ARTICLE   OPEN

Two-step machine learning enables optimized nanoparticle synthesis

Flore Mekki-Berrada[1,9], Zekun Ren[2,9], Tan Huang[1,9], Wai Kuan Wong[1], Fang Zheng[1], Jiaxun Xie[1], Isaac Parker Siyu Tian[2], Senthilnath Jayavelu[3], Zackaria Mahfoud[4], Daniil Bash[4], Kedar Hippalgaonkar[4,5], Saif Khan[1], Tonio Buonassisi[6], Qianxiao Li[7,8] and Xiaonan Wang[1✉]

In materials science, the discovery of recipes that yield nanomaterials with defined optical properties is costly and time-consuming. In this study, we present a two-step framework for a machine learning-driven high-throughput microfluidic platform to rapidly produce silver nanoparticles with the desired absorbance spectrum. Combining a Gaussian process-based Bayesian optimization (BO) with a deep neural network (DNN), the algorithmic framework is able to converge towards the target spectrum after sampling 120 conditions. Once the dataset is large enough to train the DNN with sufficient accuracy in the region of the target spectrum, the DNN is used to predict the colour palette accessible with the reaction synthesis. While remaining interpretable by humans, the proposed framework efficiently optimizes the nanomaterial synthesis and can extract fundamental knowledge of the relationship between chemical composition and optical properties, such as the role of each reactant on the shape and amplitude of the absorbance spectrum.

INTRODUCTION

In recent years, machine learning (ML) methods have been applied to solve various problems in materials science, such as drug discovery[1,2], medical imaging[3], material synthesis[4,5], functional molecules generation[6,7], and materials degradation[8]. Since the generation of experimental data in materials science is costly and time-consuming, ML algorithms have been mainly developed based on computational data or, when available, experimental datasets gathered from the literature[9]. However, once material is suggested by the ML algorithm, the material synthesis can turn out to be difficult, or even impossible. The recent development of microfluidic high-throughput experimental (HTE) platforms now allows the generation of a large amount of experimental synthesis data with small amounts of material[10–13]. The integration of the ML algorithms in a loop with these flow chemistry platforms would ensure that ML algorithms suggest only those materials that can be synthesized. Such attempts have been made in nanomaterial synthesis[14,15]. However, these studies are limited to optimization problems and focus on sparse datasets, while large datasets would be needed to extract knowledge on how the chemical composition and process parameters influence the final outcome[16].

In this paper, we propose a two-step ML framework that can drive an HTE platform from the very start of the screening process (sparse dataset) to more resolved screening states (large dataset), to target predetermined optical properties, and without any a priori knowledge of the model complexity, extract knowledge on how the chemical process impacts the optical properties of the synthesized material. Wet chemical nanoparticle synthesis is notoriously challenging to tune because of the intrinsic nonlinear competition between nucleation of seed particles and growth of pre-existing seeds in the solution[17,18]. Therefore, silver nanoparticle (AgNP) synthesis was chosen to demonstrate the efficiency of the framework. The AgNP synthesis is carried out using a droplet-based microfluidic platform with five input variables, as shown in Fig. 1 and detailed in "Methods". Due to surface plasmon resonances, AgNPs have a characteristic optical fingerprint in the UV–Vis range that depends on their size and shape distributions. In this study, we select as the optical target, the theoretical absorbance spectrum of triangular nanoprisms with 50 nm long edges and 10 nm in height, calculated by plasmon resonance simulation using discrete dipole scattering (DDSCAT).

Conventional Bayesian optimization (BO) is often chosen for driving HTE loops because of its ability to efficiently explore the parameter space and target specific material properties, even when initiated with a sparse dataset[5,19]. However, BO does not give general insights into the reaction process. Moreover, its performance depends on the initial choice of model hyperparameters and on the definition of the loss—usually a single-valued output parameter into which the measured material properties are reduced. To extract knowledge from the data, other studies use neural networks to train a regression model and perform inverse design from a fixed dataset[16,20,21]. While a neural network can learn complex functions even from a full optical spectrum, it has many hyperparameters and requires a large training dataset, which makes it difficult to integrate into a machine-driven experimental loop with limited initial data and expensive evaluations, and is inefficient to use at the early stage of sampling to explore the parameter space.

The proposed two-step framework (Fig. 1) combines the optimization assets of BO with the regression ability of a deep neural network (DNN). In a first step, after performing a first

---

[1]Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore, Singapore. [2]Singapore-MIT Alliance for Research and Technology SMART, Singapore, Singapore. [3]Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [4]Institute of Materials Research & Engineering, Singapore, Singapore. [5]Department of Materials Science and Engineering, Nanyang Technological University, Singapore, Singapore. [6]Massachusetts Institute of Technology, Cambridge, MA, USA. [7]Department of Mathematics, National University of Singapore, Singapore, Singapore. [8]Institute of High Performance Computing, Singapore, Singapore. [9]These authors contributed equally: Flore Mekki-Berrada, Zekun Ren, Tan Huang. ✉email: chewxia@nus.edu.sg
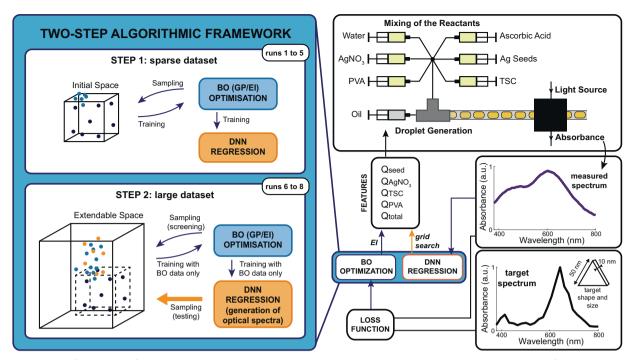
Fig. 1 Algorithmic framework for a high-throughput experimental loop. The two-step optimization algorithmic framework (blue box) consists of a first HTE loop (runs 2–5) in which the BO (blue points) is sampling the parameter space to train a DNN and a second loop (runs 6–8) in which the DNN (orange points) is allowed to sample the parameter space to validate its regression function. The conditions suggested by the BO and the DNN are tested on a droplet-based microfluidic platform. The absorbance spectrum of each droplet is measured and compared to the target spectrum through the loss function before feeding the BO, while the fully resolved absorbance spectrum is provided to the DNN.

experimental run of 15 conditions using Latin HyperCube (LH) sampling, the optimization process is initiated using a batch mode BO with local penalization (LP)[22,23]. The BO algorithm with Gaussian process (GP) as a surrogate model (see "Methods") is used to explore the parameter space, where boundaries are initially set by the experimenter, and find the chemical conditions that lead to the target spectrum. The definition of the loss function (see Eq. (7) in "Methods") takes into account both the shape and the intensity of the absorbance spectrum. At each run, the BO algorithm picks the next batch of 15 conditions to test, based on a balance between minimizing loss (exploitation) and minimizing uncertainty (exploration), as determined by the decision-making policy (acquisition function). In parallel, an offline DNN is trained using the experimental data generated by the BO sampling. In a second step, starting from the sixth run, while the BO continues suggesting 15 more conditions with the same hyperparameters and feeding the DNN with more data around the targeted spectrum, the DNN is used to produce the simulated spectra for all the process variables on a parameter space grid. In this way, the DNN is able to suggest 15 more conditions that minimize the loss function by ranking the predicted values in the grid. The DNN architecture and grid optimization are described in "Methods". The conditions suggested by both DNN and BO are tested on the HTE platform. In the subsequent runs, only experimental data generated by the BO sampling are used to train the DNN, which allows a direct comparison between the BO and the DNN performance. The ML-driven HTE loop is stopped when the target spectrum has been optimized, either by the BO or the DNN and when the DNN regression is sufficiently accurate and stable to extract knowledge on the chemical synthesis. The detailed data flow of the framework can be found in Supplementary Fig. 1.

