image analysis techniques can obtain more information on the crystal products and can adapt to more complex environments than instrumental techniques.

We have reviewed high-efficiency image segmentation via deep learning strategies involving image classification, object detection, semantic segmentation, and instance segmentation, and concluded the main characteristics of artificial neural network algorithms [24]. Image recognition strategies that employ CNN algorithms have proven useful for analyzing sustainable emulsion manufacturing and crystallization processes (Fig. 6(a)) [116]. Mask region-based convolutional neural network (Mask R-CNN) is a modified CNN algorithm with high universality, good flexibility, and fast segmentation ability, containing the following procedures: database establishment, labelling, training, image processing, mask processing, and process analysis (Fig. 6(b)).

Table 2 [117–128] lists recent cases of ML-based image processing in crystallization development. ML-based image processing techniques effectively recognize the object characteristics and monitor the crystallization process. Compared with other algorithms, neural network models are more often combined with image processing techniques for crystallization monitoring.

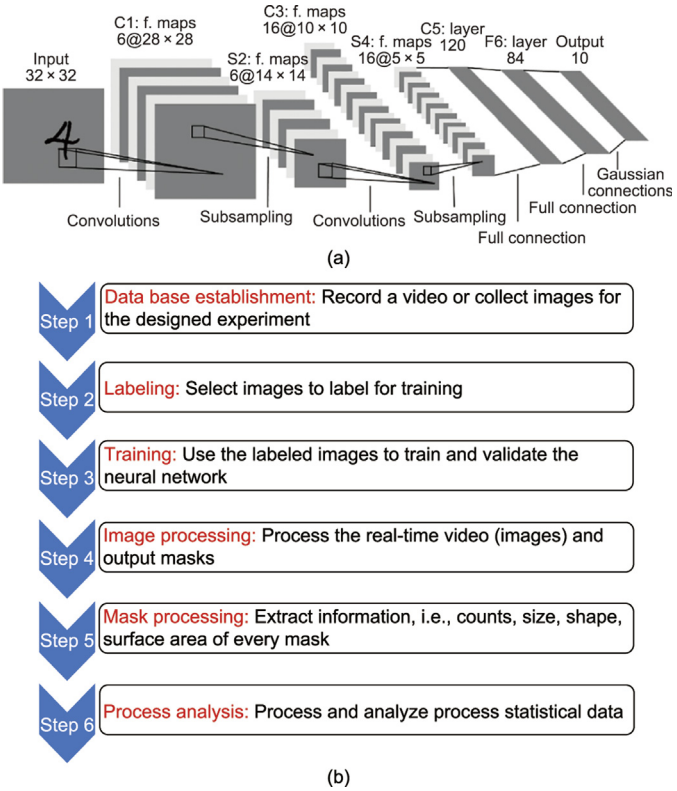

Fig. 6. (a) Example of a CNN implementation. (b) Procedures of Mask R-CNN. C1, C3, and C5 are convolutional layers; S2 is subsampling layer; F6 is fully connected layer; and f. maps stands for feature maps. (a) Reproduced from Ref. [116] with permission.

Table 3 summarizes the main differences between deep learning-based and other algorithms employed in image processing. Deep learning-based models increase the processing capacity, accuracy, and processing speed from those of other algorithms.

Although ML-based algorithms can accurately classify images and extract the crystal size information, they cannot easily identify agglomerations and long-needle crystals when lied in highly concentrated solution. It is also challenging to extract the 3D information of products in complex crystallization systems. These limitations would be mitigated by an advanced imaging device and an image recognition database covering information on crystals and particles. For example, the camera angles could be optimized to maximally capture the 3D particle information [24].

3.3. ML-based control of crystallization processes

The monitoring and precise control of crystallization processes is restricted by complex mechanisms, nonlinear and stochastic crystallization dynamics, insufficient evidence of crystal nucleation, and a lack of real-time crystal information [129–132]. PAT tools provide product-analysis and online-monitoring support, efficiently resolving the insufficiency of real-time information during crystal growth [132,133]. Data-driven monitoring and control are important for optimizing crystallization processes.

3.3.1. Crystallization control strategies

Process analysis techniques are widely adopted in crystallization control, process optimization, and industrial automation. The development of PAT technology has remarkably advanced the online monitoring of crystallization, allowing accurate measurements of solution properties, characterization of crystal quality, simulations of crystallization, and process control. The guidelines of drug crystal development have changed from "quality by test" to "quality by design" and are moving toward "quality by control" with the emergence of intelligent manufacturing (Fig. 7(a)) [134]. Similarly, the production mode has changed from batch to continuous and finally to smart manufacturing. Process control defines the regulation of technical variables to obtain the desirable outputs in real-time while ensuring safe and stable operation. Crystallization control can be divided into model-based control, model-free control, or hybrid control strategies (Fig. 7(b)). Model-based control, including open-loop optimal control and model predictive control, is restricted by the complexity of crystallization processes and obscure kinetic mechanisms. The main model-free control strategies are supersaturation and direct nucleation control [135,136]. Various PAT tools, such as focused beam reflectance measurements, particle video microscopes, and attenuated total reflectance–Fourier transform infrared (ATR-FTIR) spectroscopy, can be integrated into crystallization [137]. Fig. 7(c) [138] shows the monitoring of polymorphic concentration via online tools. The polymorphic concentration can be directly detected from the Raman and ultraviolet/visible spectra and applied as feedback control. Traditional control models are time-consuming and easily affected by disturbances. AI-based models can improve the performance of process control. Dutta and Upreti [139] depicted a map of keywords in which neural networks and process control are the two most notable labels and are highly connected. Particle swarm optimization, machine learning, and fuzzy control form another notable group of keywords. Among the various algorithms, an artificial neural network is commonly used in complex crystallization processes.

3.3.2. ML-based control of crystallization

Crystallization can be divided into batch and continuous operations. Continuous crystallization is more easily controlled and

Table 2
Cases of ML-based image processing for crystallization monitoring.

Algorithms or models	System	Highlights	Ref.
Artificial neural network	Sucrose	Calculate particle size distributions; measure the kinetics parameters	[117]
Regional CNN	*L*-Glutamic acid	Monitor the quality of crystal growth; calculate the crystal growth rate	[118]
Full CNN	Calcium oxalate	Improve the recognition rate; increase the recognition speed	[119]
Deep CNN	Sugar	Recognize crystallization images; monitor crystallization	[120]
	Macromolecular	Classify and label crystallization outcomes	[121]
CNN	Biomolecular	Achieve high-density screening	[121]
	Sodium chloride	Enhance the monitoring quality; reduce the imaging processing time; contribute to process control	[122]
	Particle	Extract the particle features	[126]
Artificial neural network, regression tree, and support vector machine	Plate-like particle	Characterize particle size and shape	[123]
Logistic regression model	Protein	Classify protein crystallization images; reduce the number of high-risk false negatives	[124]
Multilayer perceptron neural network	Protein	Improve image recognition; reduce the number of missing crystals	[125]
Decision tree	Nanocrystal	Control the morphology of nanocrystals	[127]
Single-shot alignment network	Taurine	Estimate the size distribution and crystal counts	[128]

Table 3
Comparisons between the deep learning-based and other algorithms used in image processing.

