Machine learning-assisted colloidal synthesis: A review

D G Gulevich, Igor Nabiev, P S Samokhvalov

Machine learning–assisted colloidal synthesis: A review

D.G. Gulevich.[a,*], I.R. Nabiev,[a,b,c,d,*] P.S. Samokhvalov[a,b]

[a] *Laboratory of Nano-Bioengineering, National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409 Moscow, Russian Federation*

[b] *Life Improvement by Future Technologies (LIFT) Center, 143025 Moscow, Russian Federation*

[c] *Laboratoire de Recherche en Nanosciences, LRN-EA4682, 51 rue Cognacq Jay, Université de Reims Champagne-Ardenne, 51100 Reims, France*

[d] *Department of Clinical Immunology and Allergology, Institute of Molecular Medicine, Sechenov First Moscow State Medical University (Sechenov University), 119146 Moscow, Russian Federation*

* To whom correspondence should be addressed. E-mails: gulevich8d7@gmail.com or igor.nabiev@univ-reims.fr

Abstract

Artificial intelligence (AI) technologies, including machine learning and deep learning, have become ingrained in both everyday life and in scientific research. In chemistry, these algorithms are most commonly used for the development of new materials and drugs, recognition of microscopy images, and analysis of spectral data. Finding relationships between the parameters of chemical synthesis and the properties of the resultant materials is often challenging because of the large number of variations of the temperature and time of synthesis, the chemical composition and ratio of precursors, etc. Applying machine and deep learning to the organization of chemical experiments will considerably reduce the empiricism issues in chemical research. Colloidal nanomaterials, whose morphology, size, and phase composition are influenced directly not only by the synthesis conditions, but the reagents or solvents purity and other indistinct factors are highly demanded in optoelectronics, catalysis, biological imaging, and sensing applications. In recent years, AI methods have been increasingly used for determining the key factors of synthesis and selecting the optimal reaction conditions for obtaining nanomaterials with precisely controlled and reproducible characteristics. The purpose of this review is to analyze the current progress in the AI-assisted optimization of the most common methods of production of colloidal nanomaterials, including colloidal and hydrothermal syntheses, chemical reduction, and synthesis in flow reactors.

Keywords: colloidal nanomaterials, machine learning, hot-injection synthesis, hydrothermal synthesis, chemical reduction.

1. Introduction

Artificial intelligence (AI) is a branch of computational science. Intense research in this field started back in the 1940's to 1950's. The term *artificial intelligence* was coined at a conference at Dortmund College in 1956 [1]. Currently, the term refers to computer systems with abilities primarily associated with those of the human intelligence. These include the capacity for learning, reasoning, problem solving, speech and image recognition, and language translation [2]. The first empirical test determining whether or not a computer program exhibits AI (or, more generally, answering the question, "Can machines think?") was devised by Alan Turing in 1950 [3]. It is an

imitation game consisting in conversation between a human evaluator and a machine. If, when the conversation is over, the evaluator cannot determine whether the partner in conversation has been a machine or a human being, the computer system is considered to have passed the test.

Regarding the data processing by means of AI, two approaches have emerged: the "bottom-up" approach using the artificial neural network model based on reproducing the biological basis of natural intelligence and the "top-down" one simulating intellectual activity through a sequence of reasoning, which is termed *symbolic AI.* The latter approach was prevailing in the first wave of developments in this field. Certainly, the rate of progress in the field of AI has been nonlinear, influenced by many factors, mainly the computation capacity and memory limitations of the computers, the designing and accumulation of relevant databases, and the development of computational algorithms. For example, the peak performance of one of the first Cray-1 supercomputers of Cray Research Inc (1975) was 250 million floating point operations per second (250 megaFLOPS) [4], whereas speech recognition tasks require a performance in the gigaFLOPS range. By comparison, the peak performance of ATERUI II, which has been used to build the most detailed model of the Universe, is 3.087×10^{15} FLOPS [5]. In May 2022, the Frontier supercomputer of the Oak Ridge National Laboratory (ORNL), USA, was rated first among the most high-performance computers, having crossed the 1 exaFLOPS barrier. Frontier, which had 8,730,112 processor cores, reached a performance of 1.102 exaFLOPS [6]. Early high expectations of AI, largely unmet because of the aforementioned limiting factors, led to a decrease in interest in, and funding of this field in the 1970s. This period in history is known as the "winter of AI". It was relatively brief, a new upsurge in the field starting as early as the 1980s. Since that time, the term *machine learning* (ML) has been used increasingly often. ML is a subdivision of AI (Fig. 1 a). This term refers to a set of mathematical methods and algorithms designed to solve problems by extracting patterns from input data. ML is a subset of AI, mathematical, statistical, and calculational methods for developing algorithms capable of solving a problem indirectly by finding consistent patterns in input data. The learning consists of the following steps: (1) transferring data to the algorithm; (2) training the model; (3) testing and deploying the model; and (4) using an expanded model for automated problem solving based on prediction. Deep learning (DL) is a subset of ML based on artificial neural networks. One of the differences of DL from other ML approaches is the use of more complex neural networks containing hidden layers, which makes it impossible to determine which criterion has been the key one in making a "decision". In addition, unlike ML, DL requires larger data sets for analysis and, accordingly, a higher computational capacity.

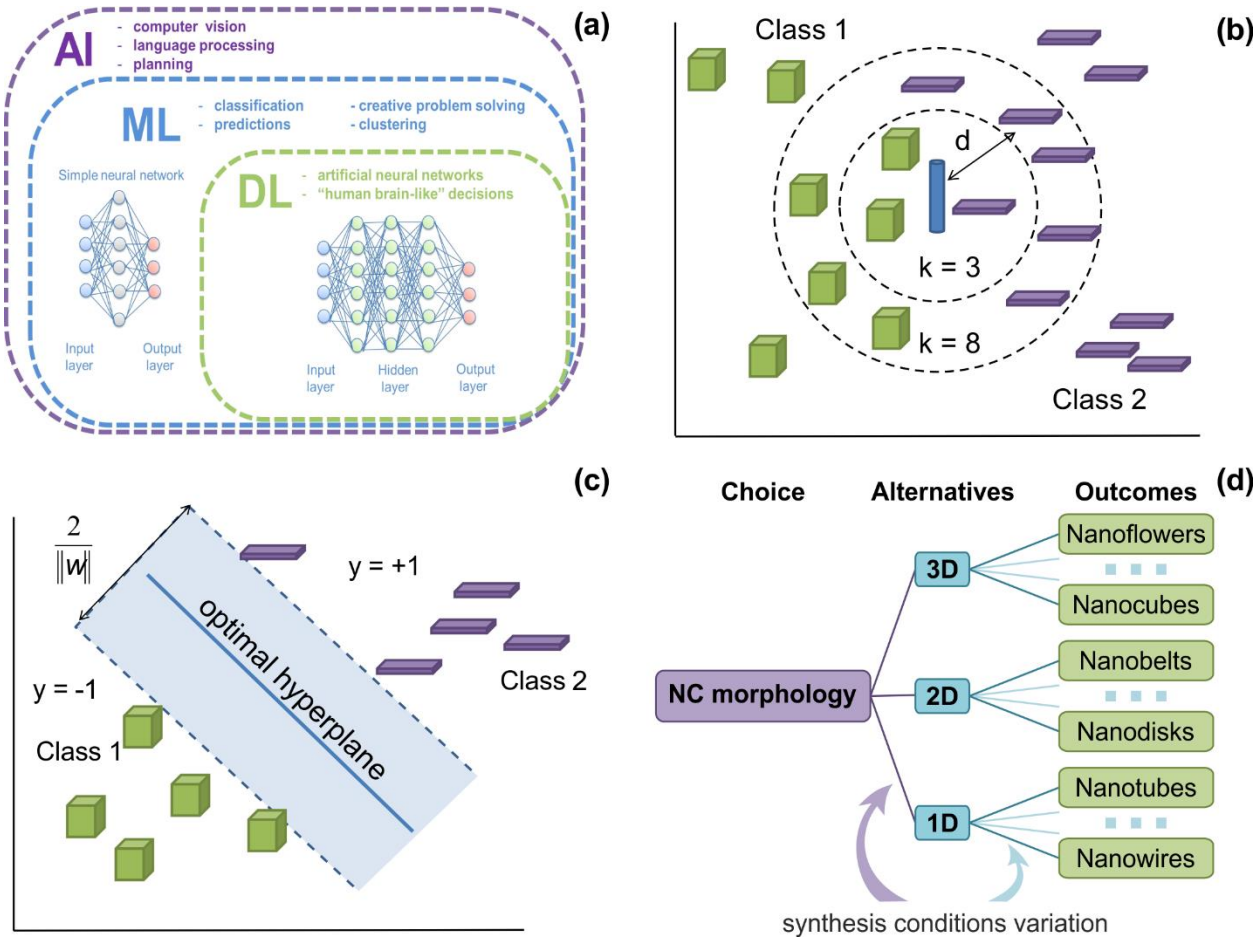

Fig. 1. (a) Components of the artificial intelligence (AI) concept and schematic representation of basic machine learning concepts: (b) k-Nearest Neighbors (kNN); (c) Support Vector Machine (SVM) (objects are divided into classes based on the values f(x) = 1 or -1. The classifying function is f(x) = sgn(<w·xi>+ b), where w is the normal vector to the separating hyperplane, *b* is the auxiliary parameter [7]; and (d) Decision Tree (DT)

Machine learning is subdivided into classical learning (supervised or unsupervised), reinforcement learning, ensemble methods, and neural network methods. Supervised learning deals with regression and classification tasks. Here, the "supervisors" are preset input data and the corresponding output data serving as training sets, which are used to train the algorithm to identify dependences and subsequently to make predictions for new input data. This group of algorithms includes the k-nearest neighbors (kNN) method, support vector machine (SVM), decision trees (DT), naive Bayes classifiers, etc. [8]. To improve the efficiency of these computational systems, they can be combined. This approach, referred to as the ensemble method, typically employs ensembles of DTs. These methods will be described in more detail in the next section. There are three ways to build ensembles: bagging (where the output data obtained by multiple training of a single algorithm are averaged), stacking (the output data are generated by a solver algorithm using the results of training several different algorithms), and boosting (sequential training of algorithms; each new data set is generated from the data that have been processed incorrectly by the preceding algorithm). Unsupervised learning is much less common. In this case, the algorithm independently searches for relationships between the objects of study. This approach is used to solve the tasks of generalization, clustering, and association rule learning. *Kusaba et al*. [9] reported an interesting example of using this ML approach. They tested whether AI could reconstruct the periodic table of elements by analyzing the pattern of changes in the physical and chemical properties of elements. For this purpose, they used an unsupervised periodic table generator algorithm based on

generative topographic mapping, which automatically translated multidimensional data into a table whose layout could be varied on demand. The algorithm was given 39 features (electronegativity, melting point, etc.) of the first 54 elements to work with. The output was two versions of the table, two-dimensional and conical three-dimensional. In both cases, the result was highly similar to the familiar form of the table: alkali, alkaline earth metals, halogens, and noble gases were distinctly identified and placed into separate groups. A slight disorder occurred in groups VI and VII of the conic model, as well as in the position of He. Nevertheless, the test of the ability of the machine algorithm to repeat the result of human intellectual work has shown how high the level of progress in this field is, and what great prospects it offers. Another type of ML that has been actively developed in recent years is termed reinforcement learning. This method allows an algorithm to learn by trial and error in an interactive environment, using feedback from its own actions and experience. In 2006 a new ML concept was developed, which is termed *deep learning* (DL). DL is based on neural networks with a more complex architecture containing so-called hidden layers. Although the mathematical model of neural networks was known as early as the second half of the 20th century, technical limitations precluded a wide spread of DL until 2012, when a neural network algorithm proved to be the most successful in image recognition for the first time [10]. The use of DL requires larger databases and, correspondingly, a higher computational capacity of the computer, which makes it considerably more expensive than ML.

Let us consider in more detail the milestones of AI application to chemistry. In general, ML requires three components: raw data; a list of attributes, i.e., the parameters used for classification or prediction; and the data-processing algorithm itself. Each field of science is characterized by a large body of accumulated experimental data growing exponentially. Regarding chemistry, about one million natural and synthesized compounds were known back in 1960 [11]. As of January 2023, as many as 115,668,216 compounds and 307,434,809 substances were registered, according to PubChem [12]. More specialized databases on inorganic and organometallic compounds, as well as reactive and spectroscopy data, have been formed since the 1980s [11].

One of the first examples of successful use of ML methods in chemistry is the studies by *Kowalski et al.* on the recognition of mass spectra of organic molecules [13] and the prediction of molecular and structural formulas [14] carried out in the late 1960s. For example, with a training set of 400 oxygen-containing organic molecules, the efficiencies of recognition and prediction of mass spectra were 94% and 79%, respectively [13]. The efficiencies of prediction of molecular and even structural formulas using the databases of mass spectra of 387 hydrocarbons and 243 oxygen- or nitrogen-containing organic substances were, on average, 90 and 88%, respectively [14]. Fig. 2 shows how the publication activity in the field of chemical applications of ML has changed with time. It clearly reflects the drastic increase in the number of publications since 2015, when a breakthrough in the development of neural networks occurred. Analysis of the subjects and research areas of the array of research papers shows that the largest number of publications deal with the design of new materials and drugs and the analysis and recognition of spectra (41.4, 26.7, and 15.2%, respectively).

The use of AI for the analysis and optimization of chemical synthesis began in the field of retrosynthesis of organic compounds, where the task was to propose a synthesis pathway for complex molecules based on their structural formulas. The computer-assisted synthesis design was pioneered by the Harvard group, which, in 1969, presented the OCSS software analyzing and generating reaction pathways [15]. In this study, the logic of the decision-making by the algorithm was based on "externally" assigned laws of organic chemistry. The first algorithms that independently extracted the rules of synthesis planning from the results of clustering data from

reaction databases appeared in the 1990s. The best examples are SYNCHEM, RETROSYN (1990), KOSP (1999), and ARChem Route Designer (2009) [1].