Herein, we first demonstrate that the proposed two-step algorithmic framework efficiently optimizes the nanomaterial synthesis to get the desired plasmon resonance. The optimization performance is validated experimentally by TEM imaging of the synthesized AgNPs. Next, by extracting both BO and DNN regression functions, we show how the optimization process remains interpretable by humans. Lastly, once the stability and accuracy of the DNN regression function are established, we use the DNN to extract fundamental knowledge on how the chemical composition and the spectral properties of the nanoparticles are related.

RESULTS AND DISCUSSION

Optimization performance

To evaluate the optimization performance of the framework, we follow the evolution of the loss over successive experimental runs. Each run consists of 15 chemical conditions. For each condition, the optical spectra of 20 droplet replicas are recorded and used to update the algorithms, and the median loss of the 20 replicas is calculated to handle the outliers. As the median loss is used to update the BO, we define the condition leading to the lowest median loss among all the conditions of a run as the best performer. In Fig. 2a, we report the loss value obtained for each replica, as well as the statistical distribution of the replicas for the best performing condition. In the first step of the framework, the median loss of the BO best performing condition quickly decreases in the first runs, before reaching a plateau starting from run 4.

The best conditions suggested by the BO accumulate at the border of the parameter space (lower Q_{seed}, higher Q_{AgNO_3}), suggesting that better configurations may be found beyond the parameter space boundaries. Active learning campaigns traditionally fix the boundaries of the parameter space, which requires prior expert knowledge on chemical synthesis. In our study, we face the reality of an experimental campaign when exploring a parameter space never explored before and the flexibility of a parameter space expansion is essential for getting closer to the
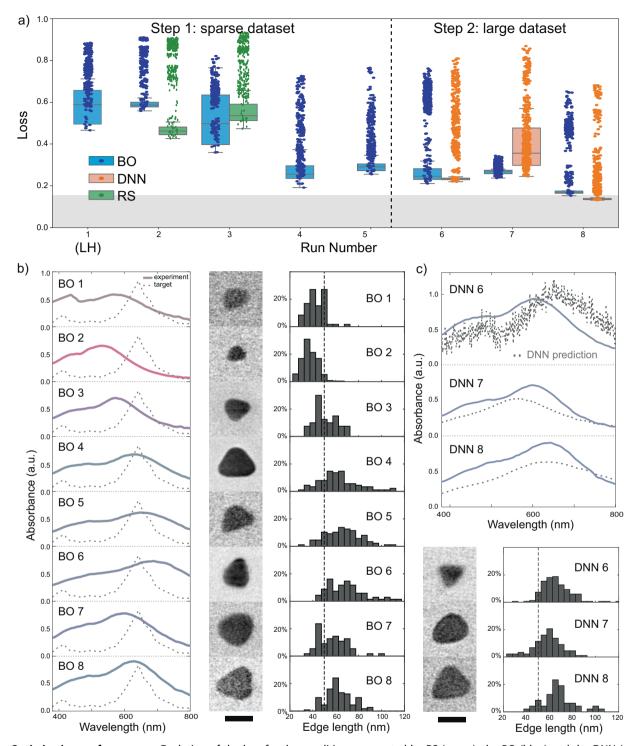
Fig. 2 Optimization performance. a Evolution of the loss for the conditions suggested by RS (green), the BO (blue) and the DNN (orange): each point represents a droplet. For each run, the condition giving the lowest loss is identified as the best performer. The box plots show the distribution of the loss values for all the droplet samples of the run, and the grey zone indicates the minimal loss value to reach before stopping the HTE loop. The absorbance spectra of the BO (**b**) and DNN (**c**) best performers, with the associated size distribution of triangular prisms in solution, show that the chosen loss function allows a convergence of the spectra towards the target spectrum and the prisms towards a triangle edge of 65 nm. The colours of the absorbance spectra indicate the colour of the best performing droplet. The TEM images show nanoprisms with the median edge size found in the sample. The scale bars correspond to 50 nm.

target spectrum. The framework is designed to be able to extend the parameter space if the suggested conditions are too close to the set boundaries for two successive runs. This situation occurs in runs 4 and 5. Thus, starting from run 6, the flow rate constraints are relaxed to the maximal values allowed by the equipment. No

preliminary screening of the extended parameter space is performed. Both BO and DNN-based grid optimization use their knowledge on the initial parameter space to start exploring the extended space. This extension allows a further decrease in the median loss of the best performing conditions obtained by BO

sampling. DNN sampling is introduced from run 6. Interestingly, the median loss of the best performing condition obtained by the DNN in run 8 is significantly lower than the one of BO (see Fig. 2a).

To justify our two-step approach, we looked at the optimization performance of the DNN when we let it sample the space by itself using grid search, instead of using BO sampled conditions. Starting from the same set of initial conditions at run 1, we show that the DNN does not manage to find lower loss conditions as quickly as the BO (see Supplementary Discussion 2 for more details). Furthermore, we demonstrate the non-triviality of the BO approach in the first step of the framework by comparing its performance to random sampling (RS) (see Fig. 2a): one can observe on run 2 that RS performs better than BO. But on run 3, BO outperforms RS and clearly starts converging towards lower loss values, while RS still picks high loss values.

The optimization is further validated by the convergence of the BO and DNN best-performing absorbance spectra towards the target spectrum. For the BO samples (Fig. 2b), the main absorbance peak quickly shifts to reach the target value (645 nm), while the intensity of the absorbance below 600 nm decreases. The evolution of the measured spectra towards the target spectrum validates the efficiency of the loss function definition used in this study. Figure 2c reports at each run the measured spectra of the DNN best performers, as well as the spectra predicted by the DNN before sampling. While the predicted spectra in run 6 are noisy due to insufficient training in a region that just became accessible, the spectral predictions tend to be smoother in the following runs, even if the DNN is still doing extrapolation at the location of the sampled conditions.