Item	Deep learning-based algorithms	Other algorithms
Segmentation	Available	Available
Accuracy	High	Limited
Classification	Available	Limited
Detection limit	Several microns	Several microns
Processing speed	High (< 0.1 s)	Low (several seconds)
Transfer application	Easy	Hard

achieves higher productivity and better consistency than batch processes, but is comparatively inflexible and requires a complex operational environment. Batch crystallization, commonly used in the pharmaceutical industry, is a labor-intensive process. Control strategies can be integrated with AI algorithms to improve the efficiency of crystallization processes. AI can direct the control process and optimize the operation with low disturbance [22]. Table 4 [140–146] summarizes typical AI algorithms adopted in chemical engineering. Artificial neural networks and reinforcement learning algorithms are more adaptable to disturbances than expert systems and fuzzy logic. Scientists have realized that hybrid algorithms can effectively avoid the limitations of single AI models

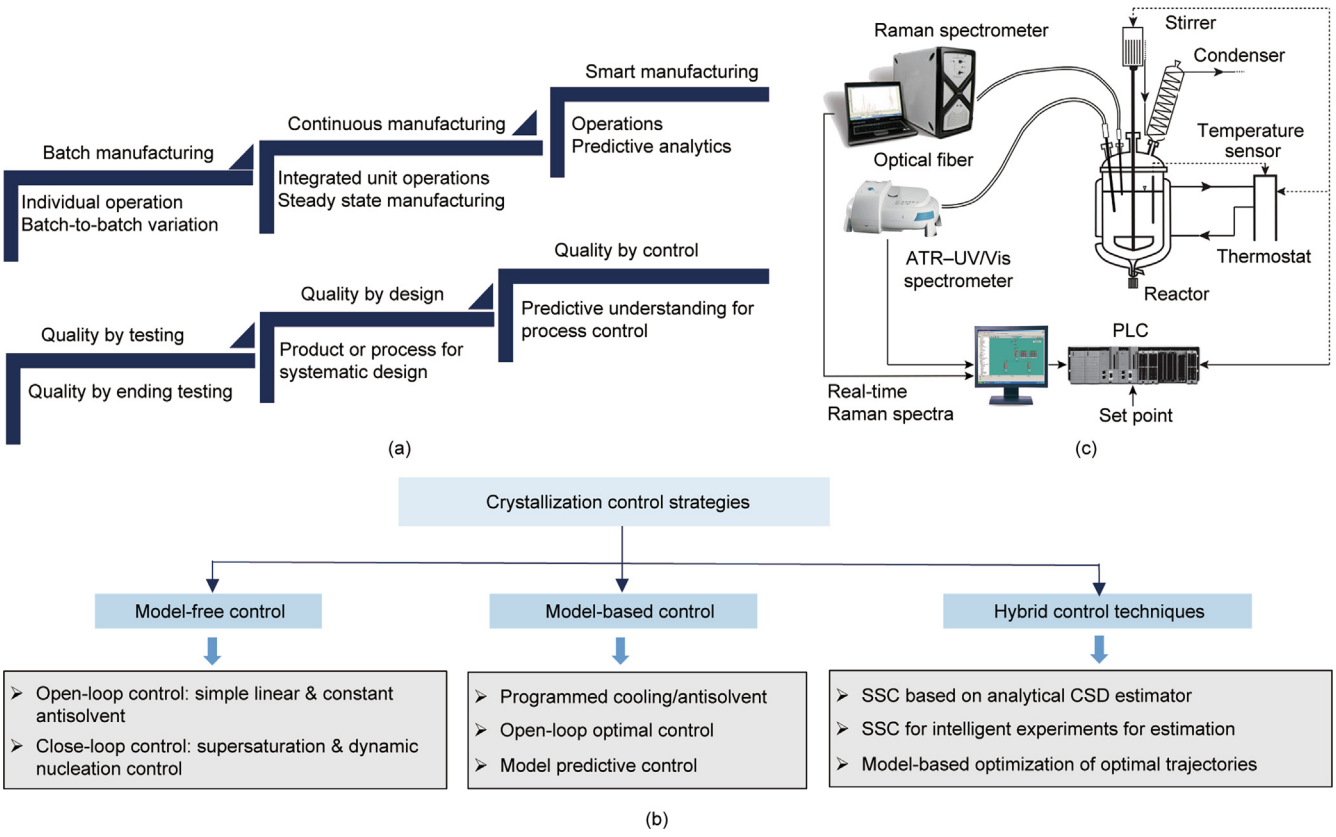

Fig. 7. (a) Guidance for developing crystal drug products, (b) crystallization control strategies, and (c) process control using PAT tools. ATR: attenuated total reflectance; UV/Vis: ultraviolet/visible; PLC: programmable logic controller; SSC: supersaturation control; CSD: crystal size distribution. (a) Reproduced from Ref. [134] with permission; (c) reproduced from Ref. [138] with permission.

Table 4
AI algorithms used in the chemical engineering field.

Technique	Highlights or features	Refs.
Expert system	Establish a base for obtaining input/output correlations; costly in the event of new data	[140,141]
Fuzzy logic	Based on multivalued logic; costly in the event of new data	[142]
Artificial neutral network	Produce knowledge from real-time data; quickly search for an optimal solution	[143,144]
Nature-inspired algorithm	Based on natural behaviors; improve the predictive performance	[145]
Reinforcement learning	Based on animal learning and disturbance adaptive control; provide optimization results	[146]

[139]. The main hybrid algorithms are evolutionary reinforcement learning, hybrid neuro-fuzzy systems, hybrid neural networks/fuzzy logic, and nature-inspired algorithms [147–149]. All these data-driven control strategies can efficiently regulate crystallization [150–152].

(1) **Concentration-based regulators.** An AI-based concentration-control model can efficiently adjust the supersaturation level of the solution environment. Zhang et al. [135,136] proposed a feedback-control strategy that regulates the feed flow rate of semi-batch crystallization to modulate the supersaturation level. In our previous work, a supersaturation control strategy is proposed to optimize the crystal morphology of *p*-aminobenzoic acid and reduce the length-to-diameter ratio of the crystals [115]. The main technical variables are the feeding modes, solution temperature, and flow rate of the antisolvent. Kamaraju and Chiu [153] designed a novel concentration-control strategy that measures the process variables of semi-batch crystallization in real-time. They simulated the solution concentration with just-in-time-learning modeling and predicted the product quality using nonlinear multiway partial least squares. Fig. 8 [154] displays a fuzzy logic-based feedback controller in which the solution concentration and cord length counts are measured by ATR-FTIR and LasenTec. The temperature and flow rate of the antisolvent are regulated based on the solution supersaturation [154–156].

AI-based supersaturation control has become a promising pathway in crystallization processes. In particular, PAT tools can improve the sensitivity of the supersaturation determination in real-time. Paengjuntuek et al. [157] regulated the batch crystallization of potassium sulphate as a model compound under neural network-based optimal control. Because the control framework is integrated into the neural network model, the supersaturation can be regulated by tuning the operating temperature. Damour et al. [158] developed an artificial neural network for nonlinear predictive control of crystallization processes. They controlled the suspension density by regulating the feeding process. Georgieva and de Azevedo [159,160] discussed the use of artificial neural networks in the modeling and optimization of a batch crystallization process. They designed an AI-based control module, by integrating an artificial neural network as a computational tool, for batch crystallization. Combined with the nonlinear dynamic process modeling and model-based predictive control, the operational parameter feed flow can be urged along the optimal supersaturation trajectory. Guo et al. [161] optimized the temperature and supersaturation level in a semi-batch crystallization process. They combined dynamic time warping with a CNN to optimize fault detection and diagnosis in the unsteady batch chemical process.

(2) **Image-based analysis.** With the exception of supersaturation control, image recognition-based ML modeling can support the control and optimization of crystallization processes. Manee et al. [162] designed an AI-based imaging-process control sensor that monitors crystallization and recognizes crystal properties in real-time. Their model can detect crystals in high-density slurries (Fig. 9(a)). A reinforcement learning strategy successfully controls the crystalline particle size in the antisolvent crystallization [163]. Manee et al. [122] applied ML strategies to the closed-loop control process, combining a CNN with a reinforcement learning framework to optimize the process control. This AI-based sensor monitors the crystal size distribution and reduces the imaging processing time. In addition, the reinforcement learning model can successfully design the crystallization under disturbances. The *in situ* analysis and image-based control system is shown in Fig. 9(b) [24]. Based on the dimensions of the crystalline particles obtained from the images, crystallization can be regulated and the operational parameters can be optimized. Öner et al. [164] measured the temperature dependence of chord length distribution during batch crystallization in real-time using a network