Extensive use of ML for optimizing the conditions of colloidal synthesis began about five years ago. To date, only 0.1% of all publications deal with this field [16]. This is because chemical synthesis is a multiparameter process, which makes its analysis using ML algorithms substantially more complicated than predicting the characteristics or spectra of individual molecules. Additional challenges arise in the formation of relevant databases, not only because conversion of published data on the synthesis conditions and methods into a format suitable for algorithm processing is sometimes difficult, but especially because failed results, which are crucial for training the algorithm, are rarely if at all published. The factors influencing the characteristics of the fabricated colloidal materials are discussed in detail in Section 2.

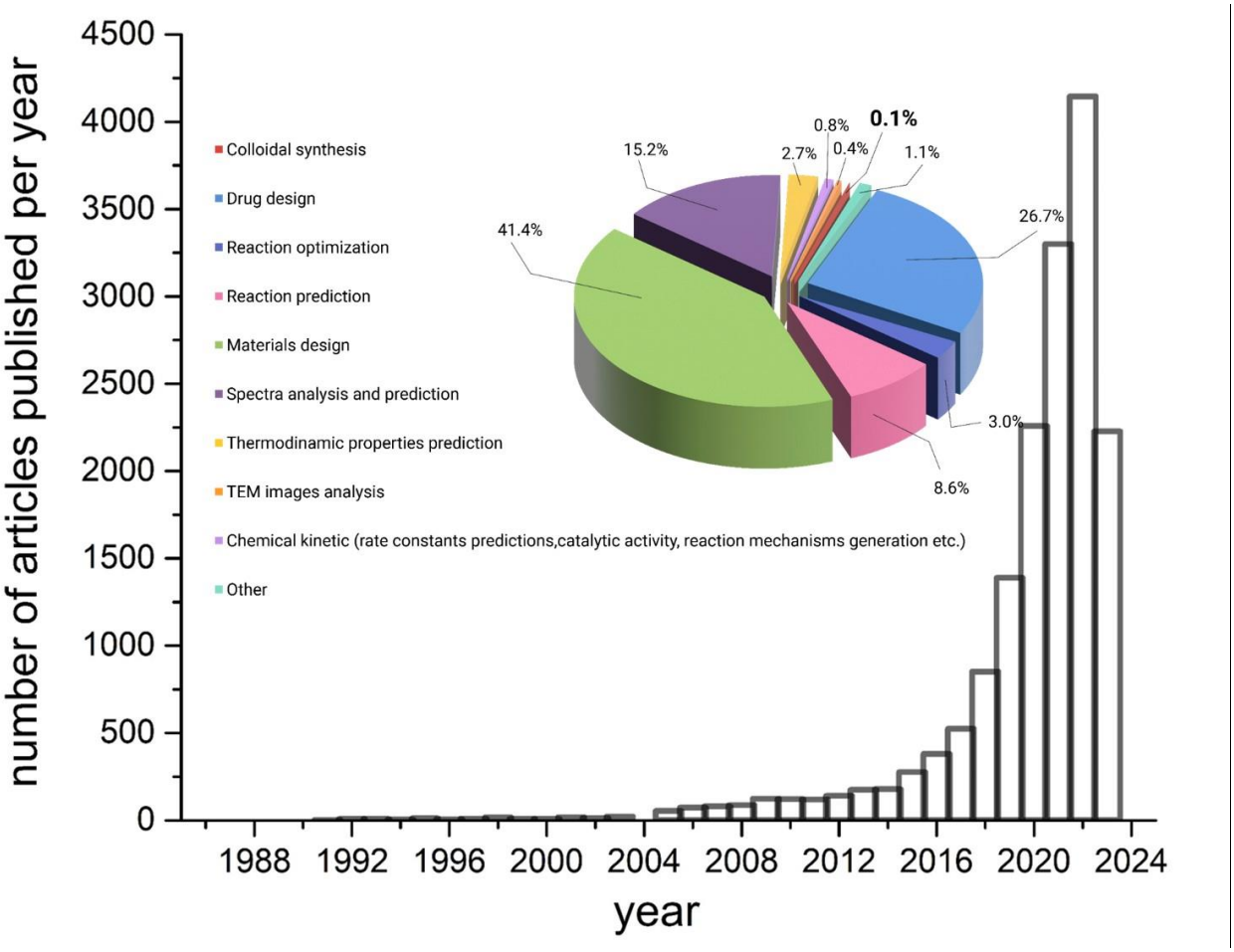

Fig. 2. Statistics of publication activity in the field of machine learning in chemistry from 1985 to July 2023. The pie chart reflects the areas of chemical research where machine learning methods are used most commonly. Source: Web-of-Science [16].

After a brief historical overview, we now turn to the ML algorithms per se and examples of their application to chemistry. Here, the kNN, SVM, DT, random forest (RF), ensemble, and neural network methods are most common.

The *k-nearest neighbors* (kNN) method is one of the simplest yet powerful ones used for both classification and regression (i.e. establishing the relationship between variables that explains the result). In this method, the distance (*d,* Figure 1b) of the classified object to each of the objects of the training sample is calculated, using the main and secondary parameters selected from all available data. First, the distances in terms of the main parameters are calculated; if ambiguity is detected, then the calculations in terms of secondary parameters are performed. Therefore, the key

point of the kNN method is the selection of the metric and the actual number of nearest neighbors *k*.

After the calculations, the object in question is assigned to the class to which most of its nearest neighbors belong. The selection of the parameter *k,* which is set by the researcher, largely affects the final result. If the *k* is small, the algorithm is susceptible to noise, because it is probable that the only nearest neighbor will prove to be incorrectly classified and, hence, the solution will be wrong. If *k* approaches the total number of neighbors *N,* the accuracy of the classification may also be low, because the result will be determined by the most common class of neighbors, with the effect of distance minimized. Thus, extreme values of *k* are unsuitable. The kNN classifier has become widely used in chemistry. The working principle of this algorithm makes it especially promising for determining the relationships between the physical or chemical characteristics and the molecular structure of substances (quantitative structure–property relationship, QSPR). The kNN algorithm has been widely used for predicting the characteristics of various substances upon the assumption that "similar molecules exhibit similar properties" [17, 18]. The list of substance's properties that can be accurately predicted by this method includes the melting temperatures, thermodynamic functions, spectra, solubility, and catalytic and biological activities [17, 19–22].

The simplicity and versatility of the kNN classifier are among its undoubted advantages; however, the working principle of the algorithm entails a number of limitations, such as the need to filter the raw data, susceptibility to errors, difficulty in selecting the *k* value, and increased time and memory consumption in dealing with large data sets. For example, classification of one object based on the distances to 100 nearest neighbors in terms of 100 parameters requires 10,000 calculations.

Support vector machine (SVM) is another algorithm that is used in both regression and classification tasks and works well with small data sets. The principle of the algorithm is to build a hyperplane that divides the objects under study, characterized by a set of descriptors, into classes (Fig. 1c). The task is performed using Lagrange multipliers and represents a quadratic optimization operation [7]. A successful solution is to build a hyperplane where the distances (support vectors) between the closest objects from different classes are maximal. Finding such a hyperplane minimizes the errors of classification; however, it is susceptible to noise if an outlier from the training data set becomes a support vector. SVM is currently widely used in chemistry, primarily in determining QSPRs [23–25]. For example, *Wang et al.* [26] have successfully used SVM to predict the decomposition temperatures of organic peroxides by training the algorithm on a data set of 40 compounds, with molar mass, number of peroxo groups, dipole moment, length and dissociation energy of the oxygen–oxygen bond, etc. used as descriptors. Although SVM is relatively simple, it ensures a high prediction accuracy, often comparable with that of neural networks [23].

The *decision tree* (DT) method is also extensively used in both classification and regression studies. The method is named so because of its hierarchical structure, consisting of "nodes", where the compliance of the classified object with the decision rule is checked, and "leaves" or subsets of data assigned to specific classes upon recursive partitioning. The principle of the method is to partition the parameter space for solving the multiple classification problem. The partitioning criteria are selected on the basis of either the information entropy principle, when the information gain of the resulting node relative to the initial node is evaluated, or the statistical approach, when the probability of incorrect recognition of an example from the training set is predicted. The advantages of the DT method are fast learning and efficient generation of rigorous classification rules. Although the algorithm cannot be considered highly accurate, this parameter can be significantly improved using ensemble methods, as mentioned earlier. Upon combining the results of classification using several DTs, the *random forest* (RF) method can be used to make a more

accurate final prediction. RFs are parallel combinations of DTs; i.e., they are an example of *bagging*. The DT structure is capable of revealing complex interactions between descriptors. The RF algorithm averages the predictions made by the ensemble of DTs to incorporate different trends detected by each tree individually, which results in a complex and reliable model.

The RF algorithm has proved to be more accurate than kNN and SVM, e.g., in determining the upper flammability limits of organic compounds after training the model on a data set of 55 structures [25]. The DT and RF algorithms are widely used in structural chemistry. A striking example is the study by *Oliynyk et al*. [27], where a model for searching new magnetic intermetallides (Heusler compounds) was trained on a data set of 1948 compounds within less than a minute, after which the algorithm made a full set of predictions for more than 400,000 compounds on the basis of 22 descriptors within 45 min. Fig. 1d shows the structure of an individual tree from an RF ensemble.

The DT and RF algorithms have also been shown to be efficient in analyzing and predicting the results of chemical synthesis [28].

Among ensemble methods, the *gradient boosting machine* (GBM) deserves special attention. The main principle of GBM is sequential addition of new models to the ensemble of used methods. At each iteration, a new weak base model is trained, with the error of the whole ensemble learned earlier taken into account. GBM is also typically applied to DTs (*gradient boosting decision trees*). In this case, errors are detected using the gradient of the loss function, which is set by the researcher. The algorithm can also be applied to both classification and regression tasks. The combination of the high accuracy of the results and fast learning has determined wide use of GBM for various fields, from estimating *quantitative structure–activity relationships* [29] to planning chemical synthesis [30] and solving material science tasks [31, 32].

Neural networks constitute a separate branch of AI capable of processing considerably larger amounts of data and are used for prediction, classification, and recognition. Neural networks are so named because their functioning resembles that of the brain.

In this case, "neurons" are computational elements or nodes that process and combine input signals according to certain rules and pass them on to the next network elements. Neural networks used in DL consist of several layers (sets of neurons with common input signals): an input layer, sequentially connected hidden layers, and an output layer. The output of the kth neuron of the $(i + 1)$th layer is calculated as a weighted sum of all contributions of the ith layer, to which the function normalizing the input signal is applied. Mathematically, the neuron model can be written in the form

$$x_k^{i+1} = f\left(\sum_{j=1}^{N} x_j^i \cdot \omega_j^k\right), \qquad (1)$$

where N is the number of preceding neurons, ω is the weight of the connection between the kth and jth neurons, and f is an activation function [33]. The higher the weight, the more prevailing the corresponding information in the next neuron. The structure of connections between the elements of a neural network is usually represented as a weight matrix. In Bayesian neural networks (BNN), probability values are used instead of weights [34]. Depending on the type of the neural network and the task to be performed, the ω_{ij} values may be either static or dynamic, i.e., adjusted during training. Neural networks, like the aforementioned ML algorithms, are subdivided into those with supervised and unsupervised types of learning. The former group usually includes multilayer neural networks, which, in contrast to the perceptron (the simplest single-layer network), contain the so-called hidden layers. In simple architectures, neurons within a layer are not connected to one another. Each node of the given layer receives and transmits signals, respectively, from each neuron of the previous layer and to each neuron of the next layer. The architecture of neural networks is selected depending on the complexity of the system while

preserving a reasonable computation time. The tasks performed by neural networks in chemical applications mostly belong to the supervised learning [35]. In this class of methods, Bayesian optimization is one of the most powerful. This is a probabilistic model analyzing the target function by learning from previous observations. This method has become extensively used in various fields where the conduction of chemical experiments is supported by neural networks [36–40].

Among the variety of multilayer deep neural network (DNN) architectures, convolutional neural networks (CNNs) are most commonly used in chemical research. CNNs are so named because their architecture contains a convolutional layer, which consists of a set of feature maps (matrices) containing a synaptic kernel or filter. The number of maps depends on the complexity of the task, but the ratio between layers is usually 1 : 2; i.e., each map of the current layer corresponds to two maps of the convolutional layer. The kernel or filter is responsible for finding specific features of the analyzed object. It is a system of separable weights, performing the convolution operation per se [41].

Convolutional neural networks have the advantage of efficient parallelization, which significantly reduces the required computational capacity. In CNNs, the convolution operation uses only a small matrix of weights, which is "moved" throughout the processed layer, an activation signal for a neuron of the next layer with the same position being formed after each shift.

The type of the ML algorithm or the neural network is determined by the specifics of the data and the study in general. In solving complex problems, such as the search for dependences between the synthesis parameters and the properties of the resulting materials, the best approach is to use several ML methods to identify the most accurate model. The efficiency of an algorithm is evaluated on the basis of the mean absolute error (MAE) and root-mean-square error (RMSE), as well as the determination coefficient R^2.

Methods of the synthesis of nanomaterials can be divided into physical and chemical ones. The former includes molecular-beam epitaxy, pyrolysis, physical vapor deposition, laser sputtering, etc. These methods make it possible to obtain nanomaterials with well-controlled chemical and phase compositions and a high degree of monodispersity. However, the procedures are usually carried out in high vacuum, and the equipment is complicated and expensive. In addition, the synthesis of nanomaterials by physical methods is poorly scalable. Chemical methods include sol–gel, colloidal, solvothermal, and hydrothermal syntheses, chemical reduction method, etc. Colloidal nanomaterials obtained by these methods are extensively used in the most advanced fields of science, their physical and chemical properties directly depending on the morphology, size, and chemical composition. These parameters are determined by the conditions of synthesis (precursors, solvents, surfactants, temperature, reaction time, additional treatment with ultrasound or microwave radiation, etc.). Variations of these parameters yield numerous possible outcomes of the synthesis, so that performing all of them experimentally would take too much time to be acceptable, given the rapid progress of modern science. The introduction of ML approaches into chemical research can significantly accelerate the obtaining of nanomaterials with desired properties. Different types of synthesis of colloidal nanomaterials adopt ML approaches more or less rapidly. The purpose of this study is to review the state of art in the field of integration of AI in planning of synthesis by the aforementioned methods.