The best performing samples were imaged by TEM (see "TEM imaging and analysis"). Two nanoparticle shapes were mainly synthesized: nanospheres and triangular nanoprisms, with a wide heterogeneity in size for both shapes. The percentage of triangular shapes stays around 30% for the different runs. Furthermore, the triangular nanoprisms edge length (Fig. 2b, c) and the nanospheres diameter (see Supplementary Figure 3) both increase over the runs, for both BO and DNN best performances. The shift of the absorbance spectra with the runs goes along with a shift of the size distribution of the synthesized triangular nanoprisms towards the desired triangular edge length. The size distribution of the triangular AgNPs synthesized by the BO (Fig. 2b) and the DNN (Fig. 2c) gets narrower with the runs, the triangle edge converging towards 65 nm, which is 30% larger than the simulated nanoparticle size that produces the target spectrum. This shift can be explained by an increase in the thickness of the triangular prisms: using TEM measurements, the prism thickness was estimated around 13 nm. In fact, the absorbance peak position is determined by the aspect ratio between the prism edge length and its thickness[24].

It is worth noting that the approach developed in this study uses the full optical absorbance spectrum. In previous studies[14,15], the optimization process was performed using only certain attributes of the absorbance spectrum, such as peak wavelength, full width at half maximum (FWHM), and peak intensity. While using limited spectral features can be a good choice for fast optimization problems, it would reduce the spectral information used for training the DNN in the case of a regression problem. Since the shape and size heterogeneity of the AgNP leads to the superposition of the absorbance peaks, the whole spectrum contains information on the full size and shape distribution of the nanoparticles. While the 1D reduction of the spectrum into a single loss value allows the BO to remain efficient, the full spectral resolution can be used by the DNN to get both an efficient optimization and allow the DNN to accurately predict the AgNP colours. This full spectral approach has two additional advantages. First, it does not require adaptation of the DNN architecture to the number of features detected, in the event that the number of features evolves during the screening of the parameter space.

Second, it would make the framework robust to a change of spectral target during the optimization process, since the loss function does not need to be adapted to the absorbance target.

Interpretability

To understand how the BO proceeds to make decisions in each successive run, the Pearson correlation matrix is calculated. Supplementary Fig. 4 shows the corresponding correlation coefficient of the shape and amplitude of the spectra pertaining to the total loss. From run 1 to run 5, the shape has a higher correlation coefficient (-0.93), compared to that of amplitude (-0.55). Thus, the spectral shape is mainly optimized in the initial parameter space, relative to the spectral amplitude. However, from run 6 to run 8, the correlation coefficient for the shape (-0.63) becomes smaller than that for the amplitude (-0.95), showing that the amplitude of the absorbance spectrum is mainly optimized during the second step of the framework.

In the following, we investigate the reasons for the good optimization performance of the DNN during the second step, while the dataset remains sparse in the extended parameter space. In the second step of the framework, both algorithms are refining their extrapolation accuracy. As the DNN is only trained with the data obtained by the BO sampling, we can compare its surrogate function with the BO's surrogate function obtained with a GP for each run. Using SHAP (SHapley Additive exPlanations), we can rank the process variables according to their importance: Q_{AgNO_3} and Q_{seed} are identified as the most important, followed by Q_{TSC}, Q_{total} and Q_{PVA} (see Supplementary Fig. 5). The {Q_{AgNO_3}, Q_{seed}} space is thus chosen to project the minimum loss obtained by the regression function over the three other process variables. The minimum loss projection obtained at the end of run 8 is shown in Fig. 3 for three different functions: the raw experimental data fitted with a Gaussian distribution, the BO regression function and the DNN regression function. Both the BO and the DNN suggested conditions converge to a similar region in the {Q_{AgNO_3}, Q_{seed}} space (Fig. 3a). The position of the global minimum is similar for both algorithms. However, the BO regression function is found to have fewer features than the DNN one (Fig. 3b, c). We report in Supplementary Movie 1 the evolution of the BO regression function from run 1 to 8, with the position of the next suggested conditions at each run. We observe that the suggested conditions are not clustered, as expected since the jitter value was chosen for BO to perform in both exploration and exploitation mode. Furthermore, we observe a lack of experimental points between the global and secondary minima due to the sudden expansion of the parameter space. This could explain the reminiscence of the local minimum observed on the projection of the BO regression function in Fig. 3b. This second minimum is not observed on the DNN regression, confirming the better ability of the DNN to fit the parameter space.

To further understand why the BO was outperformed by the DNN, we examine the minimum loss projections of the BO surrogate and the DNN in the {Q_{TSC}, Q_{total}} space over the three other dimensions. Striped features appear in the BO projection, while the DNN performs correctly in the same subspace (see Supplementary Fig. 6). The BO is unable to properly fit the Q_{total} dimension, due to the ten times higher resolution in this dimension compared to the other ones since the parameters are unnormalized before the BO training. Parameter normalization for BO surrogate leads to a better projection in the {Q_{TSC}, Q_{total}} space but we choose not to normalize that to enable more flexibility in the event of a parameter space extension. While the non-normalization of the parameter highly affects the BO performance, the DNN performs well on the Q_{total} dimension.
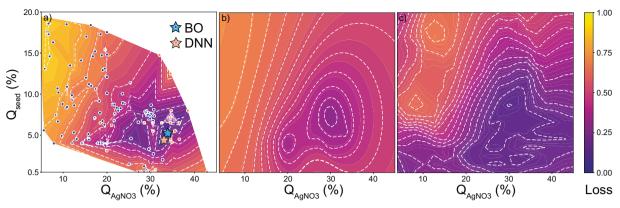
Fig. 3   Interpretability of the algorithmic decision process. 2D mapping of the minimum loss obtained with (**a**) the raw experimental data, (**b**) the BO, and (**c**) the DNN regression in the {Q_{AgNO_3}, Q_{seed}} space. Experimental conditions are represented by blue disks (BO suggested conditions from run 1 to 8) and orange disks (DNN suggested conditions from run 6 to 8). Stars represent the parameter conditions for the best loss performance obtained by the BO and the DNN. The colours correspond to the loss values as indicated on the colour bar.
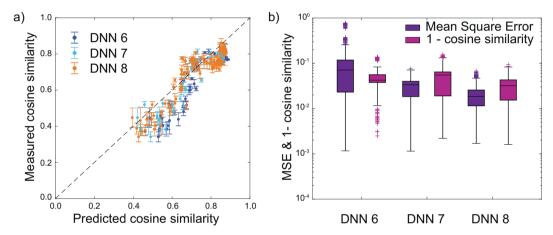
Fig. 4   DNN regression validation. a The measured cosine similarity (cosine similarity between the target spectrum and the experimental spectrum) is compared to the predicted cosine similarity (cosine similarity between the target spectrum and the DNN predicted spectrum at runs 6, 7 and 8). All the data that were not used for training the DNN are used for this validation step: all the DNN data, the BO data from the last run, and the random sampling data. The error bars represent the standard deviation of the measured cosine similarity on the droplet replicas of a condition. The alignment of the data with the diagonal shows that the DNN is able to predict the spectral shape correctly. **b** Evolution of the cosine similarity and the mean squared error (MSE) between the spectra predicted by the DNN and the spectra measured at each run, for the conditions suggested by both BO and DNN in that run. The box plots show the distribution of these values considering all the droplet samples of the run.