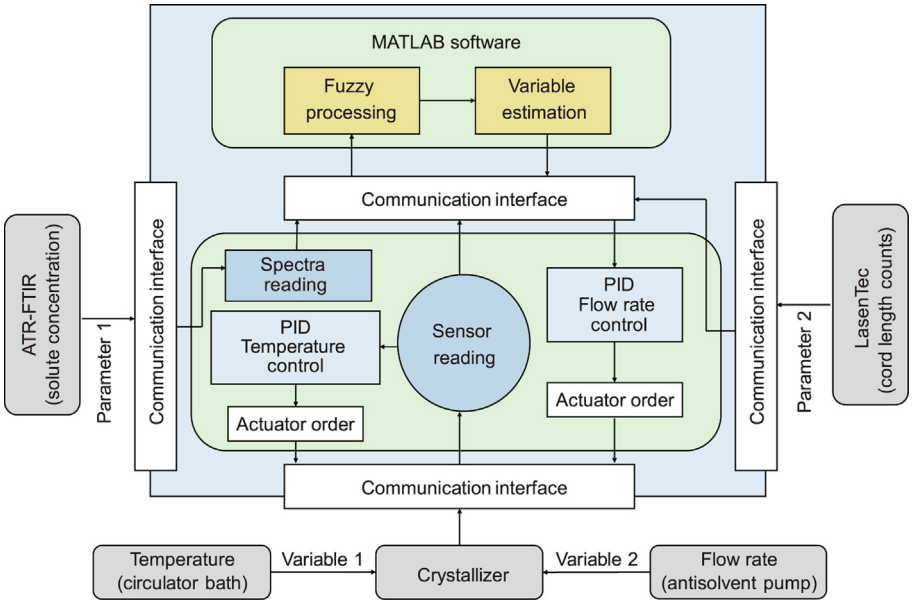

Fig. 8. Example of a feedback controller. PID: proportional–integral–derivative controller. Reproduced from Ref. [154] with permission.

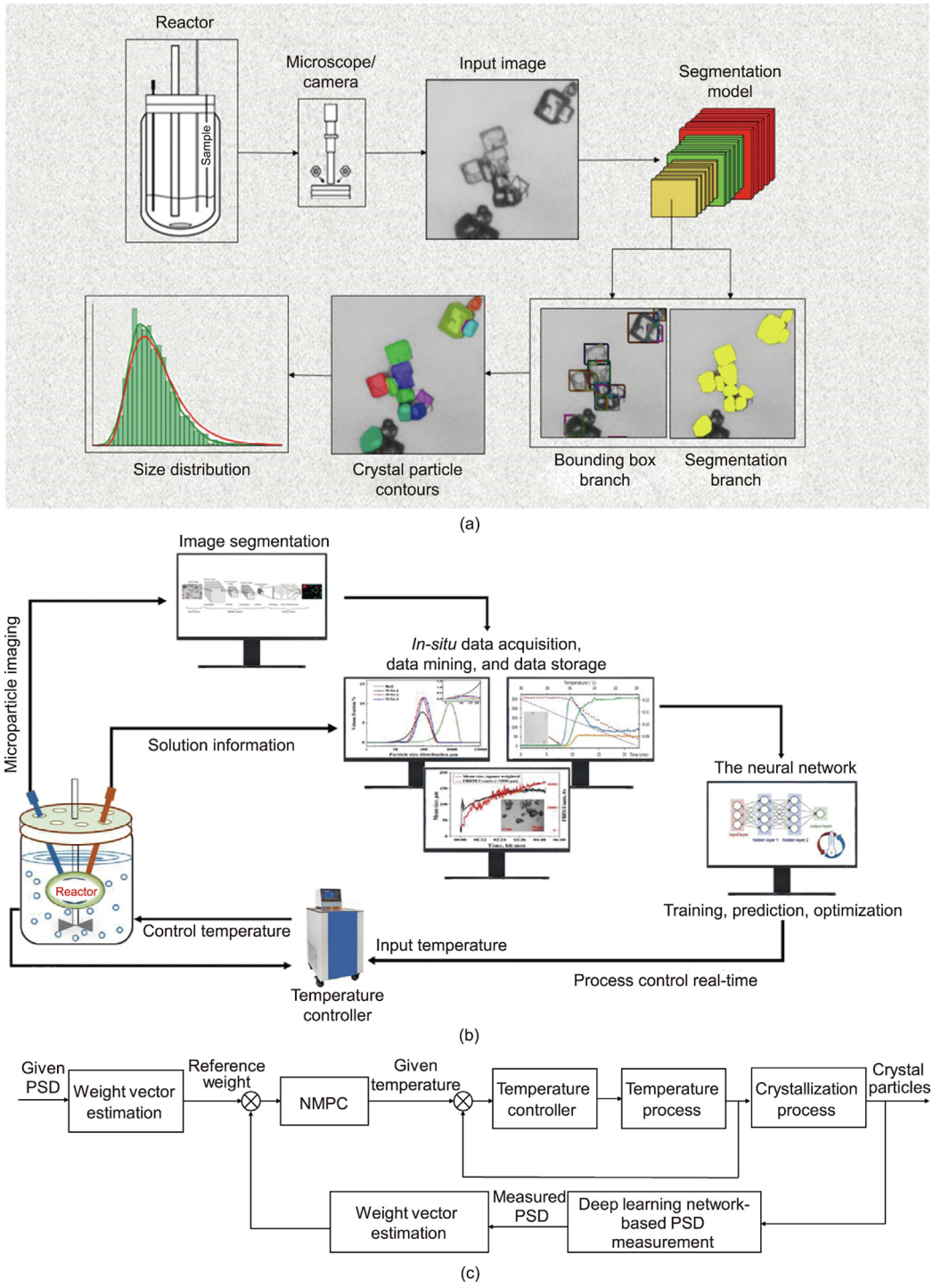

Fig. 9. (a) Workflow of a deep learning image-based sensor. (b, c) Schemes of (b) *in situ* process analysis technique and (c) feedback control system. PSD: particle-size distribution; NMPC: nonlinear model predictive control. (a) Reproduced from Ref. [162] with permission; (b) reproduced from Ref. [24] with permission; (c) reproduced from Ref. [165] with permission.

model with a radial basis function. Even in the absence of comprehensive experimental data, the proposed networks can optimize crystallization by manipulating the operational temperature. Wang et al. [165] applied deep learning-based image analysis to the feedback control of crystallization (Fig. 9(c)). They designed a deep learning-based image processing to control the particle size distribution. Improving the accuracy and efficiency of online process control is vitally important. In addition, timely handling of crystallization data and provision of accurate feedback are also important. Future research should focus on advanced online-monitoring devices and their integration into efficient ML algorithms for improved optimization and control of crystallization. The advantages and disadvantages of supervised/unsupervised ML algorithms for crystallization, and several representative examples are summarized in Table 5 [31,65,166–188].

Table 5
Comparisons of supervised and unsupervised learning used in crystallization.

Algorithm	Characteristics	Case	Refs.
Supervised learning			
Linear regression	Interpretability and simplicity	Predict the stabilities of crystallographic structures	[166]
	High computational efficiency	Design photonic crystals	[167]
	Sensitivity to outliers	Predict solubilities	[168,169]
Artificial neural network	Solve both classification and regression issues	Optimize the conditions of crystal growth	[170]
	Fast and accurate predictions	Monitor the crystallization of sugars	[171]
	Poor interpretability and numerous data	Detect crystals and inclusions	[172]
Decision tree	Low data demands	Screen the conditions of crystal growth	[173]
	Readily comprehended and easily visualized	Design co-crystallization processes	[65]
	Long training time, prone to overfitting	Direct single-crystal growth	[174]
Support vector machine	Insensitive to outliers	Predict the propensity of protein crystallization	[175]
	Avoid overfitting problems	Predict the properties of phosphate crystals	[176]
	Poor performance on large datasets	Predict solvate formation	[31]
k-Nearest neighbor	Easily understood and adaptable	Optimize the perovskite crystallization conditions	[177]
	Require homogeneous features	Estimate numbers of photonic crystal fibers	[178]
	Poor performance on imbalanced data	Adjust the parameters of the reactive force field	[179]
Unsupervised learning			
Clustering algorithm	Good performance on small datasets	Extract crystallographic structure information	[180]
	Good scalability, easily understood	Quantify the size distributions	[181]
	Large deviations on high-dimensional datasets	Determine the structures of crystalline materials	[182]
Principal component regression	Suitable for highly correlated datasets	Control the crystallization supersaturation	[183]
	Analyze multivariate collinearity problems	Select the conditions of batch crystallization	[184]
	Require more variables to improve accuracy	Predict the lattice parameters of perovskites	[185]
Dimensionality reduction	Simplify the variables and calculations	Assess the purity of crystals	[186]
	Extract information from large datasets	Detect the crystal nucleation process	[187]
	Poor treatment efficiency	Select solvents	[188]

4. Conclusions and outlook

ML has been widely utilized in the design of novel functional materials and predictive models for crystallization optimization. ML algorithms can predict the properties of products, reveal structure–function correlations, and shorten the time from discovery to application. Herein, we summarized the recent progress of ML in molecular crystal design and crystallization development. ML has modified the traditional design methods based on first-principles calculations and achieved robust process control of crystalline products.