2. Colloidal Synthesis

2.1. Principles of Colloidal Synthesis

Colloidal synthesis is one of the most common methods of obtaining nanomaterials with strictly controlled size, morphology, and composition. This concept combines several approaches.

Synthesis with a coordinating solvent is the approach where the solvent, in addition to its primary role, serves as a stabilizing agent for the colloidal nanocrystals (NCs) [42–44]. In this case, high-boiling hydrocarbons (e.g., octadecene) serve as coordinating solvents [45, 46], and fatty acids (e.g., oleic and stearic acids) or amines of a similar structure serve as additional stabilizers. Thus, varying the stabilizer is an additional factor in controlling the growth of nanoparticles, apart from varying the temperature and nucleation time, which makes it easier to obtain monodisperse NCs of a specified size. In terms of the methods of nucleation initiation, the colloidal synthesis approaches can be divided into injection (hot-injection) and non-injection (heat-up or convection) ones. The former approaches are based on rapid decomposition of organometallic reagents injected into the reaction mixture at a high temperature. This ensures burst nucleation, uniform mixing of the reagents, and a high degree of control over the monodispersity of the obtained NCs [47–49]. The main disadvantages of the injection method are the difficulties in ensuring reproducibility and scalability of the syntheses. The alternative, non-injection approach is often used to obtain large amounts of colloidal nanomaterials. In this case, the reaction medium is rapidly heated when all the precursors have already been added into it. However, this complicates the selection of the precursors, because they should become highly reactive only when the temperature reaches the value that provides predominant growth of NCs of the desired size, remaining unreactive at lower temperatures [50–52]. Otherwise, the obtained NC ensembles would be highly polydisperse.

2.2. Advances of Colloidal Synthesis (Phase-, Morphology- and Size-Controlled Synthesis)

2.2.1. Size control

The particle size considerably affects the physical, optical, and electrical characteristics of nanomaterials compared to their bulk counterparts [53]. Therefore, strict control of colloidal nanoparticle size is crucial in the fabrication of nanomaterials for a wide range of applications, from optoelectronics and photocatalysis to biophysics and medical diagnosis. The NC growth occurs in parallel with the nucleation and is prevailing if the degree of supersaturation is insufficient for massive formation of new nuclei. The thermodynamic basis of the NC growth is the tendency of the system to reduce the surface energy and, hence, form larger particles of the newly formed phase. The kinetics of these processes are complex and specific for each particular system. The NC growth rate, determined by mass transfer processes, depends not only on the supersaturation level, reaction time, and temperature, but also on the viscosity of the reaction medium, which in turn is ruled by the concentration of precursors and other components. In addition, the NC growth rate is also influenced by the ratio between the precursor and stabilizer concentrations, as well as the chemical nature of the stabilizer [49, 54–58]. The type of surface ligands is one of the most important factors determining the growth kinetics of the NCs [59]. The effect of the stabilizer molecules depends on the strength of ligand binding to the nucleus surface and the solubility of the resultant complexes. If the binding is strong, not only the free energy of the nucleus is changed, but also the probability and rate of particle collision are decreased due to the spatial inaccessibility of their surface. The combined effect is a decrease in the nucleation rate. A high solubility of the intermediate complexes may also interfere with the formation of particles of the new phase. In this case, the nucleation rate decreases, while the growth rate of nuclei, on the contrary, increases. For example, *Abe et al*. [60] showed that a 16-fold increase in the ratio between the cadmium precursor and oleic acid concentrations, the other parameters being unchanged, leads to a 40% increase in the diameter of the resultant CdSe quantum dots. A similar trend was also observed for other $A^{II}B^{VI}$ quantum dots [61, 62] and stabilizer molecules with different functional groups [60]. However, the dependence of the size of the formed NCs on the concentration of the

added surfactant is often nonmonotonic [60, 63] and can be considerably changed by using a multicomponent system of stabilizers [63].

According to the classical nucleation theory, the effect of temperature on the size of the synthesized NCs is in a decrease in the critical radius of the nuclei and an increase in their number with increasing temperature. In addition, at higher temperatures, the organometallic precursors decompose faster, which results in a higher concentration of nuclei in the reaction medium. However, if we consider the whole NC growth dynamics, these relationships become more complicated. The temperature affects both the nucleation rate and the growth rate. The latter effect is partly determined by an increased solubility of the precursors and a more intense mass transfer. The dominating temperature effect is different for each specific type of reaction. For example, for colloidal perovskite NCs (PNCs) and lead selenide quantum dots, the NC growth has been found to prevail upon increasing the temperature of synthesis [61, 64, 65].

2.2.2. Shape and phase control

Morphology strongly influences the optoelectronic properties of semiconductor colloidal NCs. Therefore, its variation within a class of compounds holds great promise in the fields of photovoltaics, catalysis, etc. It is believed that the morphology of NCs is primarily determined by the predominant growth of the crystallographic facets that are the most energetically favorable under the given conditions. The surface energy distribution, depending on the class of compounds in question, is influenced by many factors, the most critical of which are temperature [66, 67] and the type of surface ligands, as well as their concentration [68–72]. The *seed-mediated growth* technique [73] is also used to induce the growth of NCs in the specific direction, and the morphology during synthesis can additionally be influenced by multiple injections of the precursor [74]. The literature describes examples of obtaining different NC shapes within the same class of materials, from rods and disks to tetrapod and multipronged structures, or even arrow-shaped forms [75–78]. It is noteworthy that the NC shape can be effectively controlled by changing the reaction temperature alone, with all other synthesis parameters unchanged. Colloidal nanoparticles of the "classical", almost spherical shape are formed under thermodynamic control of growth [75]. If the synthesis is carried out in the kinetic mode, higher-energy facets grow more rapidly and NCs of more complex morphologies are formed.

Specific adsorption of the stabilizer molecules allows the growth rate of the selected facets to be controlled by varying the composition of the functional groups and the carbon backbone of organic ligands, using a multicomponent mixture of the ligands, and changing the ratio between the ligands and the precursors. For example, the synthesis of CdSe NCs in the presence of hexyl phosphonic acid at a low concentration yielded quasi-spherical crystals, whereas an increase in the ligand concentration led to the formation of nanorods [79]. Another study showed that a gradual twofold increase in the proportion of oleic acid in the mixture with oleylamine and carbon disulfide resulted in the formation of different morphological types of NbS_2 NCs: nanorods, hexagonal nanoplatelets, and nanodiscs [69]. A density functional theory study on the specific adsorption of ligands on the surface of CdSe quantum dots [80] showed that, e.g., in the case of a (0001) surface, the surface affinity increased in the row phosphines < phosphine oxides ≈ carboxylic acids < amines < phosphonic acids. It should be noted that the strength of adhesion between a given ligand at a given concentration and the crystallographic facet also depends on temperature [66].

The type and concentration of surface ligands, along with the temperature and time of colloidal synthesis, are also important factors determining the phase composition of nanomaterials [72, 81]. For example, the use of fatty acids as stabilizers in the synthesis of CdTe quantum dots

led to the formation of the wurtzite phase, whereas the use of phosphonic acid resulted in NCs with a zinc blende structure [72].

The composition and structure of surface ligands have an integrated effect on the properties of colloidal NCs. In addition to their strong effect on the size, phase composition, and morphology of nanomaterials, they also determine the solubility and stability of colloidal NCs and affect the transport of free charge carriers. For example, ligands with a long carbon backbone (8 to 18 C atoms) increase the colloidal stability compared to short-chain molecules, but they may cause disruption of charge transfer between NCs [56], which is critical in many optoelectronic applications of colloidal NCs.

3. Application of Machine Learning Methods to Colloidal Synthesis

3.1. Machine Learning in Optimizing Hot-Injection Syntheses of Colloidal Nanomaterials

As noted in the previous section, variation of all the main parameters of colloidal synthesis of nanomaterials (the temperature, types and concentrations of precursors and organic stabilizers, synthesis time, type of solvent, etc.) directly or indirectly affects the size, morphology, and phase composition of the synthesized NCs. This leads to a huge number of the possible outcomes of the synthesis techniques, which would take a long time to test experimentally. Further problems are related to the reproducibility of experiments and the correct choice of the step of variation of synthesis parameters. This is necessary for detecting the possible deviations from monotonic changes in the results, which often are where new discoveries can be made. Therefore, the application of ML methods to the problems of colloidal synthesis optimization is urgently called for in order to obtain nanomaterials with strictly controlled characteristics.

The first problem to be solved for applying mathematical algorithms to optimizing or fine-tuning colloidal synthesis is the collection of initial data sets. Since the first studies on the synthesis of CdS quantum dots by the hot injection method (in the 1980s) until now, some 19,233 research papers on colloidal synthesis have been published [82]. Although the amount of data is obviously sufficient, there are few, if any, data on failed syntheses, which would have been extremely useful for training the algorithms. The training data set on the relationships between the synthesis conditions and the properties of the materials obtained can be compiled both by collecting experimental data obtained in one's own laboratory and by extracting the necessary data from published papers. This extraction is also carried out by special software and ML algorithms. At the first stage, articles containing the corresponding keywords should be found, and the downloaded files should be converted from a specific format (usually PDF) into the plain text. Next, the paragraphs containing the actual data on the synthesis conditions and parameters should be found and extracted from the texts, and the words matching the search terms should be "vectorized". For this purpose, a logistic regression classifier can be used to assign binary categorical labels to the paragraphs, e.g., labels 1 and 0 to the paragraphs related and not related to the synthesis per se. The accuracy of such an algorithm is as high as 95% [83].

Neural networks transforming text words into "mathematical objects" are trained using the rules of category assignment that are set by the researcher when performing this operation manually [83]. There are examples of highly accurate extraction of necessary information (87–93%) not only from the main text, but also from tables and figure captions [84].

Machine learning methods have been increasingly used for optimizing colloidal synthesis and predicting the properties of the obtained materials in recent years. Although publications on this subject are still as few as 0.1% of all articles on AI applications in chemistry (Fig. 2), important and interesting results have already been obtained. The main lines of research in this field are the

search for the composition of colloidal NCs with the desired electronic and structural characteristics [85–90] and the selection of the synthesis parameters for obtaining NCs with the desired sizes, degrees of monodispersity, spectral characteristics [90–94], and morphologies [95, 96].

Due to the unique optical properties of ABX_3 PNCs, where A and B are mono- and divalent cations, respectively, and X are halide anions, they have become widely used in various fields, including the development of light-emitting diodes, photodetectors, solar cells, and photocatalysts [97]. Each of these applications has its own specific requirements for the properties of the materials and, accordingly, their composition. A and B can be cations of various elements or compounds compatible with the perovskite structure, and X can also be a mixed halide or a more complex anion. In this connection, ML techniques are extremely promising for finding optimal combinations of the composition of PNCs with desired electrophysical properties and the conditions of their synthesis. For example, *Saidi et al*. [98] used a hierarchical CNN with 64 and 128 filters in the first and second convolutional layers, respectively, and a fully connected layer of 100 neurons at the output of the second layer to predict the band gap of perovskites of different compositions. Nineteen cations (Cs^+, $[NH_4]^+$, $[CH_3NH_3]^+$, $[CH_3PH_3]^+$, $[CH_3AsH_3]^+$, $[(H_3N)(NH_2)]^+$, $[(CH_2)_3NH_2]^+$, etc.) were used as A^+; lead and tin cations as B^{2+}; and ten combinations of halogen (Cl, Br, and I) anions, as X^-. Twenty-nine elemental features (ionization energy, electron affinity, Goldschmidt tolerance factor, etc.) and six precursor-based features, as well as structural features (lattice constant, etc.), were used to predict the band gaps for a number of compositions with an RMSE as low as 0.02 eV. The SVM, DT, RF, GBM, nonnegative matrix factorization, kernel ridge regression (KRR), and alternating conditional expectation (ACE) methods have also been successfully used to determine the relationship between the PNC band gap and composition [99–102].

Despite the numerous advantages of perovskite nanomaterials, they also have a considerable drawback of a low stability of their luminescence characteristics. *Higgins et al*. [103] used nonnegative matrix factorization and Gaussian process regression to demonstrate that ML efficiently accelerated the experimental search for PNC compositions with the highest stability. The stability of perovskites can be improved by postsynthetic treatment that compensates for the partial desorption of surface ligands. *Yu et al*. [104] developed a model to determine the characteristics of ligands containing amino groups that most strongly affected the stability of $MAPbI_3$ perovskites (out of a total number of 31 characteristics). The model provided an accuracy of 84 and 86% when used together with the SVM and kNN methods, respectively. In addition to the requirements for long-term stability, the engineering of high-performance photodetectors, light-emitting diodes, and solar cells requires that a stable PNC thin film could be obtained. An interesting example of solving this problem is the use of Bayesian optimization (BO), whose effectiveness has been demonstrated in the search for the optimal combination of the solvent and the composition of mixed-halogen organic–inorganic hybrid perovskites [105]. Analysis of the combinations of 240 PNC compositions and 8 solvents from which the film was deposited predicted the following best composition–solvent combinations for obtaining stable thin films: $FAPbI_2Cl$, $FAPbI_2Cl$, and $CsPbI_3$ in tetrahydrothiophene-1-oxide and $CsPbICl_2$ in DMSO. The relationship between the synthesis parameters and the morphology of the resultant colloidal NCs is also successfully identified by ML methods [96, 106]. Braham et al. [97] showed that the SVM classifier/regressor made it possible to analyze the effect of the type of ligands on the formation of $CsPbBr_3$ nanoplatelets and to predict their thickness down to a submonolayer accuracy. The analysis was performed on the basis of experimental data on hot-injection colloidal syntheses of

53 samples. Different reaction parameters were analyzed independently: the reaction temperature was varied at a fixed Pb^{2+} to alkylamine concentration ratio; in turn, the concentration of surface ligands and the length of their carbon chain (from 4 to 18 C atoms) were varied at a fixed temperature (100°C). The results made it possible to establish a set of combinations of several parameters for obtaining two-dimensional PNCs with precisely specified thickness and, hence, photoluminescence characteristics. For example, according to the model, the synthesis of $CsPbBr_3$ PNCs at relatively low injection temperatures (50 and 82°C) and high concentrations of long-chain alkylamines (with low diffusion coefficients) is expected to yield mostly platelets no thicker than three monolayers (Fig. 3a, 3b).