Knowledge extraction

The complexity of the relation between chemical composition and optical performance can be explored by performing a principal component analysis (PCA). It is found that neither a linear nor a kernel PCA can help in reducing the parameter space (see Supplementary Discussion 3). This indicates that there are complex nonlinear relationships between the chemical parameters and the optical spectrum. Some information though can be extracted from the SHAP analysis: we observe that high Q_{AgNO_3}, low Q_{seed}, low Q_{TSC} and high Q_{total} values have a negative correlation with the final loss. This gives information not only about the future directions for designing the experimental setup but also about the region where the target spectrum could be reached.

The correlation matrix can also help to understand how the flow rate ratios Q_{AgNO_3}/Q_{seed} and Q_{TSC}/Q_{AgNO_3} affect the spectral outcome (see Supplementary Fig. 4). While the ratio between silver nitrate and silver seed flowrates—and therefore concentrations in the droplets—has a greater impact on the spectral amplitude, the ratio between trisodium citrate and silver nitrate concentration has a greater influence on the shape of the absorbance spectrum. This extracted insight is non-trivial and in

agreement with the prior literature on the role of trisodium citrate on anisotropic growth in AgNPs synthesis[25].

To go further and determine which colour palette can be achieved with this chemical process and establish a map for the accessible colours, we use the trained DNN to generate spectra in the parameter space. Before extracting any information from the DNN, the accuracy and stability of the DNN regression should be quantified. While neural networks usually use a fixed dataset which is generally separated in two for training and validation purposes, the two-step framework integrates the DNN in the HTE loop, with a dataset that expands at each run. The DNN is trained online with the data previously sampled by the BO and the validation step is performed with the data selected by the grid optimization for the following experimental run. Thus, the accuracy of the DNN can be investigated by comparing the absorbance spectra predicted by the DNN to the spectra measured in the following run. One way to qualitatively represent the prediction accuracy of the DNN is to report the cosine similarity between the measured and the target spectrum as a function of the cosine similarity between the DNN-predicted and the target spectrum (Fig. 4a). The data points gather around the
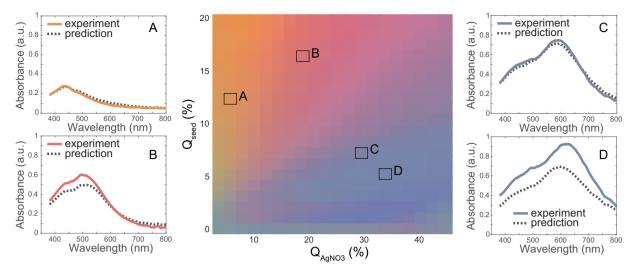
Fig. 5 Knowledge extraction on the silver nanoprism synthesis. Map of the AgNPs colours predicted by the DNN$_{teacher}$ in the $\{Q_{AgNO_3}, Q_{seed}\}$ space, for a fixed value $Q_{PVA} = 16\%$, $Q_{TSC} = 6.5\%$ and $Q_{total} = 850$ μL/min, corresponding to the conditions of the DNN best performance 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.

diagonal, meaning that the DNN predictions are as close to the target as the measured spectra. The accuracy can also be quantitatively estimated for each condition in two different ways: the cosine similarity between the predicted and measured spectra determines the accuracy of the shape of the absorbance spectrum, while the mean squared error (MSE) gives an estimation of the error in the absorbance amplitude at each wavelength. Figure 4b shows, from run 6 to run 8, a slight improvement of the cosine similarity and a more significant decrease in the mean MSE values. The DNN prediction becomes progressively more accurate in terms of spectral amplitude and in terms of spectral noise. In run 8, the MSE between predicted and measured spectra becomes lower than our target value, arbitrarily fixed to 0.02, and the HTE loop is stopped.

The stability of the DNN with respect to the variations among the runs can be investigated by tracking the evolution of the regression function in the $\{Q_{AgNO_3}, Q_{seed}\}$ space from run 1 to run 8 (Supplementary Fig. 7). Whereas the BO projection changes gradually even when the parameter space is expanded (run 5), the DNN mapping changes drastically for the first 5 runs. This shows that DNN is not stable with a small training dataset. The stability of the DNN regression function is evaluated by measuring the cosine similarity between two successive runs of DNN on the $\{Q_{AgNO_3}, Q_{seed}\}$ plane of the parameter space, while the Q_{TSC}, Q_{total} and Q_{PVA} values of the DNN are fixed to the best performance conditions in run 8 (see Supplementary Fig. 8). In both initial and extended spaces, we observe a clear increase of the stability within the runs.

Once the stability and the accuracy of the DNN are established, a final DNN is trained with all the data generated during the experimental runs. This DNN surrogate model is used to generate spectra over the whole parameter space. A software was developed to navigate continuously in the parameter space and display the predicted absorbance spectra obtained at a specific condition (see Supplementary Movie 2). Furthermore, using the CIE 1931 colour spaces, each absorbance spectrum can be converted to the colour that the human eye would see while observing the generated droplets of nanoparticles. Figure 5 shows the colours seen by the DNN surrogate model on the $\{Q_{AgNO_3}, Q_{seed}\}$ plane of the parameter space. For four different regions of the parameter space, the absorbance spectra predicted by the DNN are compared to experimental spectra obtained for similar conditions to illustrate the relevance of this representation. The

diversity of colours obtained reflects the complex link between the absorbance spectrum and droplet chemical composition.

Discussion on ML initialization

Recently, a number of studies on active-learning guided synthesis have demonstrated the ability of BO to explore a large parameter space with continuous and/or discrete variables[14,15,26] and accelerate materials discovery. Using a similar HTE platform as in the present study, Bezinge et al.[14] incorporated a Kriging-based algorithm in the loop to optimize towards certain optical properties and managed to extract some knowledge on the spectral amplitude and FWHM at a targeted emission wavelength. However, their proposed method uses a large data set of 40 experimental points for a 3D parameter space to initialize the ML algorithm, which is far above ten times the dimensionality of their parameter space. Ref. [27] is another recent example of successful optimization for the synthesis of metal halide perovskite materials in which a large initial experimental data set of more than 8000 conditions obtained by RS is used to train support vector machines and neural networks. How much experimental data is necessary to accurately train a DNN or any other regressor remains an open question in ML. Since the complexity of the experimental parameter space is unknown at the start of an experimental campaign, one cannot predict the number of experimental data required to accurately train a regressor. Our present work proposes an experimental approach to bypass this initialization problem by using BO to continue feeding the DNN while the DNN is not trained enough for doing inverse design.

Sequential or batch optimization?

Only one experimental run is required to start the proposed two-step algorithm. The batch size of an experimental run was chosen to minimize the experimental cost. By increasing the batch size to 15, the time and solvents required during the cleaning process of our HTE platform were significantly reduced (see Supplementary Discussion 4). To avoid the clustering of data points in the same batch, we selected LP as the acquisition function, as it iteratively penalizes the points in the neighbourhood of those already picked by the acquisition function. Gonzalez et al.[22] have demonstrated that LP outperforms a wide range of baselines in batch BO. With such an acquisition function, the computational cost for evaluating the loss in the whole parameter space is similar for a batch size

of 15 points or for a batch size of 1 data point. The presented two-step approach can be adapted to other HTE platforms by choosing the appropriate batch size.

Incorporating DNN in a HTE loop

The neural network was chosen for its regression ability when the dataset is large enough. Neural networks have shown their potential to help extract fundamental knowledge from the literature[28] and from experimental data[29]. Umehara et al.[29] demonstrated the ability for CNN-gradient analysis to extract a list of observations on optimization and composition-property relationships in large parameter space with high dimensionality. Our workflow is an attempt to incorporate this DNN ability to extract knowledge in a loop with an HTE platform that rapidly samples the parameter space. Epps et al.[30] recently proposed a similar approach by incorporating an ensemble of 500 neural networks in the HTE loop. This approach allows testing the DNN hyperparameters during the HTE loop. Its efficiency was demonstrated for a simple DNN architecture (3 output parameters, 2 layers and up to 25 nodes). However, using such an ensemble of neural networks to predict full absorbance spectra would dramatically increase the computation time. The solution that we proposed, consisting of a single DNN with an architecture adapted to the full optical spectral resolution, attempts to get a good compromise between the computation cost to find the best choice of hyperparameters and the regression accuracy.

Experimental uncertainty

In this study, the simulated target spectrum is not expected to be reachable because of the nanoparticles polydispersity, since we deliberately chose a chemistry route that is not optimal for producing monodisperse nanoparticles. Thus, a lot of information is hidden in the full absorbance spectrum and we chose to use the whole spectrum to train our algorithms.

The uncertainty in the experimental data was carried into both BO and DNN algorithms by individually using the absorbance spectra of each droplet replicates to train both algorithms. The surrogate model used for BO is a GP model which produces a probabilistic estimate for the loss values. Although the DNN is trained with all droplet replicates, there is no probabilistic method for the DNN.

Choice of the surrogate model and of the acquisition function

In the proposed two-step approach, BO was chosen for its efficiency at handling sparse datasets, as well as for its simplicity, since it requires little hyperparameters tuning. GP performance is not necessarily the best choice to learn complex systems[30] but it allows a quick optimization in our case, while the DNN takes the role of learning the complexity of the parameter space.

Other approaches have been recently proposed for similar optimization and regression problems. Williams et al.[31] conducted a study similar to this work, in which the regression algorithm was trained with a full-factor grid search and different regressors were ranked on their accuracy, showing in their case a better performance with random forest. For the chemical synthesis considered in our study, the use of BO to sample the space for DNN seems to allow a more accurate prediction of the position of the lowest loss than using a pure grid search approach (see Supplementary Discussion 2). Another study from Häse et al.[32] reported the efficiency of Bayesian neural networks (BNN) to optimize nonlinear chemical reaction networks. However, BNNs usually require a larger set of data than GP to become stable. Some other studies suggest that BO with an adaptive kernel might discover finer regression features[33]. However, most prior work focuses on either optimization which lacks interpretability and

transferability when the target changes[5] or inverse design using regression which uses a static dataset[20].

Since each parameter space has its own complexity, there is no single optimal choice of algorithm for all experimental systems. However, once the two-step algorithmic approach has been performed, one could use the experimental data to compare different types of regressor and acquisition functions and determine which combination would have performed the best in terms of accuracy of the localization of the best performer and in terms of accuracy of the regression around the best performer. The possibility to change the acquisition function in the second step of the framework should also be considered, especially in the event that a higher regression accuracy is required close to the target. Algorithm selection using information criteria such as Akaike information criterion[34] and Bayesian information criterion[35] could also be used to maximize time- and resource-efficiency of closed-loop laboratories, e.g., by leveraging co-evolution, physics-fusion, and related strategies[36,37].

In conclusion, we demonstrated the performance of a two-step framework algorithm that combines BO and DNN in a loop with an HTE platform, to optimize the synthesis of silver nanoprisms. The optimization process is accelerated by the offline introduction of the DNN after a few runs of targeted sampling with a BO algorithm. By following the evolution of the loss function and of the regression function over the runs, we could determine at which run the DNN starts to better predict the region around the target position in the parameter space. The process is interpretable, and knowledge can be extracted. The feature importance shows that, even if each parameter plays a role, the silver nitrate and silver seeds remain the most influential parameters for targeting silver nanoprisms. The correlation matrices give information on how the parameters and their ratios affect either the shape of the amplitude of the absorbance spectra. Moreover, absorbance spectra can be predicted all around the target to understand the sensitivity of the optical properties of the synthesized nanomaterial on process parameters. In addition to this, this framework trains a transferable algorithm, since the final trained DNN can then be used to optimize the synthesis towards a different target. Furthermore, the inverse design could be performed using the final DNN to synthesize nanoparticles with optical properties that are different from our initial target.

The developed methodology is generally applicable to other materials synthesis in an HTE loop and can be adapted to other types of HTE platforms. The set of experimental data collected during this study could be used in further work to determine how other acquisition functions and regressor would have performed in such a two-step framework, and compare it to other single step frameworks like BNN.

METHODS

Bayesian optimization

BO has many advantages which make it suitable to kick-start the sampling of the parameter space. As the response surface between the process variables and the targeted loss is unknown, the optimization of process variables can be treated as optimization of a black-box function. BO has been shown to outperform other global optimization methods on various benchmark functions[38]. A GP is chosen as the surrogate model for the BO, considering that the parameter space is continuous. An important aspect of defining the GP model is the kernel and its related hyperparameters. This controls the shape of the regression function[39], which corresponds to the fitting of the response surface between the process variables and the targeted loss.

We select a BO with GP surrogate model for the following reasons: first, implementation of BO with GP is less sensitive to the initial choice of hyperparameter selection of the algorithm[39]. The functional relationship f between the process parameters and the absorbance spectrum is expensive to evaluate and possibly noisy. This eliminates most of the exhaustive search methods such as grid sampling and RS. Ref. [40] has

shown that BO requires a smaller initial dataset and fewer iterations to reach the optimal than a genetic algorithm. There are five different process variables in the experiments. This falls into the sweet spot for BO[41]. There are a number of surrogate models that can be selected for BO such as GP, tree-based algorithms and NN. Tree-based algorithms are not adapted in this study as the process variables are continuous[32]. GP is selected since the number of hyperparameters are much smaller than the NN. Moreover, the uncertainty of a fitted GP is known, as such, it is easy to make a trade-off between exploration and exploitation. Driven by these considerations, BO coupled with GP is used to actively sample the chemical space.

$$
\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \sim N \left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{bmatrix}, \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{bmatrix} \right) \tag{1}
$$

GP is defined in Eq. (1). We can denote this equation using the notation: $f(\mathbf{X}) \sim GP(m(\mathbf{X}), k(\mathbf{X}, \mathbf{X}'))$, where $\mathbf{X} = \{x_1, \ldots, x_n\}$ is the vector of process variables, $m(\mathbf{X})$ the mean function, $k(\mathbf{X}, \mathbf{X}')$ the covariance matrix between all possible $(\mathbf{X}, \mathbf{X}')$ pairs. We use Matern 52 kernel in the covariance matrix[23] and implement the batch BO with LP[22] to suggest a batch of 15 data points to align with the experimental setup.

We use expected improvement (EI) as the acquisition function to select the next experimental conditions that trade-off exploration and exploitation.

$$
EI(\mathbf{X}) = \max(0, E(f(\mathbf{X}) - f(\mathbf{X}^+))) \tag{2}
$$

where $f(\mathbf{X}^+)$ is the value of the 15 best samples and $\mathbf{X}^+$ is the location of that 15 data points.

$$
\mathbf{X_{new}} = \arg\max EI(\mathbf{X}) \tag{3}
$$

The suggested points for the next experiments are the points that maximize the expected improvement. EI can be analytically expressed as:

$$
EI(\mathbf{X}) = (\mu(\mathbf{X}) - f(\mathbf{X}^+) - \xi)\Phi(Z) + \sigma(\mathbf{X})\varphi(Z) \tag{4}
$$

$$
Z = \frac{\mu(\mathbf{X}) - f(\mathbf{X}^+) - \xi}{\sigma(\mathbf{X})} \tag{5}
$$

where $\mu(\mathbf{X})$ and $\sigma(\mathbf{X})$ are the mean and the standard deviation of the GP posterior at $\mathbf{X}$. Φ and φ are the cumulative density function and probability density function of a normal distribution. ξ is the jitter value which determines the exploration to exploitation ratio. The higher ξ, the more explorative the BO is. In this study, we fix the jitter value at 0.1.

Neural network

Supplementary Fig. 2 shows the architecture of the neural networks used in this work. The architecture was chosen to catch the complexity of the system while keeping a reasonable computation time. The input layer is composed of five nodes, followed by four hidden layers (with 50 nodes, 100 nodes, 200 nodes, and 500 nodes). The output layer is composed of 421 nodes, which are corresponding to the UV–Vis spectral data points. As an exploratory work without much knowledge about the parameter space, we choose ReLU for all the activation functions for ease of convergence, and the cost function is an MSE. The weight and bias are updated at each run of the HTE loop. The number of the initially hidden layers is determined by Eq. (6), which is investigated by Stathakis et al.[42], where m is the number of output nodes and N is the number of data points. In this work, m is 421, and N is determined by the data points of each run (around 300). Goodfellow et al.[43] demonstrated empirically that using deep networks with many layers may be a heuristic approach to configure networks for challenging and complex predictive modelling problems.

$$
\text{Number of nodes required for the last hidden layer} = 2\sqrt{(m+2)N} \tag{6}
$$

Since the initial DNN is trained by few and under representative data, the obtained function is not well trained. We incorporate grid search over the whole parameter space for the selection of the best recipes with a minimal loss, with the aim to enforce a regularization term during the optimization process. Starting from run 6, the DNN joined the optimization process. DNN 6 was constructed following the above-mentioned specifications. Afterwards, DNN 6 was trained and established with experimental data suggested by BO runs 1 to 5. DNN 6 was used as a mapping function between the five process variables and the corresponding UV–Vis spectrum. Grid search of the 5D parameter space was conducted to generate each spectrum corresponding to each data point: Q_{AgNO_3}, Q_{TSC}

and Q_{seed} in the range of [0.5:80]% with point interval of 5%, Q_{PVA} in the range of [10:40]% with point interval of 5%, and Q_{total} in the range of [200:1000] μL/min with point interval of 100 μL/min. After that, the loss of each spectrum can be calculated according to our defined loss function and ranked in ascending order. The DNN 6 would select the 15 combinations of the five variables which have the top 15 minimal losses and recommend them to the experimentalists to carry out the synthesis and collect the actual spectral data. Further, the DNN 7 was trained with data of BO runs 1 to 6 and conducted grid optimization; DNN 8 was trained with data of BO runs 1 to 7 and conducted grid optimization.

Definition of the loss function

The loss function is defined as:

$$
Loss = 1 - \delta(A_{measured}^{max}) \frac{|A_{measured}(\lambda).A_{target}(\lambda)|}{\|A_{measured}(\lambda)\|.\|A_{target}(\lambda)\|} \tag{7}
$$

$$
\text{with } \delta(A) = \begin{cases} \frac{A}{0.7} & \text{if } A < 0.7 \\ 1 & \text{if } 0.7 \le A \le 1.2 \\ 0 & \text{if } A > 1.2 \end{cases}
$$

where $A_{measured}^{max}$ is the maximal value of the measured absorbance spectrum. The cosine similarity between the measured and targeted spectra quantifies the shape similarity of the two spectra. Thus, by using the cosine similarity in the definition of the loss function, the shape of the absorbance spectrum can be optimized. However, to avoid saturation or noisy optical measurements, the BO and DNN suggestions should remain in the detection range of the UV–Vis spectrometer. The amplitude function δ is designed for this purpose. It forces both BO and DNN to suggest conditions with a maximal absorbance higher than half the detection limit of the spectrometer.

SHAP analysis

To quantify the most significant process parameters and their impact on model performance, we use SHAP (SHapley Additive exPlanations) algorithm to evaluate both the BO and the DNN. SHAP algorithm is a game-theoretic approach that provides the additive feature importance measure of any ML model[44]. The objective is to carry out a prediction task for a single data point of the dataset. The gain is the actual prediction for this data point minus the average prediction for all data points in this dataset. It is assumed that all the feature values of a data point contribute together to the gain. In this work, the feature values (Q_{AgNO_3}, Q_{PVA}, Q_{TSC}, Q_{seed} and Q_{total}) worked together to achieve the predicted value in terms of loss. Our goal is to explain the difference between the actual prediction value and the average prediction values in the whole dataset. Specifically speaking, the value of the difference should be partially assigned to the five features, and the partially assigned values should represent each feature's importance at the specified data point.

Materials

Silver nitrate (99.9%) was purchased from Strem Chemicals Inc. Silver seeds (10 nm, 0.02 mg/mL in aqueous buffer stabilised with sodium citrate), sodium citrate tribasic dihydrate (TSC) (≥99.0%, ACS) and polyvinyl alcohol (PVA) (Mowiol 8-88, $M_w \sim 67{,}000$) were purchased from Sigma Aldrich. The silver seeds were used as received. L-(+)-ascorbic acid (AA) (99+%) was purchased from Alfa Aesar. Silicone Oil (PMX-200, 10 cSt) was purchased from MegaChem Ltd. and used as received. Ultrapure water (18.2 MO at 25 °C) was obtained from a Milli-Q purifier.

Experimental design

Silver seed (0.02 mg/mL), TSC solution (15 mM), AA solution (10 mM), PVA solution (5 wt%), water, silver nitrate solution (6 mM) and silicone oil were loaded in Hamilton glass syringes. The chemical reactants were chosen based on the literature[45,46]. Their loading concentration was estimated based on the concentrations found in the literature, and considering that a ten times higher concentration of AgNPs is required to measure the absorbance spectrum in a 1 mm optical chamber.

The syringes containing the aqueous phases were all connected to a 9-port PEEK manifold (Idex) through PTFE tubes. The manifold output and the oil syringe were connected to a PEEK Tee-junction (1 mm thru-hole), allowing the controlled generation of monodisperse droplets.

Nanoparticles are synthesized in aqueous sub-microliter droplets (see Fig. 1). In such a flow system, the concentration of each reactant is directly proportional to the flow rate ratio Q_i (%) between the flow rate of the reactant and the total aqueous flow rate. By adjusting the flow rate of the solvent (water), the flow rate ratios Q_{seed} of silver seeds, Q_{AgNO_3} of silver nitrate, Q_{TSC} of trisodium citrate and Q_{PVA} of PVA are independently controlled by varying the flow rate of the corresponding solutions using LabView automated syringe pumps. The flow rate ratio Q_{AA} of ascorbic acid is kept constant. The mixing of the reactants inside the droplet depends on the speed of the droplet, which is directly proportional to the total flow rate Q_{total} (µL/min) of both oil and aqueous phases. The absorbance spectra of the droplets are measured inline, and the five controlled variables Q_{seed}, Q_{AgNO_3}, Q_{TSC}, Q_{PVA} and Q_{total} are used as input parameters for the two-step optimization framework and the absorbance spectra as the output.

The boundaries of the parameter space were defined by the range of accessible flow rates for each solution. Since the mixing inside the droplets is directly linked to the total flow rate in the reaction tube, the sum of all flow rates (Q_{total}) was varied between 200 and 1000 µL/min. The total aqueous flow rate was kept equal to the flow rate of the oil in order to keep the droplet volume constant. The ascorbic acid flow rate was kept equal to 10% of the oil flow rate at all time. The flow rate of the five other aqueous phases could be selected by the ML algorithm within a certain range of percentage of the total aqueous flow rate. First, in the initial parameter space, the silver seeds, silver nitrate and TSC flow rates were kept between 4 and 20%, and the PVA solution between 10 and 40%. Then, in extended parameter space, while the PVA flow rate was still kept between 10 and 40%, the flow rate restrictions for silver seeds, silver nitrate and TSC were partially released, so that the only limitations left were: (1) that the sum of the silver seeds, silver nitrate, TSC and PVA flow rates should stay below 90% of the aqueous flow rate, (2) that the silver seeds, silver nitrate and TSC flow rates should remain above 0.5% of the total aqueous flow rate.

After the Tee-junction, the droplets were forced to flow in a 1.25 m long PFA tube (1 mm ID). One meter after the Tee-junction, the PFA reaction tube entered a customized optical chamber. For each condition suggested by the ML algorithm, droplets were generated until the first droplet exits the optical chamber. The total flow rate was then decreased to 30 µL/min and the absorbance spectra of 20 consecutive droplets were recorded at 1.4 fps with a spectrometer (Flame-T-UV–Vis, Ocean Optics) combined with a Deuterium-Halogen light source (DH-2000-BAL, Ocean Optics).

TEM imaging and analysis

To validate the integration of the full absorbance spectra in the HTE loop, we used TEM (JEM-2100F) imaging to measure the size dispersity of the synthesized nanoparticles. Since TEM imaging is time-consuming when compared to inline absorbance measurements, we observed the nanoparticles only for the best performing conditions of each run. Statistical analyses on the TEM images were performed with MATLAB, using several hundreds of particles for each condition.

Simulation of the target spectrum

The target spectrum was simulated by DDSCAT[47] using the optical constants of silver[48]. The simulation was performed for a 50 nm wide and 10 nm thick triangular prism, using 13,398 dipoles, and averaging the results on two incident polarisations and eight different angular orientations of the target around the axis perpendicular to the direction of propagation of the incident light. To limit the computation time, the target spectrum was calculated for 25 wavelengths equally spaced between 380 and 800 nm.

DATA AVAILABILITY

The data generated and analyzed during the current study can be found in our repository (https://github.com/acceleratedmaterials/AgBONN).

CODE AVAILABILITY

Details of our code implementation can be found in our repository (https://github.com/acceleratedmaterials/AgBONN).

REFERENCES

1. Altae-Tran, H., Ramsundar, B., Pappu, A. S. & Pande, V. Low data drug discovery with one-shot learning. *ACS Cent. Sci.* **3**, 283–293 (2017).
2. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. & Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **23**, 1241–1250 (2018).
3. Erickson, B. J., Korfiatis, P., Akkus, Z. & Kline, T. L. Machine learning for medical imaging. *RadioGraphics* **37**, 505–515 (2017).
4. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
5. Yamawaki, M., Ohnishi, M., Ju, S. & Shiomi, J. Multifunctional structural design of graphene thermoelectrics by Bayesian optimization. *Sci. Adv.* **4**, eaar4192 (2018).
6. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
7. Olivecrona, M., Blaschke, T., Engkvist, O. & Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminform.* **9**, 48 (2017).
8. Nash, W., Drummond, T. & Birbilis, N. A review of deep learning in the study of materials degradation. *npj Mater. Degrad.* **2**, 37 (2018).
9. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: status, challenges and perspectives. *Adv. Sci.* **6**, 1900808 (2019).
10. Trivedi, V. et al. A modular approach for the generation, storage, mixing, and detection of droplet libraries for high-throughput screening. *Lab Chip* **10**, 2433–2442 (2010).
11. Knauer, A. et al. Screening of plasmonic properties of composed metal nanoparticles by combinatorial synthesis in micro-fluid segment sequences. *Chem. Eng. J.* **227**, 80–89 (2013).
12. Lignos, I. et al. Synthesis of cesium lead halide perovskite nanocrystals in a droplet-based microfluidic platform: fast parametric space mapping. *Nano Lett.* **16**, 1869–1877 (2016).
13. Epps, R. W., Felton, K. C., Coley, C. W. & Abolhasani, M. Automated microfluidic platform for systematic studies of colloidal perovskite nanocrystals: towards continuous nano-manufacturing. *Lab Chip* **17**, 4040–4047 (2017).
14. Bezinge, L., Maceiczyk, R. M., Lignos, I., Kovalenko, M. V. & Demello, A. J. Pick a color MARIA: adaptive sampling enables the rapid identification of complex perovskite nanocrystal compositions with defined emission characteristics. *ACS Appl. Mater. Interfaces* **10**, 18869–18878 (2018).
15. Salley, D. et al. A nanomaterials discovery robot for the Darwinian evolution of shape programmable gold nanoparticles. *Nat. Commun.* **11**, 2771 (2020).
16. Shabanzadeh, P., Yusof, R. & Shameli, K. Neural network modelling for prediction size of silver nanoparticles in montmorillonite/starch synthesis by chemical reduction method. *Dig. J. Nanomater. Bios.* **9**, 1699–1711 (2014).
17. Thanh, N. T. K., Maclean, N. & Mahiddine, S. Mechanisms of nucleation and growth of nanoparticles in solution. *Chem. Rev.* **114**, 7610–7630 (2014).
18. Lee, J., Yang, J., Kwon, S. G. & Hyeon, T. Nonclassical nucleation and growth of inorganic nanoparticles. *Nat. Rev. Mater.* **1**, 16034 (2016).
19. Yuan, B. et al. Machine-learning-based monitoring of laser powder bed fusion. *Adv. Mater. Technol.* **3**, 1800136 (2018).
20. Kumar, J. et al. Machine learning enables polymer cloud-point engineering via inverse design. *npj Comput. Mater.* **5**, 73 (2019).
21. Voznyy, O. et al. Machine learning accelerates discovery of optimal colloidal quantum dot synthesis. *ACS Nano* **13**, 11122–11128 (2019).
22. Gonzalez, J., Dai, Z., Hennig, P. & Lawrence, N. Batch Bayesian optimization via local penalization. In: *Proc. 19th International Conference on Artificial Intelligence and Statistics.* 51, 648–657 (2016).
23. Dai, Z. et al. GPyOpt: a Bayesian optimization framework in python. http://github.com/SheffieldML/GPyOpt (2016).
24. Shuford, K. L., Ratner, M. A. & Schatz, G. C. Multipolar excitation in triangular nanoprisms. *J. Chem. Phys.* **123**, 114713 (2005).
25. Aherne, D., Ledwith, D. M., Gara, M. & Kelly, J. M. Optical properties and growth aspects of silver nanoprisms produced by a highly reproducible and rapid synthesis at room temperature. *Adv. Funct. Mater.* **18**, 2005–2016 (2008).
26. Burger et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
27. Li, Z. et al. Robot-accelerated perovskite investigation and discovery. *Chem. Mater.* **32**, 5650–5663 (2020).
28. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
29. Umehara, M. Analyzing machine learning models to accelerate generation of fundamental materials insights. *npj Comput. Mater.* **5**, 34 (2019). & al.
30. Epps, R. W. et al. Artificial chemist: an autonomous quantum dot synthesis bot. *Adv. Mater.* **32**, 2001626 (2020).

31. Williams, T., McCullough, K. & Lauterbach, J. A. Enabling catalyst discovery through machine learning and high-throughput experimentation. *Chem. Mater.* **32**, 157–165 (2020).

32. Häse, F., Roch, L. M., Kreisbeck, C. & Aspuru-Guzik, A. Phoenics: a Bayesian optimizer for chemistry. *ACS Cent. Sci.* **4**, 1134–1145 (2018).

33. Wang, H. & Li, J. Adaptive Gaussian process approximation for Bayesian inference with expensive likelihood functions. *Neural Comput* **30**, 3072–3094 (2018).

34. Sakamoto, Y., Ishiguro, M. & Kitagawa, G. Akaike information criterion statistics (D. Reidel, Dordrecht, 1986).

35. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).

36. Wolpert, D. H. & Macready, W. G. Coevolutionary free lunches. *IEEE Trans. Evolut. Comput.* **9**, 721–735 (2005).

37. Rohr, B. et al. Benchmarking the acceleration of materials discovery by sequential learning. *Chem. Sci.* **11**, 2696–2706 (2020).

38. Jones, D. R. A taxonomy of global optimization methods based on response surfaces. *J. Glob. Optim.* **21**, 345–383 (2001).

39. Rasmussen C. E. Gaussian processes in machine learning. In: *Advanced Lectures on Machine Learning* (Springer-Verlag, Berlin Heidelberg, 2004).

40. Wang, Z. L., Ogawa, T. & Adachi, Y. Influence of algorithm parameters of Bayesian optimization, genetic algorithm, and particle swarm optimization on their optimization performance. *Adv. Theory Simul.* **2**, 1900110 (2019).

41. Frazier, P. I. Bayesian optimization. In *Recent Advances in Optimization and Modeling of Contemporary Problems*, 255–278 (2018).

42. Stathakis, D. How many hidden layers and nodes? *Int. J. Remote Sens.* **30**, 2133–2147 (2009).

43. Goodfellow, I., Bengio, Y. & Courville, A. Deep Learning: Adaptive Computation and Machine Learning (MIT Press, Cambridge, 2016).

44. Lundberg, S. & Lee, S. I. A unified approach to interpreting model predictions. In *Proc. 31st Conference on Neural Information Processing Systems* (NIPS, 2017).

45. Singh, A. K. et al. Nonlinear optical properties of triangular silver nanoparticles. *Chem. Phys. Lett.* **481**, 94–98 (2009).

46. Potara, M., Gabudean, A.-M. & Astilean, S. Solution-phase, dual LSPR-SERS plasmonic sensors of high sensitivity and stability based on chitosan-coated anisotropic silver nanoparticles. *J. Mater. Chem.* **21**, 3625–3633 (2011).

47. Draine, B. T. & Flatau, P. J. Discrete dipole approximation for scattering calculations. *J. Opt. Soc. Am. A* **11**, 1491–1499 (1994).

48. Johnson, P. B. & Christy, R. W. Optical constants of the nobel metals. *Phys. Rev. B* **6**, 4370–4379 (1972).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

F.M.B., Z.R., T.B., and S.A.K. conceived this study; F.M.B., W.K.W., F.Z., and J.X. designed and supervised the autonomous experiment; Z.R., T.H., I.P.S.T., S.J., T.B., Q.L., and X.W. developed machine-learning algorithms; F.M.B., D.B. conducted the plasmon resonance simulations; F.M.B., Z.M. conducted TEM; F.M.B., Z.R., T.H., and T.B. wrote the manuscript; K.H., S.A.K., T.B., Q.L., and X.W. supervised the research. F.M.B., Z.R., and T.H. are the co-first authors of this paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.