We first discussed the use of ML in the design of multicomponent crystalline products, covering the formation of co-crystals and solvates, crystallographic structure and property predictions, and the synthesis of colloidal nanocrystals. Although multicomponent crystalline products are promising agents in drug development, their complex interactions are poorly understood and efficient screening strategies are lacking. Therefore, the design of such materials is limited in scope. ML-based molecular design strategies greatly reduce the uncertainty and operational costs of traditional approaches based on experiments and simulation calculations. Next, we reviewed applications of ML algorithms in crystallization development. ML models can predict drug solubilities, the MSZW, particle agglomeration behaviors, and spherical crystal products. ML-based online image processing can extract information about crystal products and track crystallization *in situ* using PAT tools. The features of ML-based image analysis and comparisons with other algorithms were also given. Finally, AI-based process control strategies were summarized. AI-based concentration control can detect supersaturation of a solution and image-based control strategies can capture the information of particles. Both feedback control routes effectively regulate the crystal product properties by changing the operational parameters. In further studies, the following issues should be paid attention.

(1) Data-driven ML algorithms are based on probability and statistics. They are black box models without sufficient professional knowledge. Although ML is potentially suitable for multiscale nonlinear mapping, huge-data association mining, and decision-making, ML models are incapable of deep logical reasoning and making physical connections. Future ML models must be constructed based on physical constraints and experimental results.

(2) Researchers often select the best-performing ML models for drug discovery and materials-science testing. However, the best regression or classification algorithm for a given application scenario is difficult to construct in a single ML model. To improve the predictive performance of ML models, developing specific ML-based routes and coupling efficient algorithms are required.

(3) ML-based online crystal-image processing may be unsuitable for complex systems containing particle aggregations, needle-like crystals, or bubble flow solutions. Therefore, designing advanced online-monitoring devices and expanding the sizes of databases are also important.

CRediT authorship contribution statement

Shengzhe Jia: Writing – original draft, Validation, Methodology, Investigation. **Yiming Ma:** Validation, Investigation. **Yuechao Cao:** Validation, Investigation, Conceptualization. **Zhenguo Gao:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Sohrab Rohani:** Writing – review & editing, Methodology, Conceptualization. **Junbo Gong:** Writing – review & editing, Supervision, Conceptualization. **Jingkang Wang:** Writing – review & editing, Resources, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (22008173, 21938009, and 21676179), and the Major Key Technology Project of Shandong

Provincial Key Research and Development Program (2021CXGC010514). Yiming Ma acknowledges the support of the China Scholarship Council.