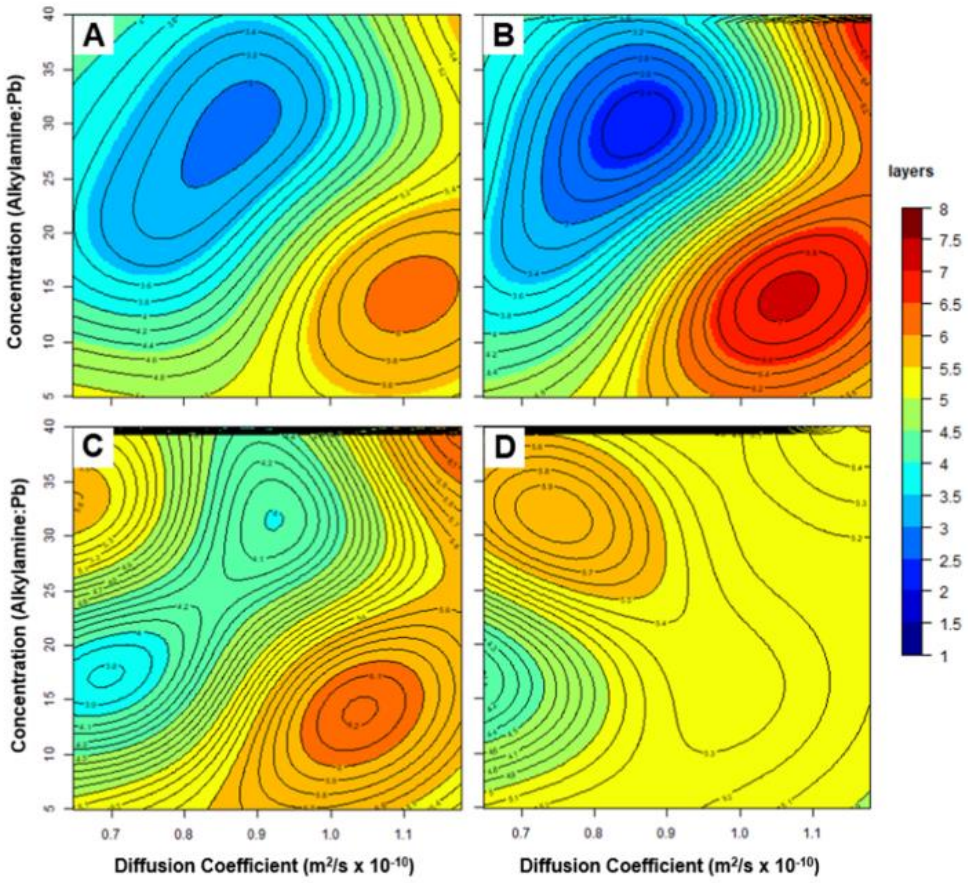

Fig. 3. Contour plot slices of the SVM regression used to analyze the effect of the type of ligands on the formation of $CsPbBr_3$ nanoplatelets at temperatures of (A) 50 °C, (B) 82 °C, (C) 120 °C, and (D) 150 °C. Reprinted with permission from Ref. [96]. Copyright 2019, American Chemical Society.

The SVM method was used to precisely determine the conditions for the preparation of two-dimensional organic–inorganic perovskites [106], including the necessary set of chemical properties of the ligands. Neural network models are extremely efficient in predicting the parameters of successful synthesis of metal halide perovskites with the desired photoluminescence quantum yield and photoluminescence peak width and position. For example, an ensemble neural network trained on the data on 1000 real syntheses was used to automatically test 150 decision-making strategies for over 600,000 simulated PNC synthesis experiments and ion exchange reactions in a microfluidic flow reactor within a single training period, which is equivalent to 7.5 years of permanent experimental work and consumption of about 400 L of reagents [107]. An ensemble neural network was also used to automate the synthesis of colloidal $CsPbBr_3$ NCs and

the subsequent ion exchange reaction yielding ten reproducible products with specified spectral characteristics [108]. Eight input parameters of the synthesis in a microfluidic flow reactor followed by the anion exchange reaction without an intermediate purification step formed a large parameter space of about 2×10^7 possible combinations; however, the neural network selected the optimal parameters after training on as few as 250 real experiments.

Deep neural networks are also successfully used for identifying causal relationships between the parameters of colloidal synthesis and the properties of quantum dots. Liu et al. [95] used a neural network with a dimension of $n \times 8 \times 8 \times m$ to determine the relationship of the type and concentration of the surface ligands, as well as the temperature, with the morphology of CdSe/CdS core/shell quantum dots (Fig. 4). Preliminary results showed that the synthesis of these quantum dots with oleic acid, octadecyl phosphonic acid, and hexyl phosphonic acid as ligands resulted in structures with complex morphology called "tadpoles". The ligand concentrations and the temperature served as the n values in the neural network architecture; m corresponded to the number of "tails" of the CdSe/CdS quantum dots.

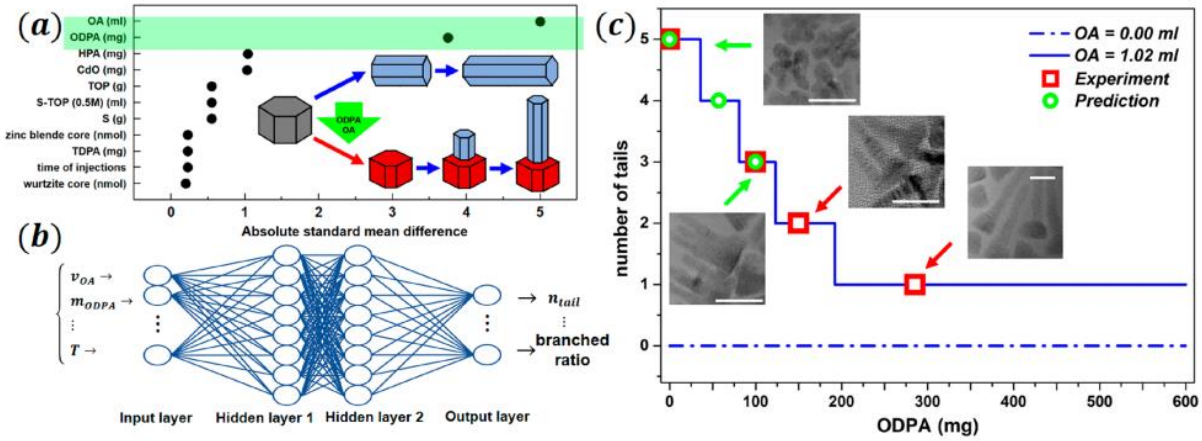

Fig. 4. Deep neural networks use for identifying causal relationships between the parameters of colloidal synthesis and the properties of quantum dots. (a) Causal relationship for the transition from quantum rod to tadpole. The standardized mean difference (SMD) for oleic acid (OA) and octadecyl phosphonic acid (ODPA) are the highest. The figure inserted shows that the growth route can be altered by the addition of OA and ODPA. (b) Detailed structure of the neural network. The *Hidden layer 1* and the *Output layer* both take a *tanh* activation function and a linear operation afterward, while the others are purely linear. The dimension of the neural network is $n \times 8 \times 8 \times m$, where n is the number of controls, and m is the number of QD features. In the experiment present, $n = 4$ refers to the amount of OA, ODPA, HPA, temperature, and $r = 1$ refers to the number of tails, respectively. (c) Neural network predicted causal relationship between the amount of ODPA and the number of tails at 320°C, trained with the experimental data. The solid red rectangles refer to the experimental data with tail numbers 1 and 2 and their corresponding TEM images; the green circles refer to the predicted structures made by the neural network with 3, 4, and 5 tails. The scale bars in the figures denote represent 20 nm. Reprinted with permission from Ref. [96]. Copyright 2020, American Chemical Society.

The model used in that study confirmed the experimental data on the parameters for obtaining structures with the number of "tails" from 0 to 2 and predicted the possible formation of heterostructures containing three, four, and five "tails", whose directed synthesis was subsequently carried out.

Voznyy et al. [91] tested a neural network that had two hidden layers with 20 nodes in a study with eight input parameters (the ambient temperature, the temperature and injected volume of the precursors, etc.) and two output parameters (the bandgap and the half width at half maximum of the luminescence spectrum). Training of this neural network on the data on 2300 syntheses of PbS quantum dots using a common personal computer took about 6 min per cycle of 15,000 epochs (runs throughout the training data set). As a result, the parameter space of the synthesis of PbS quantum dots with model predictions for almost 1,000,000 data points was simulated in less than a minute [91]. This analysis, which was aimed at finding the key synthesis factors responsible for the size of the quantum dots and the degree of their monodispersity, yielded detailed data on the effects of the oleic acid concentration, the Pb to S ratio, the precursor injection temperature, and the addition of lead and cadmium chlorides on these parameters. The resultant model made it possible to visualize the effect of each parameter, to reveal complex nonlinear dependences, and to identify the parameters that had the strongest effects on the properties of the synthesized quantum dots (Fig. 5).

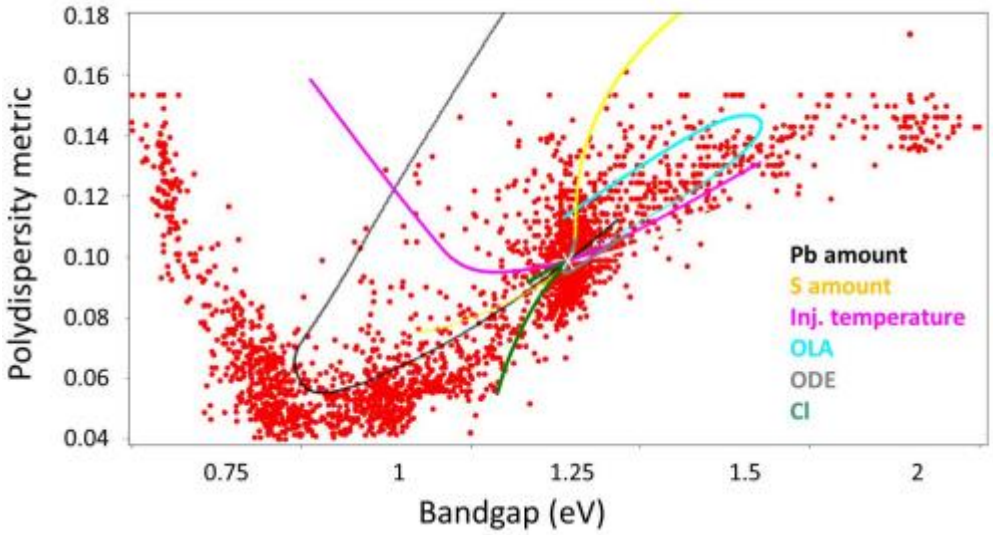

Fig. 5. ML model predictions for the behavior of wavelength and polydispersity of PbS QDs when one of the input parameters swipes the entire range of allowed values. X marks the conventional synthesis employed for colloidal QDs with 950 nm excitons. Cl is the single most noticeable parameter improving monodispersity. However, Pb and injection temperature are not parallel and thus do not cancel each other, allowing for a positive combined effect on monodispersity. The curving tail of the black line (Pb) in the lower bandgap range reflects that when Pb/S ratio is too high, the synthesis gets blocked, as represented by the topmost red point at 2 eV. Reprinted with permission from Supporting Information to Ref. [91]. Copyright 2019, American Chemical Society.

Predicted parameters for synthesizing a number of quantum dots with different bandgaps that would have the highest possible monodispersity have been tested experimentally. For example, to synthesize monodisperse lead sulfide quantum dots with an absorption maximum at 620 nm, which is a nontrivial task, the following parameters are required: 3.35, 15, 0.54, and 0.168 mL of lead oleate, octadecene, and $(TMS)_2S$ solutions, respectively, and an injection temperature of 53°C [91]. Hitting this combination of parameters by the classical trial-and-error method, sequentially synthesizing a number of samples without the use of artificial intelligence, seems to be highly improbable.

Training the GBM algorithm on a set of data on 842 syntheses [109] yielded the parameters of injection colloidal synthesis (the synthesis time, injection temperature, and types and concentrations of precursors and surface ligands) required for synthesizing quantum dots from semiconductors of groups II–VI and IV–VI of a strictly specified size. The R^2 values for the radii of CdS, CdSe, PbS, and ZnSe quantum dots were 0.74, 0.93, 0.99, and 0.80, respectively [109]. The DT and RF methods exhibited a lower prediction accuracy.

3.2. Machine Learning in Optimizing Colloidal Synthesis of Metal Nanoparticles

Chemical reduction is one of the most common methods of synthesizing colloidal nanoparticles of metals (gold, silver, iron, cobalt, nickel, etc., as well as alloy compositions) in solution because these reactions are easy to conduct and the size and morphology of the resultant samples can be controlled by varying the reaction parameters. Nucleation in this case occurs when the metal salt solution is reduced to zero-valent atoms that form crystallization centers and diffuse into the solution to form aggregates. The aggregates pass into the growth stage when they reach the critical radius. The size and morphology of the metal nanoparticles obtained in this way can be controlled by proper selection of the solvent, reducing agent, and type of surface stabilizing ligands, as well as by the use of seeds (preliminarily obtained smaller NCs) [110].

Li et al. [111] successfully used ML methods to develop an efficient synthesis of gold nanoclusters with a specified shape/size. They reported an example of an efficient model trained on a small set of data on 27 syntheses described in the literature and 27 syntheses carried out by the authors themselves. Seventeen parameters of the Au(III) reduction reaction were considered, including the concentrations of $HAuCl_4$, $NaBH_4$, and stabilizing ligands, as well as temperature, pH, and solvent polarity. Among the six ML methods tested, including SVM, DNN, Siamese neural network (SNN), and their combinations with the graph convolutional neural network (GCNN), the GCNN + DNN and GCNN + SNN combinations exhibited the highest accuracy. The SNN architecture contains two or more neural networks that have the same weights and parameters but use different input vectors for calculating comparable output vectors. This specific structure makes SNNs one of the most popular types of neural networks for processing small samples [112]. The main difference between the standard CNNs and GCNNs is that the former operate with ordinary (Euclidean) structured data, whereas in the latter, nodes are disordered (irregular on non-Euclidean structured data).

Recurrent neural networks, which are much more rarely used for the optimization of colloidal nanoparticle synthesis, are the basis of the *Deep Reaction Optimizer* algorithm. It has proved to be efficient in determining the optimal $AgNO_3$, $NaBH_4$, and sodium citrate concentrations for obtaining colloidal silver nanoparticle solution with the desired position of the plasmon resonance maximum in the optical extinction spectrum [113]. Du et al. [114] performed a large-scale study on determining the relationship between the spectral characteristics and the structural parameters of plasmonic nanoparticles by DL methods. The parameters of a model based on CNNs were optimized on the basis of 1,017,578 samples extracted from the literature. Hyperparameters (number of neurons) were established using data on 127,197 samples. The same amount of data was used to estimate the accuracy of the model, which was 99.4 ± 0.6% for spherical nanoparticles and 98 ± 2% for nanorods. The combination of BO with a DNN containing four hidden layers with 50, 100, 200, and 500 nodes, respectively, was used to determine the effect of each component of the reaction mixture on the optical properties of silver nanoparticles synthesized in a microfluidic reactor [115]. The input parameters were the ratios of the fluxes (Q) of the reagents, including silver NC seeds (Q_{seeds}), into the output layer consisting of 421 nodes

corresponding to the UV–Vis spectral data points. The spectra calculated using this algorithm closely matched the experimentally measured spectra (Fig. 6).

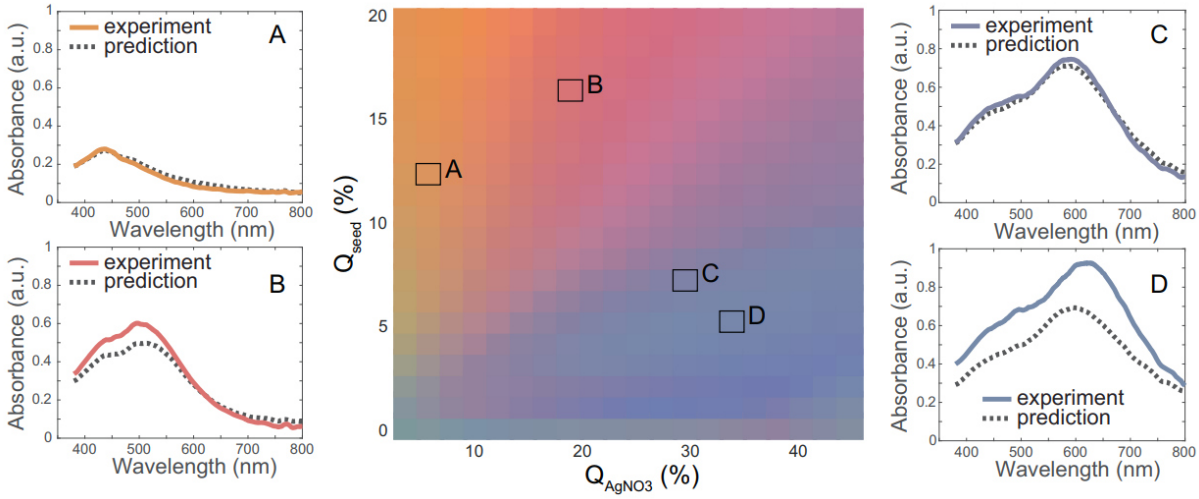

Fig. 6. Knowledge extraction on the silver nanoprism synthesis. Map of the AgNPs colours predicted by the DNN teacher in the {Q_{AgNO3} , Q_{seed}} space, for a fixed value of polyvinyl alcohol flux Q_{PVA} = 16%, trisodium citrate Q_{TSC} = 6.5% and Q_{total} = 850 µL/min, corresponding to the conditions of the DNN best performance in run 8. Predicted spectra (dotted grey line) were extracted in four regions of the space (A, B, C, and D) and compared to experimental spectra (plain coloured line) obtained for similar conditions [115].

This two-step method proved to be more efficient than the neural network alone and made it possible to determine how the ratio between the silver nitrate ($QAgNO_3$) and seed flux rates affected the spectrum parameters (Figure 6).

Bimetallic nanoparticles can be obtained by co-reduction of two metal precursor salts. In this case, the metal with a higher redox potential forms the core, and the shell is formed when the reduction of the second metal begins. The sequence of reduction of the metals in the reaction mixture can be controlled by adding ligands that form more stable complexes with the metal whose redox potential is higher, thereby preventing the formation of M^0 clusters. In addition to the redox potentials, the compositions of the core and shell are also affected by the ratio of atomic radii and the differences in the coordination numbers and cohesion energies. The components in the bimetallic nanoclusters can form not only the "ideal" core/shell structure, but also relatively disordered ones. In terms of the data representation format, DT classifier algorithms are convenient models for identifying the key factors that determine the preferred composition of the core or shell in the given metal pair. The accuracies of predicting the composition of bimetallic clusters by the RF, random tree and reduced error pruning tree (C45 modification) methods have been compared [116]. For this purpose, data on 903 bimetallic nanoclusters were treated using the density functional theory methods to identify 641 metal pairs with unambivalent differentiation of the core and shell compositions (in the other cases, the simulation predicted the formation of either mixed nanoclusters or Janus structures). Application of the aforementioned ML methods to the selected sample showed that the key factors determining the chemical composition were the difference in cohesion energy, the atomic radii, and the coordination numbers. The RF method was found to be the most accurate (about 90%).

3.3. Machine Learning in Optimizing Hydrothermal and Solvothermal Syntheses of Colloidal Nanomaterials

Hydrothermal and solvothermal syntheses are two other common methods for obtaining colloidal nanomaterials. In this approach, the morphology of the synthesized NCs is controlled by varying the temperature and pressure in the reactor within a wide range and by changing the composition of the reaction medium. Conducting the reaction in a sealed autoclave at a temperature above the boiling point of the solvent makes it possible to reach a pressure considerably higher than the atmospheric one. This ensures the formation of nanomaterials with a high degree of crystallinity. Hydrothermal synthesis is used to obtain various nanomaterials, from carbon quantum dots and wide-bandgap semiconductor oxides to magnetic nanoparticles [117]. Variation of the time, temperature, and, hence, pressure of the hydrothermal synthesis and the composition of the reaction medium allows nanomaterials with a wide range of nanostructure morphologies to be fabricated [118, 119]. For example, *Sutar et al*. [120] extracted experimental data from studies published in the period from 2005 to 2020 to identify 298 different sets of experimental conditions for hydrothermal synthesis of zinc oxide yielding particles with different morphologies. Then, they applied the DT, RF, and neural network methods to the extracted data to determine a set of "rules" and consistent relationships that ensure obtaining ZnO nanoparticles with the desired morphology, as well as solar cells based on them with the desired power conversion efficiency. Various synthesis conditions, including the type of the target structure, precursors, temperature and time of the hydrothermal reaction and the presence or absence of seeds, were used as input parameters for the model. The output power conversion efficiencies of ZnO-based solar cells predicted by the RF algorithms and neural networks were sufficiently accurate, with determination coefficients of 0.7232 and 0.7447, respectively. ML-based analysis showed that the power conversion efficiency was the highest if zinc acetate was used as a precursor, two- and three-dimensional ZnO nanostructures were formed, the reaction time was short, and the temperature of the reaction medium was high. The DT algorithm made it possible to accurately identify the parameters of hydrothermal synthesis that determined specific morphologies of the resultant zinc oxide nanoparticles. For example, if seeds were added during the growth of ZnO NCs, one-dimensional nanostructures were obtained with a probability of 87%; in the absence of seeds, this structure was formed through decomposition of zinc nitrate in the temperature range from 168 to 190°C with a probability of 35%. Application of the DT algorithm to data from 22,065 journal articles and 27 varying synthesis parameters [121] yielded the key parameters (temperature and NaOH concentration) of the hydrothermal synthesis of TiO_2 for obtaining nanoparticles in the form of nanotubes with an accuracy of 82%.

Han et al. [122] used ML methods (regressors based on XGBoost gradient boosting, SVM, and the Gaussian process) to optimize the hydrothermal synthesis of fluorescent carbon quantum dots, which are used to detect Fe^{3+} ions. The models were trained on the results of 391 experiments, the input parameters being the temperature, heating rate, reaction time, volume of ethylenediamine (a stabilizing agent), and weight of hydroquinone (a carbon precursor). The XGBoost-based regression algorithm exhibited the highest R^2 and the lowest RMSE. The use of ML made it possible to determine that the amounts of ethylenediamine and hydroquinone were among key factors affecting the fluorescence quantum yield of carbon quantum dots. The authors also determined the range ethylenediamine concentrations within which this parameter was decisive, as well as the threshold values of precursor loading at which successful synthesis became unfeasible.

Gradient boosting (using, e.g., the XGBoost algorithm) is an effective approach to developing supervised regression models. The model allows for fast parallel data processing and can regularize the loss function, which processes the results by comparing them with known data, thereby facilitating the estimation of the prediction quality for a given training example. In another study [123], XGBoost was used to optimize the synthesis of carbon quantum dots at room temperature. This model proved to be more accurate than the SVM, kNN, DT, RF, and CNN models. Using the results of 400 experimentally performed syntheses as input data, the XGBoost algorithm predicted the optical characteristics of the carbon quantum dots as dependent on the input data (synthesis time, solvent, and precursor concentrations) with a high accuracy ($R^2 > 0.96$).

Machine learning methods are being rapidly integrated into all stages of the fabrication, study, and use of colloidal nanomaterials, which promotes both research and technological progress. Table 1 shows examples of the use of AI at all stages of obtaining colloidal nanomaterials, from prediction of their properties to practical applications in optoelectronics.

Table 1. Machine learning methods in colloidal nanomaterial synthesis and related fields.

Application area		Object/subject/purpose of research	Machine learning and deep learning methods	References
Colloidal nanomaterial synthesis	Injection synthesis	ABX_3 perovskite nanocrystals	Support vector machine	[98, 106]
			Ensemble neural network	[107, 108]
		$A^{II}B^{VI}$ and $A^{IV}B^{VI}$ semiconductor quantum dots and core/shell structures	Deep neural network	[91, 95]
			Gradient boosting machine	[92]
	Hydrothermal synthesis	ZnO nanoparticles	Decision tree, random forest, neural network	[120]
		TiO_2 nanoparticles	Decision tree	[121]

		Carbon quantum dots	XGBoost, multilayer perceptron, support vector machine, Gaussian process regressors	[122]
			XGBoost regressor	[123]
	Chemical reduction	Au nanoclusters	Support vector machine, deep neural network, Siamese neural network, graph convolutional neural network + deep neural network, graph convolutional neural network + Siamese neural network	[111]
		Ag nanoparticles	Recurrent neural network	[113]
			Bayesian optimization + deep neural network	[115]

Characterization of colloidal nanomaterials	Analysis of spectroscopy data	Finding analytical relationship between the descriptors of absorbance spectra and the structural parameters of Au nanoparticles	Light gradient boosting machine, radial-basis functions	[124]
		Analysis of time-resolved photoluminescence spectra. Estimation of the decay rates of photoluminescence of AgInSe and $MAPbI_3$ colloidal nanocrystals	LumiML software	[125]
		Predicting the effect of air humidity on the intensity of photoluminescence of $MAPbBr_3$ and $MAPbI_3$	Echo state network	[126]
	Microscopy image recognition	Qualitative and quantitative recognition of the morphology and size of nanoparticles in electron microscopy images	Bayesian deep learning model	[127]
			Modified super-resolution convolutional neural network	[128]
		Quantitative analysis of holographic microscopy images. Determining the size, morphology, and composition of colloidal nanoparticles	Haar cascade classifiers + convolutional neural network	[129]
		Analysis of XRD data on perovskite nanocrystals	Deep feedforward neural network	[130]
Prediction of the properties of colloidal nanoparticles		Predicting the bandgap widths of ABX_3 perovskite nanocrystals and $A^{IV}B^{VI}$ quantum dots	Kernel ridge regression, alternating conditional expectations, gradient boosting	[99]

		machine, decision tree	
		Support vector regression	[131]
	Predicting the physical and chemical properties and biological activities of Au, Pt, and Pd nanoparticles	Convolutional neural network	[132]
	Predicting the optical properties of carbon quantum dots	Deep convolutional neural network	[133]
	Determining relationships between various structural, morphological, and geometrical characteristics of Ag nanoparticles and their Fermi energy	ANN regression, logistic regression, random forest, principal component analysis, k-fold cross validation	[134]
Nanophotonic applications	Designing photonic–plasmonic nanostructures based on Ag nanowires	Physics-guided neural network	[135]
	Designing thin-film nanophotonic heterostructures	Convolutional neural network	[136-138]
	Designing plasmonic sensors	Least absolute shrinkage and selection operator	[139]
	Determining the key factors in the stability of blue OLED. Analysis of current density–voltage–luminance relationships, impedance spectra, and operational lifetime	Convolutional neural network, decision tree	[140]

4. Conclusions and Perspective

The progress in ML and DL methods has made it possible to use them for analyzing complicated multiparameter processes, such as the synthesis of colloidal nanomaterials. This area is relatively new, and the number of such studies compared to those on AI integration in the analysis of spectra and microscopy images, construction models of organic retrosynthesis, and solution of QSPR tasks is still small; however, interesting and important results have already been obtained in constructing models of colloidal NC synthesis. Among the variety of colloidal NCs, the current focus in this area is on II–VI and IV–VI quantum dots, ABX_3 perovskites, plasmonic nanoparticles, and oxide semiconductors. This approach ensures a high accuracy of finding consistent relationships between the synthesis parameters and the properties of the nanomaterials, determining the compositions with the best structural and electrophysical properties and stability indices, revealing hidden composition–structure–property relationships, and establishing the conditions for efficient postsynthetic stabilization and interaction with solvents. Analysis of published studies has shown that neural networks are the most common tools for performing these tasks, but relatively simple classical ML methods and algorithms, such as SVM, GBM, RF, and KRR, are also highly effective. The accuracy of the predictions made by the algorithms has been confirmed by both experimental and computational methods. Given the success of earlier applications of ML approaches to the recognition of spectra and electron microscopy images, the use of mathematical algorithms at all stages of obtaining and characterization of colloidal nanomaterials seems extremely promising. AI makes it possible to do in a matter of minutes what would take months or years of experimental work of a researcher. Therefore, further effective integration of AI in this field is extremely promising. This could reduce empiricism in chemical synthesis, increase the efficiency of research, and promote the discovery of new materials and mechanisms of chemical reactions, thus leading to a major breakthrough in science. It is no overstatement to say that the potential of ML in chemistry is limitless. At the moment, however, there are aspects that could be optimized. Analysis of the published studies leads to the conclusion that finding and extracting the necessary data for training the algorithm is a fairly serious problem. Compiling databases of a new type specifically designed for the use in building predictive synthesis models seems promising. These databases should contain reliable verified data on the synthesis conditions and characteristics of nanomaterials of different classes, be unified, and be open to researchers from all over the world.