References

[1] Dobbelaere MR, Plehiers PP, Van de Vijver R, Stevens CV, Van Geem KM. Machine learning in chemical engineering: strengths, weaknesses, opportunities, and threats. Engineering 2021;7(9):1201–11.
[2] Levenspiel O. Modeling in chemical engineering. Chem Eng Sci 2002;57(22–23):4691–6.
[3] Qiao C, Zhao T, Yu X, Qing L, Bao B, Zhao S, et al. On the relation between dynamical density functional theory and Navier–Stokes equation. Chem Eng Sci 2021;230:116203.
[4] Wang CY. Flow due to a stretching boundary with partial slip—an exact solution of the Navier–Stokes equations. Chem Eng Sci 2002;57(17):3745–7.
[5] Gbadago DQ, Moon J, Kim M, Hwang S. A unified framework for the mathematical modelling, predictive analysis, and optimization of reaction systems using computational fluid dynamics, deep neural network and genetic algorithm: a case of butadiene synthesis. Chem Eng J 2021;409:128163.
[6] Schweidtmann AM, Esche E, Fischer A, Kloft M, Repke JU, Sager S, et al. Machine learning in chemical engineering: a perspective. Chemie Ingenieur Technik 2021;93(12):2029–39.
[7] Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 2018;15(3):504–8.
[8] Glaser JI, Benjamin AS, Farhoodi R, Kording KP. The roles of supervised machine learning in systems neuroscience. Prog Neurobiol 2019;175:126–37.
[9] Artrith N, Butler KT, Coudert FX, Han S, Isayev O, Jain A, et al. Best practices in machine learning for chemistry. Nat Chem 2021;13(6):505–8.
[10] Micol Policarpo L, da Silveira DE, da Rosa RR, Antunes Stoffel R, da Costa CA, Victória Barbosa JL, et al. Machine learning through the lens of e-commerce initiatives: an up-to-date systematic literature review. Comput Sci Rev 2021;41:100414.
[11] Ghoddusi H, Creamer GG, Rafizadeh N. Machine learning in energy economics and finance: a review. Soc Econ 2019;81:709–27.
[12] Char DS, Abràmoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. Am J Bioeth 2020;20(11):7–17.
[13] Wäldchen J, Mäder P. Machine learning for image based species identification. Methods Ecol Evol 2018;9(11):2216–25.
[14] Kolachalama VB, Garg PS. Machine learning and medical education. npj Digit Med 2018;1:54.
[15] Wang W, Siau K. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: a review and research agenda. J Database Manag 2019;30(1):61–79.
[16] Michalski RS, Carbonell JG, Mitchell TM, editors. Machine learning: an artificial intelligence approach. Berlin: Springer-Verlag; 2013.
[17] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015;521(7553):452–9.
[18] Hartikainen K, Geng X, Haarnoja T, Levine S. Dynamical distance learning for semi-supervised and unsupervised skill discovery. 2019. arXiv:1907.08225.
[19] Saravanan R, Sujatha P. A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In: Proceedings of 2018 Second International Conference on Intelligent Computing and Control Systems; 2018 Jun 14–15; Madurai, India. Piscataway: IEEE; 2018. p. 945–9.
[20] Kim KG. Book review: deep learning. Healthc Inform Res 2016;22(4):351–4.
[21] Ding J, Xu N, Nguyen MT, Qiao Q, Shi Y, He Y, et al. Machine learning for molecular thermodynamics. Chin J Chem Eng 2021;31:227–39.
[22] Venkatasubramanian V. The promise of artificial intelligence in chemical engineering: is it here, finally? AIChE J 2019;65(2):466–78.
[23] Xiouras C, Cameli F, Quilló GL, Kavousanakis ME, Vlachos DG, Stefanidis GD. Applications of artificial intelligence and machine learning algorithms to crystallization. Chem Rev 2022;122(15):13006–42.
[24] Liu J, Kuang W, Liu J, Gao Z, Rohani S, Gong J. *In-situ* multi-phase flow imaging for particle dynamic tracking and characterization: advances and applications. Chem Eng J 2022;438:135554.
[25] Gong J, Sun J, Wang J. Research progress of industrial crystallization towards intelligent manufacturing. CIESC J 2018;69(11):4505–17. Chinese.
[26] Gao Z, Rohani S, Gong J, Wang J. Recent developments in the crystallization process: toward the pharmaceutical industry. Engineering 2017;3(3):343–53.
[27] Heng T, Yang D, Wang R, Zhang L, Lu Y, Du G. Progress in research on artificial intelligence applied to polymorphism and cocrystal prediction. ACS Omega 2021;6(24):15543–50.
[28] Vermeire FH, Green WH. Transfer learning for solvation free energies: from quantum chemistry to experiments. Chem Eng J 2021;418:129307.
[29] Liu F, Chen K, Xue D. How to fast grow large-size crystals? Innovation 2023;4(4):100458.
[30] Vippagunta SR, Brittain HG, Grant DJW. Crystalline solids. Adv Drug Deliv Rev 2001;48(1):3–26.
[31] Xin D, Gonnella NC, He X, Horspool K. Solvate prediction for pharmaceutical organic molecules with machine learning. Cryst Growth Des 2019;19(3):1903–11.
[32] Bhardwaj RM, Reutzel-Edens SM, Johnston BF, Florence AJ. A random forest model for predicting crystal packing of olanzapine solvates. CrystEngComm 2018;20(28):3947–50.
[33] Johnston A, Johnston BF, Kennedy AR, Florence AJ. Targeted crystallisation of novel carbamazepine solvates based on a retrospective random forest classification. CrystEngComm 2008;10(1):23–5.
[34] Takieddin K, Khimyak YZ, Fábián L. Prediction of hydrate and solvate formation using statistical models. Cryst Growth Des 2016;16(1):70–81.
[35] Ma YM, Niu Y, Yang HY, Dai JY, Lin JW, Wang HQ, et al. Prediction and design of cyclodextrin inclusion complexes formation via machine learning-based strategies. Chem Eng Sci 2022;261:117946.
[36] Egorova O, Hafizi R, Woods DC, Day GM. Multifidelity statistical machine learning for molecular crystal structure prediction. J Phys Chem A 2020;124(39):8065–78.
[37] Chui CP, Liu W, Xu Y, Zhou Y. Molecular dynamics simulation of iron—a review. Spin 2015;5(4):1540007.
[38] Pèpe G, Fery-Forgues S, Jouanna P. Predicting crystal structure and habit of organic micro-crystals by experimentally assisted molecular modelling (EAMM). The case of *n*-octylamino-NBD. J Cryst Growth 2011;333(1):25–35.
[39] Yang P, Jia S, Wang Y, Li Z, Wu S, Wang J, et al. Dissolution behavior, thermodynamic and kinetic analysis of malonamide by experimental measurement and molecular simulation. Chin J Chem Eng 2023;53:260–9.
[40] Li Z, Jia S, Gao Y, Wang M, Hong W, Gao Z, et al. Solid–liquid equilibrium behavior and thermodynamic analysis of *p*-aminobenzoic acid using experimental measurement and molecular dynamic simulation. J Mol Liq 2021;323:114964.
[41] Doi H, Takahashi KZ, Aoyagi T. Mining of effective local order parameters for classifying crystal structures: a machine learning study. J Chem Phys 2020;152(21):214501.
[42] Wengert S, Csányi G, Reuter K, Margraf JT. Data-efficient machine learning for molecular crystal structure prediction. Chem Sci 2021;12(12):4536–46.
[43] Seko A, Togo A, Hayashi H, Tsuda K, Chaput L, Tanaka I. Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and Bayesian optimization. Phys Rev Lett 2015;115(20):205901.
[44] Faber FA, Lindmaa A, von Lilienfeld OA, Armiento R. Machine learning energies of 2 million elpasolite (ABC_2D_6) crystals. Phys Rev Lett 2016;117(13):135502.
[45] Xue D, Balachandran PV, Hogden J, Theiler J, Xue D, Lookman T. Accelerated search for materials with targeted properties by adaptive design. Nat Commun 2016;7:11241.
[46] Chibani S, Coudert FX. Machine learning approaches for the prediction of materials properties. APL Mater 2020;8(8):080701.
[47] Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A. Universal fragment descriptors for predicting properties of inorganic crystals. Nat Commun 2017;8:15679.
[48] Legrain F, Carrete J, van Roekeghem A, Curtarolo S, Mingo N. How chemical composition alone can predict vibrational free energies and entropies of solids. Chem Mater 2017;29(15):6220–7.
[49] Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett 2018;120(14):145301.
[50] Tawfik SA, Isayev O, Spencer MJ, Winkler DA. Predicting thermal properties of crystals using machine learning. Adv Theory Simul 2020;3(2):1900208.
[51] Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature 2018;559(7715):547–55.
[52] Xiouras C, Cameli F, Quillo GL, Kavousanakis ME, Vlachos DG, Stefanidis GD. Applications of artificial intelligence and machine learning algorithms to crystallization. Chem Rev 2022;122(15):13006–42.
[53] Pilania G, Wang C, Jiang X, Rajasekaran S, Ramprasad R. Accelerating materials property predictions using machine learning. Sci Rep 2013;3(1):2810.
[54] Sadat SM, Wang RY. A machine learning based approach for phononic crystal property discovery. J Appl Phys 2020;128(2):025106.
[55] Pandey A, Pokharel R. Machine learning based surrogate modeling approach for mapping crystal deformation in three dimensions. Scr Mater 2021;193:1–5.
[56] Morita K, Davies DW, Butler KT, Walsh A. Modeling the dielectric constants of crystals using machine learning. J Chem Phys 2020;153(2):024503.
[57] Wang B, Wang M, Fan Q, Wang Y, Zhang H, Yue Y. Study on prediction of crystal properties based on deep learning. J Sys Simul 2021;33(12):2854–63. Chinese.
[58] Wang B, Fan Q, Yue Y. Study of crystal properties based on attention mechanism and crystal graph convolutional neural network. J Phys Condens Matter 2022;34(19):195901.
[59] Kirman J, Johnston A, Kuntz DA, Askerka M, Gao Y, Todorović P, et al. Machine-learning-accelerated perovskite crystallization. Matter 2020;2(4):938–47.
[60] Jia S, Gao Z, Tian N, Li Z, Gong J, Wang J, et al. Review of melt crystallization in the pharmaceutical field, towards crystal engineering and continuous process development. Chem Eng Res Des 2021;166:268–80.
[61] Jia S, Yang P, Gao Z, Li Z, Fang C, Gong J. Recent progress in antisolvent crystallization. CrystEngComm 2022;24(17):3122–35.
[62] Wicker JGP, Crowley LM, Robshaw O, Little EJ, Stokes SP, Cooper RI, et al. Will they co-crystallize? CrystEngComm 2017;19(36):5336–40.