Acknowledgements

The part of the study related to the protocols of synthesis of nanomaterials was supported by the Russian Science Foundation (RSF), grant no. 18-72-10143, and the part related to the machine learning approaches to this synthesis was supported by the RSF grant no. 21-79-30048. IN acknowledges support from the French National Research Agency ANR, grant no. ANR-20-CE19-009-02, and from the Université de Reims Champagne-Ardenne.

References

[1]. T. E. H. Allen, Computers as scientist. In Machine Learning in Chemistry: The Impact of Artificial Intelligence, Cartwright, H. M., Ed. Royal Society of Chemistry, 2020.

[2] A. Barr, E. A. Feigenbaum, The Handbook of artificial intelligence, volume 1. Los Alto, Calif., William Kaufmann, 1981.

[3] A. M. Turing, Computing machinery and intelligence, Mind. 59 (1950) 433 – 460.

[4] P.V.S. Rao, Computer system architecture. New Delhi, PHI Learning Pvt Ltd., 2009.

[5] T. Ishiyama, F. Prada, A.A. Klypin, M. Sinha, R.B. Metcalf, E. Jullo, B. Altieri, S.A. Cora, D. Croton, S. de la Torre, D.E. Millán-Calero, T. Oogi, J. Ruedas, C.A. Vega-Martínez, The Uchuu simulations: Data Release 1 and dark matter halo concentrations, Mon. Not. R. Astron. Soc. 506 (2021) 4210–4231. https://doi.org/10.1093/mnras/stab1755.

[6] Top 500. The List. https://top500.org/ (accessed 30 May 2022)

[7] R. Czermiński, A. Yasri, D. Hartsough, Use of Support Vector Machine in Pattern Classification: Application to QSAR Studies, Quant. Struct. Relationships. 20 (2001) 227–240. https://doi.org/10.1002/1521-3838(200110)20:3<227::AID-QSAR227>3.0.CO;2-Y.

[8] S. Cohen, The basics of machine learning: strategies and techniques, in: Artif. Intell. Deep Learn. Pathol., Elsevier, 2021: pp. 13–40. https://doi.org/10.1016/B978-0-323-67538-3.00002-6.

[9] M. Kusaba, C. Liu, Y. Koyama, K. Terakura, R. Yoshida, Recreation of the periodic table with an unsupervised machine learning algorithm, Sci. Rep. 11 (2021) 4780. https://doi.org/10.1038/s41598-021-81850-z.

[10] T. J. Sejnowski, The Deep Learning Revolution, Ed. The MIT Press. 2018.

[11] J. Gasteiger, T. Engel, eds., Chemoinformatics, Wiley, 2003. https://doi.org/10.1002/3527601643.

[12] National Library of Medicine. Explore Chemistry. https://pubchem.ncbi.nlm.nih.gov/ (accessed 27 July 2023)

[13] B.R. Kowalski, P.C. Jurs, T.L. Isenhour, C.N. Reilley, Computerized learning machines applied to chemical problems. Multicategory pattern classification by least squares, Anal. Chem. 41 (1969) 695–700. https://doi.org/10.1021/ac60275a026.

[14] P.C. Jurs, B.R. Kowalski, T.L. Isenhour, C.N. Reilley, Computerized learning machines applied to chemical problems. Molecular structure parameters from low resolution mass spectrometry, Anal. Chem. 42 (1970) 1387–1394. https://doi.org/10.1021/ac60294a015.

[15] E.J. Corey, W.T. Wipke, Computer-Assisted Design of Complex Organic Syntheses, Science (80-.). 166 (1969) 178–192. https://doi.org/10.1126/science.166.3902.178.

[16] Web of Science. https://webofscience.com (accessed 29 July 2023).

[17] F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch, J.B.O. Mitchell, Melting Point Prediction Employing k -Nearest Neighbor Algorithms and Genetic Parameter Optimization, J. Chem. Inf. Model. 46 (2006) 2412–2422. https://doi.org/10.1021/ci060149f.

[18] S. Chinta, R. Rengaswamy, Machine Learning Derived Quantitative Structure Property Relationship (QSPR) to Predict Drug Solubility in Binary Solvent Systems, Ind. Eng. Chem. Res. 58 (2019), 3082–3092. https://doi.org/10.1021/acs.iecr.8b04584.

[19] A. Varnek, N. Kireeva, I. V. Tetko, I. I. Baskin, V. P. Solov'ev, Exhaustive QSPR Studies of a Large Diverse Set of Ionic Liquids: How Accurately Can We Predict Melting Points? J. Chem. Inf. Model. 47 (2007), 1111–1122. https://doi.org/10.1021/ci600493x.

[20] W. Jia, Z. Yang, M. Yang, L. Cheng, Z. Lei, X. Wang, Machine Learning Enhanced Spectrum Recognition Based on Computer Vision (SRCV) for Intelligent NMR Data Extraction, J. Chem. Inf. Model. 61 (2021) 21–25. https://doi.org/10.1021/acs.jcim.0c01046.

[21] M. Shen, Y. Xiao, A. Golbraikh, V.K. Gombar, A. Tropsha, Development and Validation of k -Nearest-Neighbor QSPR Models of Metabolic Stability of Drug Candidates, J. Med. Chem. 46 (2003) 3013–3020. https://doi.org/10.1021/jm020491t.

[22] W. Yang, T.T. Fidelis, W.-H. Sun, Machine Learning in Catalysis, From Proposal to Practicing, ACS Omega. 5 (2020) 83–88. https://doi.org/10.1021/acsomega.9b03673..

[23] R. Burbidge, M. Trotter, B. Buxton, S. Holden, Drug design by machine learning: support vector machines for pharmaceutical data analysis, Comput. Chem. 26 (2001) 5–14. https://doi.org/10.1016/S0097-8485(01)00094-8.

[24] X. Yu, Support vector machine-based QSPR for the prediction of glass transition temperatures of polymers, Fibers Polym. 11 (2010) 757–766. https://doi.org/10.1007/s12221-010-0757-6.

[25] S. Yuan, Z. Jiao, N. Quddus, J.S.-I. Kwon, C. V. Mashuga, Developing Quantitative Structure–Property Relationship Models To Predict the Upper Flammability Limit Using Machine Learning, Ind. Eng. Chem. Res. 58 (2019) 3531–3537. https://doi.org/10.1021/acs.iecr.8b05938..

[26] B. Wang, H. Yi, K. Xu, Q. Wang, Prediction of the self-accelerating decomposition temperature of organic peroxides using QSPR models, J. Therm. Anal. Calorim. 128 (2017) 399–406. https://doi.org/10.1007/s10973-016-5922-8.

[27] A.O. Oliynyk, E. Antono, T.D. Sparks, L. Ghadbeigi, M.W. Gaultois, B. Meredig, A. Mar, High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds, Chem. Mater. 28 (2016) 7324–7331. https://doi.org/10.1021/acs.chemmater.6b02724.

[28] H. Huo, Z. Rong, O. Kononova, W. Sun, T. Botari, T. He, V. Tshitoyan, G. Ceder, Semi-supervised machine-learning classification of materials synthesis procedures, Npj Comput. Mater. 5 (2019) 62. https://doi.org/10.1038/s41524-019-0204-1.

[29] R.P. Sheridan, W.M. Wang, A. Liaw, J. Ma, E.M. Gifford, Extreme Gradient Boosting as a Method for Quantitative Structure–Activity Relationships, J. Chem. Inf. Model. 56 (2016) 2353–2360. https://doi.org/10.1021/acs.jcim.6b00591.

[30] M.R. Maser, A.Y. Cui, S. Ryou, T.J. DeLano, Y. Yue, S.E. Reisman, Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions, J. Chem. Inf. Model. 61 (2021) 156–166. https://doi.org/10.1021/acs.jcim.0c01234.

[31] B. Zheng, G.X. Gu, Prediction of Graphene Oxide Functionalization Using Gradient Boosting: Implications for Material Chemical Composition Identification, ACS Appl. Nano Mater. 4 (2021) 3167–3174. https://doi.org/10.1021/acsanm.1c00384.

[32] G. Wang, T. Fearn, T. Wang, K.-L. Choy, Machine-Learning Approach for Predicting the Discharging Capacities of Doped Lithium Nickel–Cobalt–Manganese Cathode Materials in Li-Ion Batteries, ACS Cent. Sci. 7 (2021) 1551–1560. https://doi.org/10.1021/acscentsci.1c00611.

[33] I.N. da Silva, D. Hernane Spatti, R. Andrade Flauzino, L.H.B. Liboni, S.F. dos Reis Alves, Artificial Neural Networks, Springer International Publishing, Cham, 2017. https://doi.org/10.1007/978-3-319-43162-8.

[34] M.B. Rao, C.R. Rao, Bayesian Networks, in: 2014: pp. 357–385. https://doi.org/10.1016/B978-0-444-63431-3.00010-3.

[35] S. Alarie, C. Audet, A.E. Gheribi, M. Kokkolaras, S. Le Digabel, Two decades of blackbox optimization applications, EURO J. Comput. Optim. 9 (2021) 100011. https://doi.org/10.1016/j.ejco.2021.100011.

[36] H. Gao, S. Zhong, W. Zhang, T. Igou, E. Berger, E. Reid, Y. Zhao, D. Lambeth, L. Gan, M.A. Afolabi, Z. Tong, G. Lan, Y. Chen, Revolutionizing Membrane Design Using Machine Learning-Bayesian Optimization, Environ. Sci. Technol. 56 (2022) 2572–2581. https://doi.org/10.1021/acs.est.1c04373.

[37] B. Yildirim, J.M. Cole, Bayesian Particle Instance Segmentation for Electron Microscopy Image Quantification, J. Chem. Inf. Model. 61 (2021) 1136–1149. https://doi.org/10.1021/acs.jcim.0c01455.

[38] Y. Xie, C. Zhang, H. Deng, B. Zheng, J.-W. Su, K. Shutt, J. Lin, Accelerate Synthesis of Metal–Organic Frameworks by a Robotic Platform and Bayesian Optimization, ACS Appl. Mater. Interfaces. 13 (2021) 53485–53491. https://doi.org/10.1021/acsami.1c16506.

[39] G. Agarwal, H.A. Doan, L.A. Robertson, L. Zhang, R.S. Assary, Discovery of Energy Storage Molecular Materials Using Quantum Chemistry-Guided Multiobjective Bayesian Optimization, Chem. Mater. 33 (2021) 8133–8144. https://doi.org/10.1021/acs.chemmater.1c02040.

[40] Z. Yang, S. Suzuki, N. Tanibata, H. Takeda, M. Nakayama, M. Karasuyama, I. Takeuchi, Efficient Experimental Search for Discovering a Fast Li-Ion Conductor from a Perovskite-Type Li x La (1– x)/3 NbO 3 (LLNO) Solid-State Electrolyte Using Bayesian Optimization, J. Phys. Chem. C. 125 (2021) 152–160. https://doi.org/10.1021/acs.jpcc.0c08887.

[41] P. Kim, MATLAB Deep Learning, Apress, Berkeley, CA, 2017. https://doi.org/10.1007/978-1-4842-2845-6.

[42] J.S. Niezgoda, M.A. Harrison, J.R. McBride, S.J. Rosenthal, Novel Synthesis of Chalcopyrite Cu x In y S 2 Quantum Dots with Tunable Localized Surface Plasmon Resonances, Chem. Mater. 24 (2012) 3294–3298. https://doi.org/10.1021/cm3021462.

[43] P. Linkov, P. Samokhvalov, K. Vokhmintsev, M. Zvaigzne, V.A. Krivenkov, I. Nabiev, Optical Properties of Quantum Dots with a Core–Multishell Structure, JETP Lett. 109 (2019) 112–115. https://doi.org/10.1134/S0021364019020103.

[44] K. Yu, S. Singh, N. Patrito, V. Chu, Effect of Reaction Media on the Growth and Photoluminescence of Colloidal CdSe Nanocrystals, Langmuir. 20 (2004) 11161–11168. https://doi.org/10.1021/la049202p.

[45] C. Wang, Y. Jiang, Z. Zhang, G. Li, L. Chen, J. Jie, Phosphine-Free Synthesis of CdSe Quantum Dots in a New Co-CappingLigand System, J. Nanosci. Nanotechnol. 9 (2009) 4735–4740. https://doi.org/10.1166/jnn.2009.1099.

[46] L. Protesescu, S. Yakunin, M.I. Bodnarchuk, F. Krieg, R. Caputo, C.H. Hendon, R.X. Yang, A. Walsh, M. V. Kovalenko, Nanocrystals of Cesium Lead Halide Perovskites ($CsPbX_3$, X = Cl, Br, and I): Novel Optoelectronic Materials Showing Bright Emission with Wide Color Gamut, Nano Lett. 15 (2015) 3692–3696. https://doi.org/10.1021/nl5048779.

[47] C.B. Murray, D.J. Norris, M.G. Bawendi, Synthesis and characterization of nearly monodisperse CdE (E = sulfur, selenium, tellurium) semiconductor nanocrystallites, J. Am. Chem. Soc. 115 (1993) 8706–8715. https://doi.org/10.1021/ja00072a025.

[48] P. Samokhvalov, M. Artemyev, I. Nabiev, Basic Principles and Current Trends in Colloidal Synthesis of Highly Luminescent Semiconductor Nanocrystals, Chem. - A Eur. J. 19 (2013) 1534–1546. https://doi.org/10.1002/chem.201202860.

[49] X. Peng, J. Wickham, A.P. Alivisatos, Kinetics of II-VI and III-V Colloidal Semiconductor Nanocrystal Growth: “Focusing” of Size Distributions, J. Am. Chem. Soc. 120 (1998) 5343–5344. https://doi.org/10.1021/ja9805425.

[50] J. van Embden, A.S.R. Chesman, J.J. Jasieniak, The Heat-Up Synthesis of Colloidal Nanocrystals, Chem. Mater. 27 (2015) 2246–2285. https://doi.org/10.1021/cm5028964.

[51] Y.A. Yang, H. Wu, K.R. Williams, Y.C. Cao, Synthesis of CdSe and CdTe Nanocrystals without Precursor Injection, Angew. Chemie Int. Ed. 44 (2005) 6712–6715. https://doi.org/10.1002/anie.200502279.

[52] Y. Liu, Y. Tang, Y. Ning, M. Li, H. Zhang, B. Yang, “One-pot” synthesis and shape control of ZnSe semiconductor nanocrystals in liquid paraffin, J. Mater. Chem. 20 (2010) 4451. https://doi.org/10.1039/c0jm00115e.

[53] A.P. Alivisatos, Perspectives on the Physical Chemistry of Semiconductor Nanocrystals, J. Phys. Chem. 100 (1996) 13226–13239. https://doi.org/10.1021/jp9535506.

[54] E.A. Slejko, V. Lughi, Size Control at Maximum Yield and Growth Kinetics of Colloidal II–VI Semiconductor Nanocrystals, J. Phys. Chem. C. 123 (2019) 1421–1428. https://doi.org/10.1021/acs.jpcc.8b07754.

[55] L. Qu, X. Peng, Control of Photoluminescence Properties of CdSe Nanocrystals in Growth, J. Am. Chem. Soc. 124 (2002) 2049–2055. https://doi.org/10.1021/ja017002j.

[56] D. V. Talapin, J.-S. Lee, M. V. Kovalenko, E. V. Shevchenko, Prospects of Colloidal Nanocrystals for Electronic and Optoelectronic Applications, Chem. Rev. 110 (2010) 389–458. https://doi.org/10.1021/cr900137k.

[57] S. Abe, R.K. Čapek, B. De Geyter, Z. Hens, Tuning the Postfocused Size of Colloidal Nanocrystals by the Reaction Rate: From Theory to Application, ACS Nano. 6 (2012) 42–53. https://doi.org/10.1021/nn204008q.

[58] E.M. Chan, C. Xu, A.W. Mao, G. Han, J.S. Owen, B.E. Cohen, D.J. Milliron, Reproducible, High-Throughput Synthesis of Colloidal Nanocrystals for Optimization in Multidimensional Parameter Space, Nano Lett. 10 (2010) 1874–1885. https://doi.org/10.1021/nl100669s.

[59] C.B. Murray, D.J. Norris, M.G. Bawendi, Synthesis and characterization of nearly monodisperse CdE (E = sulfur, selenium, tellurium) semiconductor nanocrystallites, J. Am. Chem. Soc. 115 (1993) 8706–8715. https://doi.org/10.1021/ja00072a025.

[60] S. Abe, R.K. Capek, B. De Geyter, Z. Hens, Reaction Chemistry/Nanocrystal Property Relations in the Hot Injection Synthesis, the Role of the Solute Solubility, ACS Nano. 7 (2013) 943–949. https://doi.org/10.1021/nn3059168.

[61] J. Joo, J.M. Pietryga, J.A. McGuire, S.-H. Jeon, D.J. Williams, H.-L. Wang, V.I. Klimov, A Reduction Pathway in the Synthesis of PbSe Nanocrystal Quantum Dots, J. Am. Chem. Soc. 131 (2009) 10620–10628. https://doi.org/10.1021/ja903445f.

[62] J.J. Urban, D. V. Talapin, E. V. Shevchenko, C.B. Murray, Self-Assembly of PbTe Quantum Dots into Nanocrystal Superlattices and Glassy Films, J. Am. Chem. Soc. 128 (2006) 3248–3255. https://doi.org/10.1021/ja058269b.

[63] X. Huang, V.K. Parashar, M.A.M. Gijs, Nucleation and Growth Behavior of CdSe Nanocrystals Synthesized in the Presence of Oleylamine Coordinating Ligand, Langmuir. 34 (2018) 6070–6076. https://doi.org/10.1021/acs.langmuir.7b01337.

[64] S. Kim, K.-D. Park, H. Lee, Growth Kinetics and Optical Properties of CsPbBr3 Perovskite Nanocrystals, Energies. 14 (2021) 275. https://doi.org/10.3390/en14020275.

[65] W. Zhihai, W. Jiao, S. Yanni, W. Jun, H. Yafei, W. Pan, W. Nengping, Z. Zhenfu, Air-stable all-inorganic perovskite quantum dot inks for multicolor patterns and white LEDs, J. Mater. Sci. 54 (2019) 6917–6929. https://doi.org/10.1007/s10853-019-03382-2.

[66] Y. Wu, T. Wei, X. An, L.-M. Liu, Colloidal synthesis of SnS nanocrystals with dimension-dependent photoelectrochemical properties, New J. Chem. 43 (2019) 7457–7462. https://doi.org/10.1039/C9NJ00506D.

[67] P. Cheng, L. Sun, L. Feng, S. Yang, Y. Yang, D. Zheng, Y. Zhao, Y. Sang, R. Zhang, D. Wei, W. Deng, K. Han, Colloidal Synthesis and Optical Properties of All- Inorganic Low-Dimensional Cesium Copper Halide Nanocrystals, Angew. Chemie Int. Ed. 58 (2019) 16087–16091. https://doi.org/10.1002/anie.201909129.

[68] X. Peng, Mechanisms for the Shape-Control and Shape-Evolution of Colloidal Semiconductor Nanocrystals, Adv. Mater. 15 (2003) 459–463. https://doi.org/10.1002/adma.200390107.

[69] A. Mansouri, N. Semagina, Colloidal Synthesis Protocol of Shape- and Dimensionally-Controlled Transition-Metal Chalcogenides and Their Hydrodesulfurization Activities, ACS Appl. Nano Mater. 1 (2018) 4408–4412. https://doi.org/10.1021/acsanm.8b01353.

[70] W. Jung, S. Lee, D. Yoo, S. Jeong, P. Miró, A. Kuc, T. Heine, J. Cheon, Colloidal Synthesis of Single-Layer MSe 2 (M = Mo, W) Nanosheets via Anisotropic Solution-Phase Growth Approach, J. Am. Chem. Soc. 137 (2015) 7266–7269. https://doi.org/10.1021/jacs.5b02772.

[71] J. Zhang, K. Sun, A. Kumbhar, J. Fang, Shape-Control of ZnTe Nanocrystal Growth in Organic Solution, J. Phys. Chem. C. 112 (2008) 5454–5458. https://doi.org/10.1021/jp711778u.

[72] W.W. Yu, Y.A. Wang, X. Peng, Formation and Stability of Size-, Shape-, and Structure-Controlled CdTe Nanocrystals: Ligand Effects on Monomers and Nanocrystals, Chem. Mater. 15 (2003) 4300–4308. https://doi.org/10.1021/cm034729t.

[73] Y. Xia, K.D. Gilroy, H.-C. Peng, X. Xia, Seed-Mediated Growth of Colloidal Metal Nanocrystals, Angew. Chemie Int. Ed. 56 (2017) 60–95. https://doi.org/10.1002/anie.201604731.

[74] Y. Feng, Y. Wang, J. He, X. Song, Y.Y. Tay, H.H. Hng, X.Y. Ling, H. Chen, Achieving Site-Specificity in Multistep Colloidal Synthesis, J. Am. Chem. Soc. 137 (2015) 7624–7627. https://doi.org/10.1021/jacs.5b04310.

[75] P. Cheng, L. Sun, L. Feng, S. Yang, Y. Yang, D. Zheng, Y. Zhao, Y. Sang, R. Zhang, D. Wei, W. Deng, K. Han, Colloidal Synthesis and Optical Properties of All- Inorganic Low-Dimensional Cesium Copper Halide Nanocrystals, Angew. Chemie Int. Ed. 58 (2019) 16087–16091. https://doi.org/10.1002/anie.201909129.

[76] Y. Yin, A.P. Alivisatos, Colloidal nanocrystal synthesis and the organic–inorganic interface, Nature. 437 (2005) 664–670. https://doi.org/10.1038/nature04165.

[77] J. Zhang, K. Sun, A. Kumbhar, J. Fang, Shape-Control of ZnTe Nanocrystal Growth in Organic Solution, J. Phys. Chem. C. 112 (2008) 5454–5458. https://doi.org/10.1021/jp711778u.

[78] A.G. Kanaras, C. Sönnichsen, H. Liu, A.P. Alivisatos, Controlled Synthesis of Hyperbranched Inorganic Nanocrystals with Rich Three-Dimensional Structures, Nano Lett. 5 (2005) 2164–2167. https://doi.org/10.1021/nl0518728.

[79] X. Peng, L. Manna, W. Yang, J. Wickham, E. Scher, A. Kadavanich, A.P. Alivisatos, Shape control of CdSe nanocrystals, Nature. 404 (2000) 59–61. https://doi.org/10.1038/35003535.

[80] J.Y. Rempel, B.L. Trout, M.G. Bawendi, K.F. Jensen, Density Functional Theory Study of Ligand Binding on CdSe (0001), (000$\bar{1}$), and (11$\bar{2}$0) Single Crystal Relaxed and Reconstructed Surfaces: Implications for Nanocrystalline Growth, J. Phys. Chem. B. 110 (2006) 18007–18016. https://doi.org/10.1021/jp064051f.

[81] L. Ruan, W. Shen, A. Wang, A. Xiang, Z. Deng, Alkyl-Thiol Ligand-Induced Shape- and Crystalline Phase-Controlled Synthesis of Stable Perovskite-Related $CsPb_2Br_5$ Nanocrystals at Room Temperature, J. Phys. Chem. Lett. 8 (2017) 3853–3860. https://doi.org/10.1021/acs.jpclett.7b01657.

[82] Scopus. https://scopus.com (accessed 29 July 2023).

[83] E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder, E. Olivetti, Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning, Chem. Mater. 29 (2017) 9436–9444. https://doi.org/10.1021/acs.chemmater.7b03500.

[84] M.C. Swain, J.M. Cole, ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature, J. Chem. Inf. Model. 56 (2016) 1894–1904. https://doi.org/10.1021/acs.jcim.6b00207.

[85] W.A. Saidi, W. Shadid, I.E. Castelli, Machine-learning structural and electronic properties of metal halide perovskites using a hierarchical convolutional neural network, Npj Comput. Mater. 6 (2020) 36. https://doi.org/10.1038/s41524-020-0307-8.

[86] L. Zhang, M. He, S. Shao, Machine learning for halide perovskite materials, Nano Energy. 78 (2020) 105380. https://doi.org/10.1016/j.nanoen.2020.105380.

[87] K. Higgins, S.M. Valleti, M. Ziatdinov, S. V. Kalinin, M. Ahmadi, Chemical Robotics Enabled Exploration of Stability in Multicomponent Lead Halide Perovskites via Machine Learning, ACS Energy Lett. 5 (2020) 3426–3436. https://doi.org/10.1021/acsenergylett.0c01749.

[88] R.W. Epps, M.S. Bowen, A.A. Volk, K. Abdel- Latif, S. Han, K.G. Reyes, A. Amassian, M. Abolhasani, Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot, Adv. Mater. 32 (2020) 2001626. https://doi.org/10.1002/adma.202001626.

[89] I.M. Pendleton, M.K. Caucci, M. Tynes, A. Dharna, M.A.N. Nellikkal, Z. Li, E.M. Chan, A.J. Norquist, J. Schrier, Can Machines "Learn" Halide Perovskite Crystal Formation without Accurate Physicochemical Features?, J. Phys. Chem. C. 124 (2020) 13982–13992. https://doi.org/10.1021/acs.jpcc.0c01726.

[90] L. Bezinge, R.M. Maceiczyk, I. Lignos, M. V. Kovalenko, A.J. DeMello, Pick a Color MARIA: Adaptive Sampling Enables the Rapid Identification of Complex Perovskite Nanocrystal Compositions with Defined Emission Characteristics, ACS Appl. Mater. Interfaces. 10 (2018) 18869–18878. https://doi.org/10.1021/acsami.8b03381.

[91] O. Voznyy, L. Levina, J.Z. Fan, M. Askerka, A. Jain, M.-J. Choi, O. Ouellette, P. Todorović, L.K. Sagar, E.H. Sargent, Machine Learning Accelerates Discovery of Optimal Colloidal Quantum Dot Synthesis, ACS Nano. 13 (2019) 11122–11128. https://doi.org/10.1021/acsnano.9b03864.

[92] F. Baum, T. Pretto, A. Köche, M.J.L. Santos, Machine Learning Tools to Predict Hot Injection Syntheses Outcomes for II–VI and IV–VI Quantum Dots, J. Phys. Chem. C. 124 (2020) 24298–24305. https://doi.org/10.1021/acs.jpcc.0c05993.

[93] A.Y. Fong, L. Pellouchoud, M. Davidson, R.C. Walroth, C. Church, E. Tcareva, L. Wu, K. Peterson, B. Meredig, C.J. Tassone, Utilization of machine learning to accelerate colloidal synthesis and discovery, J. Chem. Phys. 154 (2021). https://doi.org/10.1063/5.0047385.

[94] K. Abdel-Latif, R.W. Epps, F. Bateni, S. Han, K.G. Reyes, M. Abolhasani, Self- Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow, Adv. Intell. Syst. 3 (2021) 2000245. https://doi.org/10.1002/aisy.202000245.

[95] R. Liu, J. Hao, J. Li, S. Wang, H. Liu, Z. Zhou, M.-H. Delville, J. Cheng, K. Wang, X. Zhu, Causal Inference Machine Learning Leads Original Experimental Discovery in CdSe/CdS Core/Shell Nanoparticles, J. Phys. Chem. Lett. 11 (2020) 7232–7238. https://doi.org/10.1021/acs.jpclett.0c02115.

[96] E.J. Braham, J. Cho, K.M. Forlano, D.F. Watson, R. Arròyave, S. Banerjee, Machine Learning-Directed Navigation of Synthetic Design Space: A Statistical Learning Approach to Controlling the Synthesis of Perovskite Halide Nanoplatelets in the Quantum-Confined Regime, Chem. Mater. 31 (2019) 3281–3292. https://doi.org/10.1021/acs.chemmater.9b00212.