[63] Mswahili ME, Lee MJ, Martin GL, Kim J, Kim P, Choi GJ, et al. Cocrystal prediction using machine learning models and descriptors. Appl Sci 2021;11(3):1323.
[64] Wang D, Yang Z, Zhu B, Mei X, Luo X. Machine-learning-guided cocrystal prediction based on large data base. Cryst Growth Des 2020;20(10):6610–21.
[65] Chabalenge B, Korde S, Kelly AL, Neagu D, Paradkar A. Understanding matrix-assisted continuous co-crystallization using a data mining approach in quality by design (QbD). Cryst Growth Des 2020;20(7):4540–9.
[66] Ferguson AL. Machine learning and data science in soft materials engineering. J Phys Condens Matter 2018;30(4):043002.
[67] Yamaguchi S. Molecular field analysis for data-driven molecular design in asymmetric catalysis. Org Biomol Chem 2022;20(31):6057–71.
[68] Margraf JT. Science-driven atomistic machine learning. Angew Chem Int Ed 2023;62(26):e202219170.
[69] Ashraf C, Joshi N, Beck DAC, Pfaendtner J. Data science in chemical engineering: applications to molecular science. Annu Rev Chem Biomol Eng 2021;12(1):15–37.
[70] Li W, Ma H, Li S, Ma J. Computational and data driven molecular material design assisted by low scaling quantum mechanics calculations and machine learning. Chem Sci 2021;12(45):14987–5006.
[71] Talapin DV, Lee JS, Kovalenko MV, Shevchenko EV. Prospects of colloidal nanocrystals for electronic and optoelectronic applications. Chem Rev 2010;110(1):389–458.
[72] Li P, Li Y, Zhou ZK, Tang S, Yu XF, Xiao S, et al. Evaporative self-assembly of gold nanorods into macroscopic 3D plasmonic superlattice arrays. Adv Mater 2016;28(13):2511–7.
[73] Coe S, Woo WK, Bawendi M, Bulović V. Electroluminescence from single monolayers of nanocrystals in molecular organic devices. Nature 2002;420(6917):800–3.
[74] Talapin DV, Murray CB. PbSe nanocrystal solids for n- and p-channel thin film field-effect transistors. Science 2005;310(5745):86–9.
[75] Wu S, Han G, Milliron DJ, Aloni S, Altoe V, Talapin DV, et al. Non-blinking and photostable upconverted luminescence from single lanthanide-doped nanocrystals. Proc Natl Acad Sci USA 2009;106(27):10917–21.
[76] Chen T, Qiu M, Peng Y, Yi C, Xu Z. Colloidal polymer-templated formation of inorganic nanocrystals and their emerging applications. Small 2023;19(44):2303282.
[77] Granda JM, Donina L, Dragone V, Long DL, Cronin L. Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature 2018;559(7714):377–81.
[78] Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, et al. Organic synthesis in a modular robotic system driven by a chemical programming language. Science 2019;363(6423):eaav2211.
[79] Zhao H, Chen W, Huang H, Sun Z, Chen Z, Wu L, et al. A robotic platform for the synthesis of colloidal nanocrystals. Nat Synth 2023;2(6):505–14.
[80] Chan EM, Xu C, Mao AW, Han G, Owen JS, Cohen BE, et al. Reproducible, high-throughput synthesis of colloidal nanocrystals for optimization in multidimensional parameter space. Nano Lett 2010;10(5):1874–85.
[81] Wicker JGG, Cooper RI. Will it crystallise? Predicting crystallinity of molecular materials. CrystEngComm 2015;17(9):1927–34.
[82] Pillong M, Marx C, Piechon P, Wicker JGP, Cooper RI, Wagner T. A publicly available crystallisation data set and its application in machine learning. CrystEngComm 2017;19(27):3737–45.
[83] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, et al. Machine-learning-assisted materials discovery using failed experiments. Nature 2016;533(7601):73–6.
[84] Venkatram S, Batra R, Chen L, Kim C, Shelton M, Ramprasad R. Predicting crystallization tendency of polymers using multifidelity information fusion and machine learning. J Phys Chem B 2020;124(28):6046–54.
[85] Ghosh A, Louis L, Arora KK, Hancock BC, Krzyzaniak JF, Meenan P, et al. Assessment of machine learning approaches for predicting the crystallization propensity of active pharmaceutical ingredients. CrystEngComm 2019;21(8):1215–23.
[86] Pereira F. Machine learning methods to predict the crystallization propensity of small organic molecules. CrystEngComm 2020;22(16):2817–26.
[87] Boobier S, Hose DRJ, Blacker AJ, Nguyen BN. Machine learning with physicochemical relationships: solubility prediction in organic solvents and water. Nat Commun 2020;11:5753.
[88] Cui Q, Lu S, Ni B, Zeng X, Tan Y, Chen YD, et al. Improved prediction of aqueous solubility of novel compounds by going deeper with deep learning. Front Oncol 2020;10:121.
[89] Jia S, Zhang K, Wan X, Gao Z, Gong J, Rohani S. Effects of temperature and solvent properties on the liquid–solid phase equilibrium of γ-pyrazinamide. J Chem Eng Data 2020;65(7):3667–78.
[90] Zhao J, Yang J, Xie Y. Improvement strategies for the oral bioavailability of poorly water-soluble flavonoids: an overview. Int J Pharm 2019;570:118642.
[91] Fernandes GJ, Kumar L, Sharma K, Tunge R, Rathnanand M. A review on solubility enhancement of carvedilol—a BCS class II drug. J Pharm Innov 2018;13(3):197–212.
[92] Loschen C, Klamt A. Solubility prediction, solvate and cocrystal screening as tools for rational crystal engineering. J Pharm Pharmacol 2015;67(6):803–11.
[93] Sheikholeslamzadeh E, Rohani S. Solubility prediction of pharmaceutical and chemical compounds in pure and mixed solvents using predictive models. Ind Eng Chem Res 2012;51(1):464–73.
[94] Ma Y, Gao Z, Shi P, Chen M, Wu S, Yang C, et al. Machine learning-based solubility prediction and methodology evaluation of active pharmaceutical ingredients in industrial crystallization. Front Chem Sci Eng 2022;16(4):523–35.
[95] Ge K, Ji Y. Novel computational approach by combining machine learning with molecular thermodynamics for predicting drug solubility in solvents. Ind Eng Chem Res 2021;60(25):9259–68.
[96] Zou F, Zhuang W, Wu J, Zhou J, Yang P, Liu Q, et al. Determination of metastable zone widths and the primary nucleation and growth mechanisms for the crystallization of disodium guanosine 5′-monophosphate from a water–ethanol system. Ind Eng Chem Res 2015;54(1):137–45.
[97] Kadam SS, Kramer HJM, ter Horst JH. Combination of a single primary nucleation event and secondary nucleation in crystallization processes. Cryst Growth Des 2011;11(4):1271–7.
[98] Myerson AS, Trout BL. Nucleation from solution. Science 2013;341(6148):855–6.
[99] Xu S, Wang J, Zhang K, Wu S, Liu S, Li K, et al. Nucleation behavior of eszopiclone-butyl acetate solutions from metastable zone widths. Chem Eng Sci 2016;155:248–57.
[100] Ma S, Li C, Gao J, Yang H, Tang W, Gong J, et al. Artificial neural network prediction of metastable zone widths in reactive crystallization of lithium carbonate. Ind Eng Chem Res 2020;59(16):7765–76.
[101] Chen X, Wang LG, Meng F, Luo ZH. Physics-informed deep learning for modelling particle aggregation and breakage processes. Chem Eng J 2021;426:131220.
[102] Lindenberg C, Schöll J, Vicum L, Mazzotti M, Brozio J. *L*-Glutamic acid precipitation: agglomeration effects. Cryst Growth Des 2008;8(1):224–37.
[103] Ålander EM, Uusi-Penttila MS, Rasmuson ÅC. Characterization of paracetamol agglomerates by image analysis and strength measurement. Powder Technol 2003;130(1–3):298–306.
[104] Faria N, Pons MN, de Azevedo SF, Rocha FA, Vivier H. Quantification of the morphology of sucrose crystals by image analysis. Powder Technol 2003;133(1–3):54–67.
[105] Ochsenbein DR, Vetter T, Schorsch S, Morari M, Mazzotti M. Agglomeration of needle-like crystals in suspension: I. measurements. Cryst Growth Des 2015;15(4):1923–33.
[106] Lins J, Harweg T, Weichert F, Wohlgemuth K. Potential of deep learning methods for deep level particle characterization in crystallization. Appl Sci 2022;12(5):2465.
[107] Sinha K, Murphy E, Kumar P, Springer KA, Ho R, Nere NK. A novel computational approach coupled with machine learning to predict the extent of agglomeration in particulate processes. AAPS PharmSciTech 2022;23:18.
[108] Ma Y, Sun M, Liu Y, Chen M, Wu S, Wang M, et al. Design of spherical crystallization of active pharmaceutical ingredients via a highly efficient strategy: from screening to preparation. ACS Sustain Chem Eng 2021;9(27):9018–32.
[109] Vriza A, Canaj AB, Vismara R, Kershaw Cook LJ, Manning TD, Gaultois MW, et al. One class classification as a practical approach for accelerating π–π co-crystal discovery. Chem Sci 2021;12(5):1702–19.
[110] Devogelaer JJ, Meekes H, Tinnemans P, Vlieg E, de Gelder R. Co-crystal prediction by artificial neural networks. Angew Chem Int Ed 2020;59(48):21711–8.
[111] Wang W, Yang T, Harris WH, Gómez-Bombarelli R. Active learning and neural network potentials accelerate molecular screening of ether-based solvate ionic liquids. Chem Commun 2020;56(63):8920–3.
[112] Liu N, Wang J, Sun S, Li C, Tian W. Optimized principal component analysis and multi-state Bayesian network integrated method for chemical process monitoring and variable state prediction. Chem Eng J 2022;430(Pt 1):132617.
[113] Doan XT, Zhou Y, Srinivasan R. Integrating multi-variate image analysis and artificial intelligence techniques with PVM for inline crystal size and shape measurements. In: 2006 AIChE Annual Meeting; 2006 Nov 12–17; San Francisco, CA, USA. Madison: Omnipress; 2006. p. 301aa.
[114] Gao Z, Zhu D, Wu Y, Rohani S, Gong J, Wang J. Motion-based multiple object tracking of ultrasonic-induced nucleation: a case study of *L*-glutamic acid. Cryst Growth Des 2017;17(10):5007–11.
[115] Gao Z, Wu Y, Bao Y, Gong J, Wang J, Rohani S. Image analysis for in-line measurement of multidimensional size, shape, and polymorphic transformation of *L*-glutamic acid using deep learning-based image segmentation and classification. Cryst Growth Des 2018;18(8):4275–81.
[116] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324.
[117] Crestani CE, Bernardo A, Costa CBB, Giulietti M. An artificial neural network model applied to convert sucrose chord length distributions into particle size distributions. Powder Technol 2021;384:186–94.
[118] Chen S, Liu T, Xu D, Huo Y, Yang Y. Image based measurement of population growth rate for *L*-glutamic acid crystallization. In: Proceedings of 2019 Chinese Control Conference; 2019 Jul 27–30; Guangzhou, China. Piscataway: IEEE; 2019. p. 7933–8.
[119] Xiang H, Chen Q, Wu Y, Xu D, Qi S, Mei J, et al. Urine calcium oxalate crystallization recognition method based on deep learning. In: Proceedings of 2019 International Conference on Automation, Computational and Technology Management; 2019 Apr 24–26; London, UK. Piscataway: IEEE; 2019. p. 30–3.
[120] Zhang J, Meng Y, Wu J, Qin J, Wang H, Yao T, et al. Monitoring sugar crystallization with deep neural networks. J Food Eng 2020;280:109965.
[121] Bruno AE, Charbonneau P, Newman J, Snell EH, So DR, Vanhoucke V, et al. Classification of crystallization outcomes using deep convolutional neural networks. PLoS One 2018;13(6):e0198883.

[122] Manee V, Baratti R, Romagnoli JA. Learning to navigate a crystallization model with deep reinforcement learning. Chem Eng Res Des 2022;178:111–23.
[123] Jaeggi A, Rajagopalan AK, Morari M, Mazzotti M. Characterizing ensembles of platelike particles via machine learning. Ind Eng Chem Res 2021;60(1):473–83.
[124] Hung J, Collins J, Weldetsion M, Newland O, Chiang E, Guerrero S, et al. Protein crystallization image classification with elastic net. In: Ourselin S, Styner MA, editors. Proceedings volume 9034, Medical Imaging 2014: Image Processing; 2014 Feb 15–20; San Diego, CA, USA. Bellingham: Society of Photo-Optical Instrumentation Engineers; 2014. p. 90341X.
[125] Sigdel M, Pusey ML, Aygun RS. Real-time protein crystallization image acquisition and classification system. Cryst Growth Des 2013;13(7):2728–36.
[126] Tian C, Cai Y, Yang H, Su M. Investigation on mixed particle classification based on imaging processing with convolutional neural network. Powder Technol 2021;391:267–74.
[127] Chen P, Tang Z, Zeng Z, Hu X, Xiao L, Liu Y, et al. Machine-learning-guided morphology engineering of nanoscale metal–organic frameworks. Matter 2020;2(6):1651–66.
[128] Wu Y, Gao Z, Rohani S. Deep learning-based oriented object detection for *in situ* image monitoring and analysis: a process analytical technology (PAT) application for taurine crystallization. Chem Eng Res Des 2021;170:444–55.
[129] Cote A, Erdemir D, Girard KP, Green DA, Lovette MA, Sirota E, et al. Perspectives on the current state, challenges, and opportunities in pharmaceutical crystallization process development. Cryst Growth Des 2020;20(12):7568–81.
[130] Févotte G. *In situ* Raman spectroscopy for in-line control of pharmaceutical crystallization and solids elaboration processes: a review. Chem Eng Res Des 2007;85(7):906–20.
[131] Nagy ZK, Févotte G, Kramer H, Simon LL. Recent advances in the monitoring, modelling and control of crystallization systems. Chem Eng Res Des 2013;91(10):1903–22.
[132] Fytopoulos AA, Kavousanakis ME, Van Gerven T, Boudouvis AG, Stefanidis GD, Xiouras C. Crystal growth, dissolution, and agglomeration kinetics of sodium chlorate. Ind Eng Chem Res 2021;60(19):7367–84.
[133] Yang Y, Zhang C, Pal K, Koswara A, Quon J, McKeown R, et al. Application of ultra-performance liquid chromatography as an online process analytical technology tool in pharmaceutical crystallization. Cryst Growth Des 2016;16(12):7074–82.
[134] Su Q, Ganesh S, Moreno M, Bommireddy Y, Gonzalez M, Reklaitis GV, et al. A perspective on quality-by-control (QbC) in pharmaceutical continuous manufacturing. Comput Chem Eng 2019;125:216–31.
[135] Zhang T, Szilágyi B, Gong J, Nagy ZK. Novel semibatch supersaturation control approach for the cooling crystallization of heat-sensitive materials. AIChE J 2020;66(6):e16955.
[136] Zhang T, Nagy B, Szilágyi B, Gong J, Nagy ZK. Simulation and experimental investigation of a novel supersaturation feedback control strategy for cooling crystallization in semi-batch implementation. Chem Eng Sci 2020;225:115807.
[137] Gao Y, Zhang T, Ma Y, Xue F, Gao Z, Hou B, et al. Application of PAT-based feedback control approaches in pharmaceutical crystallization. Crystals 2021;11(3):221.
[138] Tacsi K, Gyürkés M, Csontos I, Farkas A, Borbás E, Nagy ZK, et al. Polymorphic concentration control for crystallization using Raman and attenuated total reflectance ultraviolet visible spectroscopy. Cryst Growth Des 2020;20(1):73–86.
[139] Dutta D, Upreti SR. Artificial intelligence-based process control in chemical, biochemical, and biomedical engineering. Can J Chem Eng 2021;99(11):2467–504.
[140] Tan CF, Wahidin LS, Khalil SN, Tamaldin N, Hu J, Rauterberg GWM. The application of expert system: a review of research and applications. ARPN J Eng Appl Sci 2016;11(4):2448–53.
[141] Tan H. A brief history and technical review of the expert system research. In: IOP Conference series: materials science and engineering, volume 242, 2017 3rd International Conference on Applied Materials and Manufacturing Technology; 2017 Jun 23–25; Changsha, China. Bristol: IOP Publishing; 2017. p. 012111.
[142] De Silva CW. Intelligent control: fuzzy logic applications. Boca Raton: CRC Press; 2018.
[143] Konaté AA, Pan H, Khan N, Ziggah YY. Prediction of porosity in crystalline rocks using artificial neural networks: an example from the Chinese continental scientific drilling main hole. Stud Geophys Geod 2015;59:113–36.
[144] Conradie AE, Aldrich C. Neurocontrol of a multi-effect batch distillation pilot plant based on evolutionary reinforcement learning. Chem Eng Sci 2010;65(5):1627–43.
[145] Zang H, Zhang S, Hapeshi K. A review of nature-inspired algorithms. J Bionics Eng 2010;7(S4):S232–7.
[146] Hubbs CD, Li C, Sahinidis NV, Grossmann IE, Wassick JM. A deep reinforcement learning approach for chemical production scheduling. Comput Chem Eng 2020;141:106982.
[147] Ali JM, Hussain MA, Tade MO, Zhang J. Artificial intelligence techniques applied as estimator in chemical process systems—a literature survey. Expert Syst Appl 2015;42(14):5915–31.
[148] Wang J, Cao LL, Wu HY, Li XG, Jin QB. Dynamic modeling and optimal control of batch reactors, based on structure approaching hybrid neural networks. Ind Eng Chem Res 2011;50(10):6174–86.
[149] Boroushaki M, Ghofrani MB, Lucas C, Yazdanpanah MJ. Identification and control of a nuclear reactor core (VVER) using recurrent neural networks and fuzzy systems. IEEE Trans Nucl Sci 2003;50(1):159–74.
[150] Meng Y, Yao T, Yu S, Qin J, Zhang J, Wu J. Data-driven modeling for crystal size distribution parameters in cane sugar crystallization process. J Food Process Eng 2021;44(4):e13648.
[151] Meng Y, Lan Q, Qin J, Yu S, Pang H, Zheng K. Data-driven soft sensor modeling based on twin support vector regression for cane sugar crystallization. J Food Eng 2019;241:159–65.
[152] Alhazmi K, Albalawi F, Sarathy SM. A reinforcement learning-based economic model predictive control framework for autonomous operation of chemical reactors. Chem Eng J 2022;428:130993.
[153] Kamaraju VK, Chiu MS. Improved operation of concentration control for antisolvent crystallization processes. Org Process Res Dev 2015;19(1):178–88.
[154] Hojjati H, Sheikhzadeh M, Rohani S. Control of supersaturation in a semibatch antisolvent crystallization process using a fuzzy logic controller. Ind Eng Chem Res 2007;46(4):1232–40.
[155] Sheikhzadeh M, Trifkovic M, Rohani S. Fuzzy logic and rigid control of a seeded semi-batch, anti-solvent, isothermal crystallizer. Chem Eng Sci 2008;63(4):991–1002.
[156] Sheikhzadeh M, Trifkovic M, Rohani S. Adaptive MIMO neuro-fuzzy logic control of a seeded and an unseeded anti-solvent semi-batch crystallizer. Chem Eng Sci 2008;63(5):1261–72.
[157] Paengjuntuek W, Thanasinthana L, Arpornwichanop A. Neural network-based optimal control of a batch crystallizer. Neurocomputing 2012;83:158–64.
[158] Damour C, Benne M, Grondin-Perez B, Chabriat JP. Nonlinear predictive control based on artificial neural network model for industrial crystallization. J Food Eng 2010;99(2):225–31.
[159] Georgieva P, de Azevedo SF. Application of artificial neural networks in modeling and optimization of batch crystallization processes. J Electrón Telecomun 2006;4(6):697–706.
[160] Georgieva P, de Azevedo SF. Application of feed forward neural networks in modeling and control of a fed-batch crystallization process. Trans Eng Comput Technol 2006;12:65–70.
[161] Guo P, Rao S, Hao L, Wang J. Fault diagnosis of a semi-batch crystallization process through deep learning method. Comput Chem Eng 2022;164:107807.
[162] Manee V, Zhu W, Romagnoli JA. A deep learning image-based sensor for real-time crystal size distribution characterization. Ind Eng Chem Res 2019;58(51):23175–86.
[163] Manee V, Baratti R, Romagnoli JA. Optimal strategies to control particle size and variance in antisolvent crystallization operations using deep RL. Chem Eng Trans 2021;86:943–8.
[164] Öner M, Montes FCC, Ståhlberg T, Stocks SM, Bajtner JE, Sin G. Comprehensive evaluation of a data driven control strategy: experimental application to a pharmaceutical crystallization process. Chem Eng Res Des 2020;163:248–61.
[165] Wang L, Zhu Y, Gan C. Predictive control of particle-size distribution of crystallization process using deep learning based image analysis. AIChE J 2022;68(11):e17817.
[166] Gajera U, Storchi L, Amoroso D, Delodovici F, Picozzi S. Toward machine learning for microscopic mechanisms: a formula search for crystal structure stability based on atomic properties. J Appl Phys 2022;131(21):215703.
[167] Das T, Hossen MN, Rahman SKM, Parvin T, Ahmed K, Bui FM. In: Proceedings of 2022 2nd International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies; 2022 Apr 21–22; Bhilai, India. Piscataway: IEEE; 2022.
[168] Hewitt M, Cronin MTD, Enoch SJ, Madden JC, Roberts DW, Dearden JC. *In silico* prediction of aqueous solubility: the solubility challenge. J Chem Inf Model 2009;49(11):2572–87.
[169] Jouyban A, Acree Jr WE. Mathematical derivation of the Jouyban–Acree model to represent solute solubility data in mixed solvents at various temperatures. J Mol Liq 2018;256:541–7.
[170] Dropka N, Holena M. Application of artificial neural networks in crystal growth of electronic and opto-electronic materials. Crystals 2020;10(8):663.
[171] Chayatummagoon S, Chongstitvatana P. Image classification of sugar crystal with deep learning. In: Proceedings of 2021 13th International Conference on Knowledge and Smart Technology; 2021 Jan 21–24; Chonburi, Thailand. Piscataway: IEEE; 2021. p. 118–22.
[172] Rymarczyk T, Klosowski G, Cieplak T, Kozlowski E. Industrial processes control with the use of a neural tomographic algorithm. Prz Elektrotechniczn 2019;95(2):96–9.
[173] Dropka N, Böttcher K, Holena M. Development and optimization of VGF-GaAs crystal growth process using data mining and machine learning techniques. Crystals 2021;11(10):1218.
[174] Yao TS, Tang CY, Yang M, Zhu KJ, Yan DY, Yi CJ, et al. Machine learning to instruct single crystal growth by flux method. Chin Phys Lett 2019;36(6):068101.
[175] Le NQK, Li W, Cao Y. Sequence-based prediction model of protein crystallization propensity using machine learning and two-level feature selection. Brief Bioinform 2023;24(5):bbad319.
[176] Pierro C, Capitelli F. Inorganic phosphates investigation by support vector machine. In: Laganá A, Gavrilova ML, Kumar V, Mun Y, Tan CJK, Gervasi O, editors. Computational science and its applications—ICCSA 2004. Berlin: Springer; 2004. p. 338–49.

[177] Shekar V, Nicholas G, Najeeb MA, Zeile M, Yu V, Wang X, et al. Active meta-learning for predicting and selecting perovskite crystallization experiments. J Chem Phys 2022;156(6):064108.
[178] Kalyoncu C, Yasli A, Ademgil H. Machine learning methods for estimating bent photonic crystal fiber based SPR sensor properties. Heliyon 2022;8(11):e11582.
[179] Nakata H, Bai S. Development of a new parameter optimization scheme for a reactive force field based on a machine learning approach. J Comput Chem 2019;40(23):2000–12.
[180] Corrias M, Papa L, Sokolović I, Birschitzky V, Gorfer A, Setvin M, et al. Automated real-space lattice extraction for atomic force microscopy images. Mach Learn Sci Technol 2023;4(1):015015.
[181] Chan H, Cherukara M, Loeffler TD, Narayanan B, Sankaranarayanan SKRS. Machine learning enabled autonomous microstructural characterization in 3D samples. npj Comput Mater 2020;6(1):1.
[182] Schön CF, van Bergerem S, Mattes C, Yadav A, Grohe M, Kobbelt L, et al. Classification of properties and their relation to chemical bonding: essential steps toward the inverse design of functional materials. Sci Adv 2022;8(47):eade0828.
[183] Zhang J, Nguyan J, Xiong Z, Morris J. Iterative learning control of a crystallisation process using batch wise updated linearised models. In: Proceedings of 2009 Chinese Control and Decision Conference; 2009 Jun 17–19; Guilin, China. Piscataway: IEEE; 2009. p. 1734.
[184] Togkalidou T, Braatz RD, Johnson BK, Davidson O, Andrews A. Experimental design and inferential modeling in pharmaceutical crystallization. AIChE J 2001;47(1):160–8.
[185] Briones J, Guinto MC, Pelicano CM. Accelerated lattice constant prediction of perovskite materials (ABX_3, $A_2BB'O_6$) using partial least squares and principal component regression methods. Mater Lett 2021;298:130040.
[186] Simone E, Zhang W, Nagy ZK. Analysis of the crystallization process of a biopharmaceutical compound in the presence of impurities using process analytical technology (PAT) tools. J Chem Technol Biol 2016;91(5):1461–70.
[187] Taris A, Hansen TB, Rong BG, Grosso M, Qu H. Detection of nucleation during cooling crystallization through moving window PCA applied to *in situ* infrared data. Org Process Res Dev 2017;21(7):966–75.
[188] Diorazio LJ, Hose DRJ, Adlington NK. Toward a more holistic framework for solvent selection. Org Process Res Dev 2016;20(4):760–73.