[97] L.N. Quan, B.P. Rand, R.H. Friend, S.G. Mhaisalkar, T.-W. Lee, E.H. Sargent, Perovskites for Next-Generation Optical Sources, Chem. Rev. 119 (2019) 7444–7477. https://doi.org/10.1021/acs.chemrev.9b00107.

[98] W.A. Saidi, W. Shadid, I.E. Castelli, Machine-learning structural and electronic properties of metal halide perovskites using a hierarchical convolutional neural network, Npj Comput. Mater. 6 (2020) 36. https://doi.org/10.1038/s41524-020-0307-8.

[99] V. Gladkikh, D.Y. Kim, A. Hajibabaei, A. Jana, C.W. Myung, K.S. Kim, Machine Learning for Predicting the Band Gaps of ABX_3 Perovskites from Elemental Properties, J. Phys. Chem. C. 124 (2020) 8905–8918. https://doi.org/10.1021/acs.jpcc.9b11768.

[100] Z. Yang, Y. Liu, Y. Zhang, L. Wang, C. Lin, Y. Lv, Y. Ma, C. Shao, Machine Learning Accelerates the Discovery of Light-Absorbing Materials for Double Perovskite Solar Cells, J. Phys. Chem. C. 125 (2021) 22483–22492. https://doi.org/10.1021/acs.jpcc.1c07262.

[101] T. Wu, J. Wang, Deep Mining Stable and Nontoxic Hybrid Organic–Inorganic Perovskites for Photovoltaics via Progressive Machine Learning, ACS Appl. Mater. Interfaces. 12 (2020) 57821–57831. https://doi.org/10.1021/acsami.0c10371.

[102] K. Higgins, M. Ziatdinov, S. V. Kalinin, M. Ahmadi, High-Throughput Study of Antisolvents on the Stability of Multicomponent Metal Halide Perovskites through Robotics-Based Synthesis and Machine Learning Approaches, J. Am. Chem. Soc. 143 (2021) 19945–19955. https://doi.org/10.1021/jacs.1c10045.

[103] K. Higgins, S.M. Valleti, M. Ziatdinov, S. V. Kalinin, M. Ahmadi, Chemical Robotics Enabled Exploration of Stability in Multicomponent Lead Halide Perovskites via Machine Learning, ACS Energy Lett. 5 (2020) 3426–3436. https://doi.org/10.1021/acsenergylett.0c01749.

[104] Y. Yu, X. Tan, S. Ning, Y. Wu, Machine Learning for Understanding Compatibility of Organic–Inorganic Hybrid Perovskites with Post-Treatment Amines, ACS Energy Lett. 4 (2019) 397–404. https://doi.org/10.1021/acsenergylett.8b02451.

[105] H.C. Herbol, W. Hu, P. Frazier, P. Clancy, M. Poloczek, Efficient search of compositional space for hybrid organic–inorganic perovskites via Bayesian optimization, Npj Comput. Mater. 4 (2018) 51. https://doi.org/10.1038/s41524-018-0106-7.

[106] R. Lyu, C.E. Moore, T. Liu, Y. Yu, Y. Wu, Predictive Design Model for Low-Dimensional Organic–Inorganic Halide Perovskites Assisted by Machine Learning, J. Am. Chem. Soc. 143 (2021) 12766–12776. https://doi.org/10.1021/jacs.1c05441.

[107] R.W. Epps, A.A. Volk, K.G. Reyes, M. Abolhasani, Accelerated AI development for autonomous materials synthesis in flow, Chem. Sci. 12 (2021) 6025–6036. https://doi.org/10.1039/D0SC06463G.

[108] K. Abdel-Latif, R.W. Epps, F. Bateni, S. Han, K.G. Reyes, M. Abolhasani, Self- Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow, Adv. Intell. Syst. 3 (2021) 2000245. https://doi.org/10.1002/aisy.202000245.

[109] F. Baum, T. Pretto, A. Köche, M.J.L. Santos, Machine Learning Tools to Predict Hot Injection Syntheses Outcomes for II–VI and IV–VI Quantum Dots, J. Phys. Chem. C. 124 (2020) 24298–24305. https://doi.org/10.1021/acs.jpcc.0c05993.

[110] Y. Xia, X. Xia, Y. Wang, S. Xie, Shape-controlled synthesis of metal nanocrystals, MRS Bull. 38 (2013) 335–344. https://doi.org/10.1557/mrs.2013.84.

[111] J. Li, T. Chen, K. Lim, L. Chen, S.A. Khan, J. Xie, X. Wang, Deep Learning Accelerated Gold Nanocluster Synthesis, Adv. Intell. Syst. 1 (2019) 1900029. https://doi.org/10.1002/aisy.201900029.

[112] G. Koch, R. Zemel, R. Salakhutdinov, Siamese Neural Networks for One-shot Image Recognition, Proceedings of the 32nd International Conference on Machine Learning, Lille, France, JMLR: W&CP, 37 (2015).

[113] Z. Zhou, X. Li, R.N. Zare, Optimizing Chemical Reactions with Deep Reinforcement Learning, ACS Cent. Sci. 3 (2017) 1337–1344. https://doi.org/10.1021/acscentsci.7b00492.

[114] Q. Du, Q. Zhang, G. Liu, Deep learning: an efficient method for plasmonic design of geometric nanoparticles, Nanotechnology. 32 (2021) 505607. https://doi.org/10.1088/1361-6528/ac2769.

[115] F. Mekki-Berrada, Z. Ren, T. Huang, W.K. Wong, F. Zheng, J. Xie, I.P.S. Tian, S. Jayavelu, Z. Mahfoud, D. Bash, K. Hippalgaonkar, S. Khan, T. Buonassisi, Q. Li, X. Wang, Two-step machine learning enables optimized nanoparticle synthesis, Npj Comput. Mater. 7 (2021) 55. https://doi.org/10.1038/s41524-021-00520-w.

[116] A. Ghosh, S. Datta, T. Saha-Dasgupta, Understanding the Trend in Core–Shell Preferences for Bimetallic Nanoclusters: A Machine Learning Approach, J. Phys. Chem. C. 126 (2022) 6847–6853. https://doi.org/10.1021/acs.jpcc.2c01096.

[117] J.A. Darr, J. Zhang, N.M. Makwana, X. Weng, Continuous Hydrothermal Synthesis of Inorganic Nanoparticles: Applications and Future Directions, Chem. Rev. 117 (2017) 11125–11238. https://doi.org/10.1021/acs.chemrev.6b00417.

[118] D.Y. Nadargi, M.S. Tamboli, S.S. Patil, I.S. Mulla, S.S. Suryavanshi, Development of Ag/ZnO nanorods and nanoplates at low hydrothermal temperature and time for acetone sensing

application: an insight into spillover mechanism, SN Appl. Sci. 1 (2019) 1564. https://doi.org/10.1007/s42452-019-1573-2.

[119] Z. Bao, Y. Yuan, C. Leng, L. Li, K. Zhao, Z. Sun, One-Pot Synthesis of Noble Metal/Zinc Oxide Composites with Controllable Morphology and High Catalytic Performance, ACS Appl. Mater. Interfaces. 9 (2017) 16417–16425. https://doi.org/10.1021/acsami.7b02667.

[120] S.S. Sutar, S.M. Patil, S.J. Kadam, R.K. Kamat, D. Kim, T.D. Dongale, Analysis and Prediction of Hydrothermally Synthesized ZnO-Based Dye-Sensitized Solar Cell Properties Using Statistical and Machine-Learning Techniques, ACS Omega. 6 (2021) 29982–29992. https://doi.org/10.1021/acsomega.1c04521.

[121] E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder, E. Olivetti, Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning, Chem. Mater. 29 (2017) 9436–9444. https://doi.org/10.1021/acs.chemmater.7b03500.

[122] Y. Han, B. Tang, L. Wang, H. Bao, Y. Lu, C. Guan, L. Zhang, M. Le, Z. Liu, M. Wu, Machine-Learning-Driven Synthesis of Carbon Dots with Enhanced Quantum Yields, ACS Nano. 14 (2020) 14761–14768. https://doi.org/10.1021/acsnano.0c01899.

[123] Q. Hong, X.-Y. Wang, Y.-T. Gao, J. Lv, B.-B. Chen, D.-W. Li, R.-C. Qian, Customized Carbon Dots with Predictable Optical Properties Synthesized at Room Temperature Guided by Machine Learning, Chem. Mater. 34 (2022) 998–1009. https://doi.org/10.1021/acs.chemmater.1c03220.

[124] D.M. Pashkov, A.A. Guda, M. V. Kirichkov, S.A. Guda, A. Martini, S.A. Soldatov, A. V. Soldatov, Quantitative Analysis of the UV–Vis Spectra for Gold Nanoparticles Powered by Supervised Machine Learning, J. Phys. Chem. C. 125 (2021) 8656–8666. https://doi.org/10.1021/acs.jpcc.0c10680.

[125] N. Đorđević, J.S. Beckwith, M. Yarema, O. Yarema, A. Rosspeintner, N. Yazdani, J. Leuthold, E. Vauthey, V. Wood, Machine Learning for Analysis of Time-Resolved Luminescence Data, ACS Photonics. 5 (2018) 4888–4895. https://doi.org/10.1021/acsphotonics.8b01047.

[126] J.M. Howard, Q. Wang, M. Srivastava, T. Gong, E. Lee, A. Abate, M.S. Leite, Quantitative Predictions of Moisture-Driven Photoemission Dynamics in Metal Halide Perovskites via Machine Learning, J. Phys. Chem. Lett. 13 (2022) 2254–2263. https://doi.org/10.1021/acs.jpclett.2c00131.

[127] B. Yildirim, J.M. Cole, Bayesian Particle Instance Segmentation for Electron Microscopy Image Quantification, J. Chem. Inf. Model. 61 (2021) 1136–1149. https://doi.org/10.1021/acs.jcim.0c01455.

[128] K.T. Mukaddem, E.J. Beard, B. Yildirim, J.M. Cole, ImageDataExtractor: A Tool To Extract and Quantify Data from Microscopy Images, J. Chem. Inf. Model. 60 (2020) 2492–2509. https://doi.org/10.1021/acs.jcim.9b00734.

[129] M.D. Hannel, A. Abdulali, M. O'Brien, D.G. Grier, Machine-learning techniques for fast and accurate feature localization in holograms of colloidal particles, Opt. Express. 26 (2018) 15221. https://doi.org/10.1364/OE.26.015221.

[130] S. Sun, N.T.P. Hartono, Z.D. Ren, F. Oviedo, A.M. Buscemi, M. Layurova, D.X. Chen, T. Ogunfunmi, J. Thapa, S. Ramasamy, C. Settens, B.L. DeCost, A.G. Kusne, Z. Liu, S.I.P. Tian, I.M. Peters, J.-P. Correa-Baena, T. Buonassisi, Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis, Joule. 3 (2019) 1437–1451. https://doi.org/10.1016/j.joule.2019.05.014.

[131] Y. Zhuo, A. Mansouri Tehrani, J. Brgoch, Predicting the Band Gaps of Inorganic Solids by Machine Learning, J. Phys. Chem. Lett. 9 (2018) 1668–1673. https://doi.org/10.1021/acs.jpclett.8b00124.

[132] X. Yan, J. Zhang, D.P. Russo, H. Zhu, B. Yan, Prediction of Nano–Bio Interactions through Convolutional Neural Network Analysis of Nanostructure Images, ACS Sustain. Chem. Eng. 8 (2020) 19096–19104. https://doi.org/10.1021/acssuschemeng.0c07453.

[133] X.-Y. Wang, B.-B. Chen, J. Zhang, Z.-R. Zhou, J. Lv, X.-P. Geng, R.-C. Qian, Exploiting deep learning for predictable carbon dot design, Chem. Commun. 57 (2021) 532–535. https://doi.org/10.1039/D0CC07882D.

[134] B. Sun, M. Fernandez, A.S. Barnard, Machine Learning for Silver Nanoparticle Electron Transfer Property Prediction, J. Chem. Inf. Model. 57 (2017) 2413–2423. https://doi.org/10.1021/acs.jcim.7b00272.

[135] B. Liang, D. Xu, N. Yu, Y. Xu, X. Ma, Q. Liu, M.S. Asif, R. Yan, M. Liu, Physics-Guided Neural-Network-Based Inverse Design of a Photonic – Plasmonic Nanodevice for Superfocusing, ACS Appl. Mater. Interfaces. 14 (2022) 27397–27404. https://doi.org/10.1021/acsami.2c05083.

[136] A. Lininger, M. Hinczewski, G. Strangi, General Inverse Design of Layered Thin-Film Materials with Convolutional Neural Networks, ACS Photonics. 8 (2021) 3641–3650. https://doi.org/10.1021/acsphotonics.1c01498.

[137] R. Unni, K. Yao, Y. Zheng, Deep Convolutional Mixture Density Network for Inverse Design of Layered Photonic Structures, ACS Photonics. 7 (2020) 2703–2712. https://doi.org/10.1021/acsphotonics.0c00630.

[138] D. Mengu, M.S. Sakib Rahman, Y. Luo, J. Li, O. Kulce, A. Ozcan, At the intersection of optics and deep learning: statistical inference, computing, and inverse design, Adv. Opt. Photonics. 14 (2022) 209. https://doi.org/10.1364/AOP.450345.

[139] Z.S. Ballard, D. Shir, A. Bhardwaj, S. Bazargan, S. Sathianathan, A. Ozcan, Computational Sensing Using Low-Cost and Mobile Plasmonic Readers Designed by Machine Learning, ACS Nano. 11 (2017) 2266–2274. https://doi.org/10.1021/acsnano.7b00105.

[140] C. Chen, X. Lin, X. Wu, H. Bao, L. Wu, X. Hu, Y. Zhang, D. Yang, W. Hou, W. Cao, H. Zhong, Machine Learning Assisted Stability Analysis of Blue Quantum Dot Light-Emitting Diodes, Nano Lett. 23 (2023) 5738–5745. https://doi.org/10.1021/acs.nanolett.3c01491.

Declaration of interests

☒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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: