ADVANCED FUNCTIONAL MATERIALS
www.afm-journal.de

Predicting Aggregation Behavior of Nanoparticles in Liquid Crystals via Automated Data-Driven Workflows

*Yueyang Gao, Niamh Mac Fhionnlaoich, Max Besenhard, Arun Pankajakshan, Federico Galvanin, and Stefan Guldin**

Gold nanoparticles (AuNPs) have gained prominence as versatile nanoscale building blocks in chemical and biomedical research. Liquid crystals (LCs) offer a promising composite matrix for fundamental research and in a variety of applications. However, optimizing the solubility of AuNPs within the LC matrix remains challenging due to the interplay of multiple experimental variables, necessitating extensive combinatorial trials. In this study, an automated AuNP synthesis platform combined with a Design of Experiment (DoE) framework was employed to streamline the optimization process. A random forest model, trained on a relatively small dataset, successfully predicted nanoparticle aggregate classifications with high accuracy. Aggregate behavior was further analyzed using UV–vis spectroscopy with automated data processing for feature reduction. These steps enhanced the closed-loop optimization workflow by iteratively constructing a generalized additive model for predicting spectral characteristics. AuNPs optimized for solubility were deployed in subsequent experiments for temperature-induced hierarchical assembly driven by the phase transition of the thermotropic LC. Computer vision methods were used to quantify the reversibility of LC-AuNP composites during self-assembly, utilizing entropy values derived from a pattern recognition algorithm. This study highlights the benefits of integrating cross-disciplinary approaches to refine analytical workflows, advancing the discovery of nanomaterial systems with programmable and reconfigurable features.

1. Introduction

Nanoparticles have garnered significant interest across various technological and medical fields. Functionalization with chemical entities (e.g., thiol ligands) renders gold nanoparticles (AuNPs) a highly versatile materials system,[1,2] enabling tunable properties through co-assembly with different ligands. Depending on the matrix, AuNPs-based nanocomposites also show notable potential in stability, biocompatibility, catalytic activity, and size-dependent optical properties.[2–4] The tunable nature of nanoparticle surface characteristics is a key factor in their versatility as the ligand shell encompassing the surface not only stabilizes the NP from aggregation but also dictates their interaction with the surrounding environment. The composition of the ligand shell, their spatial distribution, and the completeness of surface coverage are pivotal in determining the stability and functionality of nanoparticles.[5] These factors are governed by the reaction conditions during the functionalization and often exhibit complex, non-linear relationships that are challenging to decipher through traditional experimental approaches. Tackling such a complex and expansive parameter space presents a substantial experimental burden. Moreover, manual exploration of these conditions is time-consuming and prone to reproducibility issues.

The advent of automation, including robotic platforms, has revolutionized high-throughput experimentation, enabling faster, less labor-intensive sample handling with enhanced reproducibility,[6] leveraging automation a promising solution to accelerate the AuNP functionalization within the high-dimensional parameter space. As an application-independent methodology which helps reduce the trial numbers,[7–9] Design of Experiment (DoE) can aid in as the initial guidance for constructing the data set since conventional methods like OFAT (one-factor-at-a-time) are inefficient during the screening analysis, especially in identifying complex intervariable effects.[10,11] Therefore, systematic exploration of the functionalization conditions can be carried out through data-driven approaches to identify the key factors and interactions with minimal resource expenditure.

Integrating data-driven workflows with advanced modeling techniques provides deeper insights into the underlying

Y. Gao, N. M. Fhionnlaoich, M. Besenhard, A. Pankajakshan, F. Galvanin, S. Guldin
Department of Chemical Engineering
University College London
London WC1E 7JE, UK
E-mail: guldin@tum.de
S. Guldin
Department of Life Science Engineering
Technical University of Munich
Gregor-Mendel-Straße 4, 85354 Freising, Germany
S. Guldin
TUMCREATE
1 CREATE Way, #10-02 CREATE Tower, Singapore 138602, Singapore
N. M. Fhionnlaoich
APC Ltd.
Building 11, Cherrywood Business Park, Loughlinstown, Dublin D18 DH50, Ireland
The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adfm.202501657

DOI: 10.1002/adfm.202501657

behaviors of AuNPs, facilitating more effective optimization. Machine learning, especially cluster and pattern recognition algorithms, have become powerful tools in scientific research for analyzing complex data sets and solving problems involving overwhelming resource consumption.[12–14] The Random Forest (RF) model has become one of the most well-established supervised algorithms due to its generalizability and flexibility, in handling both classification and regression problems, known for good performance on general datasets, with minimal requirements on data type and feature correlation.[15–17] It has been successfully used in many research areas such as monitoring physiological processes in cells,[18] identifying diseases in serum samples,[19] and predicting structural motifs in spectroscopy.[16] During the tuning process of AuNPs functionalization, the high dimensionality of the feature space normally restricts the conventional analytic workflow for optimization. Therefore, implementing RF model as a data-driven approach to identify optimal experimental factor combinations would benefit the systematic investigation of AuNPs, especially for screening active effects during the functionalization. Since the nanoparticle system has the size-dependent optical feature such as the signature of the surface plasmon resonance (SPR) peak within the visible range, UV–visible spectroscopy offers an effective tool for analyzing AuNPs.[2,3] The shift, location, and bandwidth of the peak offers valuable information including size, aspect ratio, solubility, and aggregation.[20,21] Considering the multiple factors during the analysis of spectra characteristics, a generalized additive model (GAM) is helpful since multivariate function can be represented as sums and compositions of univariate functions.[22] Furthermore, the big advantage of GAM approach is their ability to model nonlinear data while retaining interpretability, therefore greatly improving its versatility when multi-interaction effects need to be considered.[23]

AuNPs functionalized for solubility in a liquid crystal matrix was employed here as a model system. Liquid crystals (LCs) demonstrate unique properties between those of liquids and solid crystals,[24] known for the widespread application in the display technology,[25] also had various applications in optics and colloidal assembly.[26–28] However, the dense packing of mesogens pose significant challenges when incorporating other material,[29] such as nanoparticles, often leading to poor solubility or aggregation.[30–32] A prominent approach to overcoming this involves the functionalization of the AuNPs with a specifically tailored mixture of short spacer ligands and longer mesogenic ligands.[33–35] The spacer ligands create pockets within the ligand shell,[36] allowing the LC mesogens to permeate and interact with the mesogenic ligands in order to promote the solubility. This approach necessitates a balanced ratio of ligands as well as complete and random surface coverage, all of which is subject to the experimental factors during the functionalized reaction. The complexity of this system and its sensitivity of the resulting solubility makes it an ideal model for assessing the effectiveness of the automated data-driven workflow.

In this work, automated functionalization of AuNPs using a robotic platform was leveraged to optimize the ligand exchange functionalization. Additionally, an RF classification model was developed to predict AuNPs aggregation behavior. Iterated construction of the GAM approach was developed and aided by the DoE screening, with batch auto-processing of the spectra data to help accelerate the optimization of soluble AuNPs seed. This enabled further experimentation to target the regions within the design space that did not cause aggregation. Thermotropic liquid crystal 4′-pentyl-biphenyl-4-carbonitrile (5CB) was then introduced as the matrix for LC-AuNPs composites to further study the hierarchical self-assembly behavior of nanometer-sized AuNPs into micrometer-sized agglomerates. Investigation regarding the self-assembly behavior and its reversibility was realized via the computer vision-based methodology, with referenced entropy and the nearest neighbor distance obtained from the algorithm. This integrated approach opens new possibilities in analyzing nanoparticles aggregation behavior with reversible self-assembly features based on micrometer-sized periodicities in liquid crystals. Moreover, it may serve as a blueprint for the design and analysis of nanomaterial systems with applications of NPs in programmable and switchable functional composites.

2. Results and Discussion

The robotic workstation was constructed with integrated modules, allowing the automated functionalization of soluble AuNPs during the experimentation. The downstream analytic toolbox was then developed to process UV–vis spectra and characterization images automatically, which helped analyze the huge amount of data generated. Furthermore, three phases were presented for the modular investigation: 1) initial screening, 2) detailed modelling, and 3) final validation. First, a DoE-based strategy was used for an initial screening of the design space and the results were trained with a random forest (RF) model to classify aggregation behavior. The second phase involved refining the design space and developing a Generalized Additive Model (GAM) to establish a detailed understanding of how experimental conditions affect AuNP solubility in the LC. Finally, validation was performed by thermocycling the optimized AuNP-LC composite to confirm its fully reversible self-assembly, with quantification evaluation based on computer vision analysis.

2.1. Automated Solutions for AuNPs Functionalization and Data-Processing

To accelerate the sample handling during the functionalization of AuNPs, an integrated robotic platform was developed as a dedicated workstation. To realize the temperature-controlled reaction and purified separation during functionalization, the integration of different hardware modules was required, which is quite a challenging problem. For instance, centrifugation is essential in synthetic workflows of AuNPs while conventional devices are not suitable for modular applications and crucially do not offer the required control over sample positions.[37] Herein, a smart centrifuge (SC) with position control function was manufactured based on our previous prototype.[38] Updated with an adaptive plate for integration in versatile platforms, the compact and modular centrifugation facilitates the separation for automated ligand exchange procedure during AuNPs functionalization (Hardware details can be found in Figure S7, Supporting Information, Construction of Robotic Platform). As exhibited in **Figure 1a**, the Opentrons robot (OT-2) was used as a foundation for the automated platform with integrated hardware modules, the general

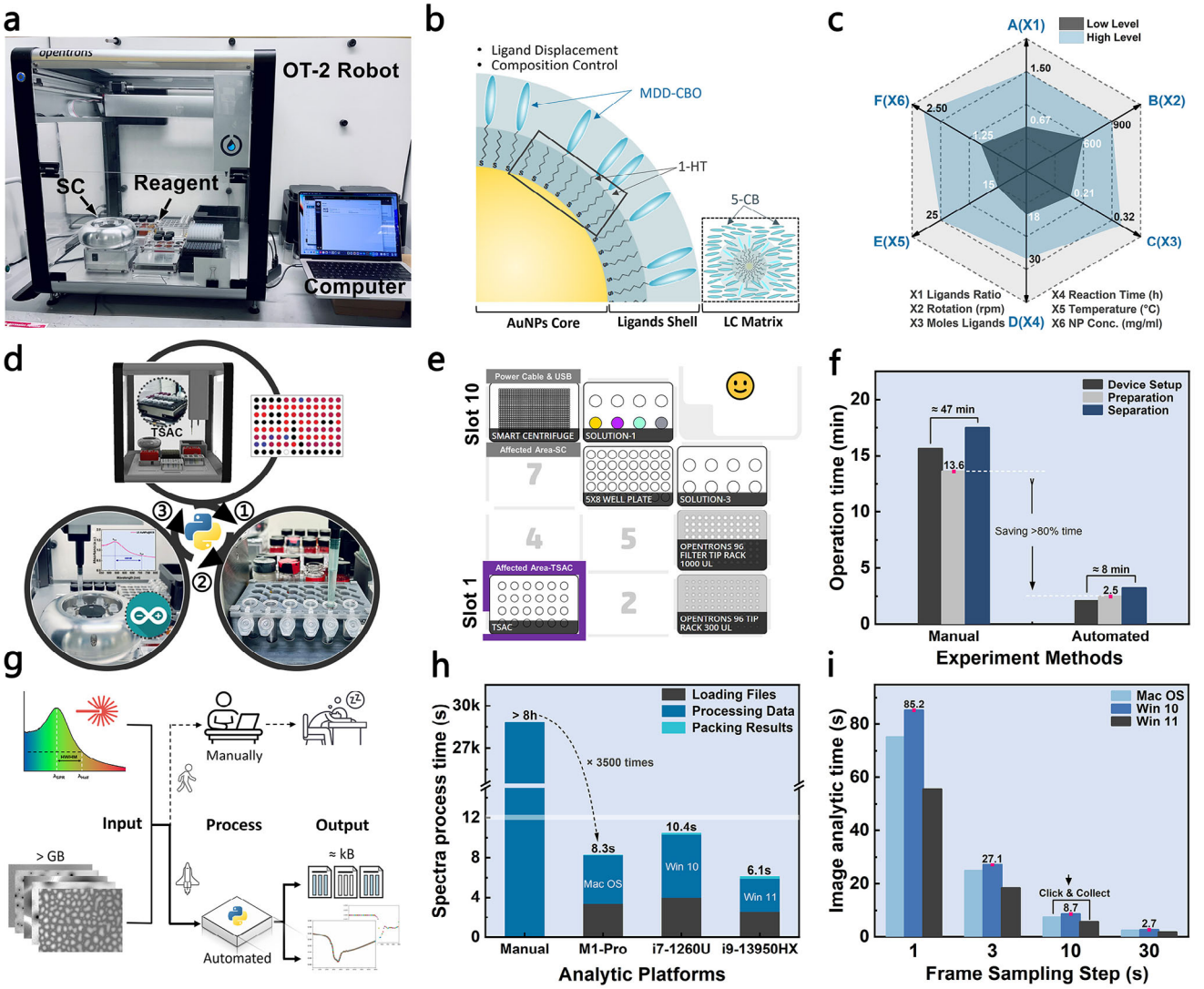

Figure 1. a) Schematic of the automated workstation including the OT-2 robot, customized modules, and a computer. b) Cross section view of the functional AuNPs within LC matrix c) Illustration of the expansive parameter space for designing soluble functionalization of AuNPs. d) Schematic representation of the automated workflow of fabrication, analysis, and optimization of AuNPs. e) Optimized deck layout of modules in the workstation for efficient setup. f) Time comparison for manual and automated methods during functionalization procedures g) Automated solutions for downstream data processing to realize high-throughput analysis. h) Time comparison for processing 50 spectra curves between manual method and different analytic platforms. i) Automated analytic time of processing 900 images during phase transition for quantification results of information entropy under different frame sampling steps, enabling good compatibility across operation systems among Mac OS, Win 10 and 11.

setup of this workstation and related software for protocol designing was summarized in a mapping structure in Figure S8 (Supporting Information). Different functional modules including a thermoshaker with automated clamping system (TSAC) in configuration allowing for reaction temperature and agitation control (Figure S9, Supporting Information), customized labware and smart centrifuge were all integrated. To ensure the precise pipetting process within integrated modules, calibrated coordinates with customized design were programmed to the workstation (The demonstration can be found in Figure S10, Supporting Information). Therefore, customized modules can be recognized by the OT-2, enabling the automated pipetting work to be completed with high efficiency.

Normally, the synthetic procedure for functionalized AuNPs needs to replace the capping layer of the Oleylamine-capped AuNPs (OAm-AuNPs) with specific ligand shell.[39] As demonstrated in Figure 1b, short spacer ligands (1-hexanethiol, 1-HT) and long mesogenic ligands (4´-(12-Mercaptododecyloxy) Biphenyl-4-Carbonitrile, MDD-CBO) were used (molecular structure can be found in Figure S1, Supporting Information) as the ligand shell to complete the surface functionalization of AuNPs core for manipulating its solubility within the LC matrix (5-CB). Since the ligand shell composition, spatial distribution, and surface coverage are pivotal in determining the stability and functionality of the AuNPs, related experimental factors during the functionalization were considered and listed in Figure 1c,

where X1-X3 were ligands mole ratio (1-HT to MDD-CBO), reaction rotation speed, and total moles of ligands; X4-X6 represented reaction time, reaction temperature, and AuNP concentration. Those key parameters were determined by the design of the experiment (DoE) as initial guidance with subsequent iterations based on the resulting solubility during the optimization. The customized workstation realized the controlled reaction for AuNPs functionalization with specific parameters; as the temperature, rotation speed, and time can be programmed within the TSAC, while concentration-related factors can be controlled by the pipetting robot. Therefore, by leveraging lab automation, the AuNP functionalization process was not only greatly accelerated but also enabled the exploration of high-dimensional parameter spaces.

Figure 1d illustrates the standard workflow for fabricating and analyzing the functionalized AuNPs, following the DoE-based initial investigation, combined with AuNPs preparation and purification based on integrated modules. By analyzing the UV–vis spectrum, screening and characterization results of the solubility could then be evaluated for iterated new experimentation rounds. The automated workstation required the joint application of several modules; therefore, the deck arrangement of the platform was optimized to improve the efficiency of setup (Figure 1e). Since the margin of customized modules could affect nearby deck slots, employing the TSAC and centrifuge into the left column as the current layout can minimize their space influence while maintaining functionality during the functionalization. Other customized labware including the SLA printed stock, the acrylic and aluminum supporting plate were also manufactured with the standard footprint to fit the deck via plug and play, enabling great flexibility for automation purposes (Figures S11–S14, Supporting Information). To evaluate the efficiency of the integrated workstation, a typical operation cycle comparison was made between conventional setup with manual experiment and automation. As shown in Figure 1f, a total operation time of 47 min was required for the manual step while similar procedures by automation only took ≈8 min, saving over 80% time over the general operation. Specifically, due to the compact and efficient layout of functional modules within the workstation, time consumption for the device setup part was greatly reduced as the manual method had to spend more time in constructing the reactor, connecting the temperature controller, testing circulating liquid, and so on, while the automated method only needs to wait for the initialization of hardware modules. Furthermore, repetitive pipetting and mixing work was another time-consuming process for the sample preparation. Herein, automated preparing the reaction solution with ligands as well as adjusting AuNPs concentration with the dual channel pipetting robot (OT-2) was 5.4 times faster than the manual process (13.6 min), let alone the precision and reproducibility.

Since the automated workstation offered a feasible solution for high-throughput experimentation, fast and efficient downstream analytic approaches were also required to accelerate the whole workflow. Aided by the workstation, liquid crystal cells with planar alignment were used to prepare the LC-AuNPs nanocomposite for further characterization. To efficiently investigate the nanocomposite, a kinetic series of bright-field (BF) images were automatically captured in real-time via a ZEISS optical microscope system to observe the dynamic diffusion and assembly behavior of AuNPs. Furthermore, the integration of a beam splitter system within the microscope allowed for the simultaneous characterization of the in situ UV–vis spectroscopy (The synchronized characterization setup is shown in Figure S15, Supporting Information). This customized procedure was realized with a fixed sampling step agent, leveraging less labor-intensive loadings during the characterization.

As exhibited in Figure 1g, UV–vis spectral curves contained valuable information as the shift, location, and bandwidth of the peak can be used to evaluate the AuNPs solubility. Besides, BF images provided instant intuitive visual results during the dynamic assembly process. However, during the typical investigation period, a large number of BF images and related spectra curves would be generated simultaneously, resulting in file sizes over GB level. Such a huge amount of input data with different formats during the analytic process is a big challengefor the downstream workflow. First, manually extracting spectral features is time-consuming and certainly not efficient. Besides, it is impractical for humans to give unbiased quantification analysis of thousands of images based on information entropy. Therefore, an automated toolbox was developed for batch processing of spectral curves for feature extraction along with quantification analysis of images based on the computer vision methodology for referenced entropy (see Supporting Information for details, set up of automated analytic platforms), enabling the reduction of data dimensionality.

Figure 1h shows the processing time of 50 spectra curves across different platforms, the analytic toolbox not only provides great compatibility in both Mac OS and Windows system, but also has a notable speed compared with manual analysis. For instance, in the M1 Pro (Mac OS) platform, the total time consumption only takes 8.26 s (including loading files then processing, and packing results), which is 3500 times faster than the manual process, indicating less than 100 ms was spent on each curve for actual extraction processing. Furthermore, the instant image analytic process was also realized when an optimal frame sampling step (FSS) was selected. Figure 1i summarizes the time consumption of the typical quantification analysis of 900 images (GB level) during the phase transition process. considering the balance between performance and speed, aA fast and efficient analytic process of 8.69 s was achieved on Win 10 platform under FSS of 10, enabling the click and collect of the formatted quantification results in kB level. Therefore, the automated toolbox for downstream data processing not only accelerated the analytic procedure but also provided formatted results for further data-driven workflow, enabling a robust foundation for high-throughput analysis.

2.2. Predicting Aggregation Behavior of AuNPs Based on Random Forest Model

An initial analytical workflow was developed to segment the design space into conditions causing or preventing AuNP aggregation, enabling better capture of underlying experimental interactions. To efficiently meet this, selecting a proper classifier model based on machine learning at the screening step would benefit the purpose of the prediction of aggregation behavior. The proposed pipeline for training the model is demonstrated

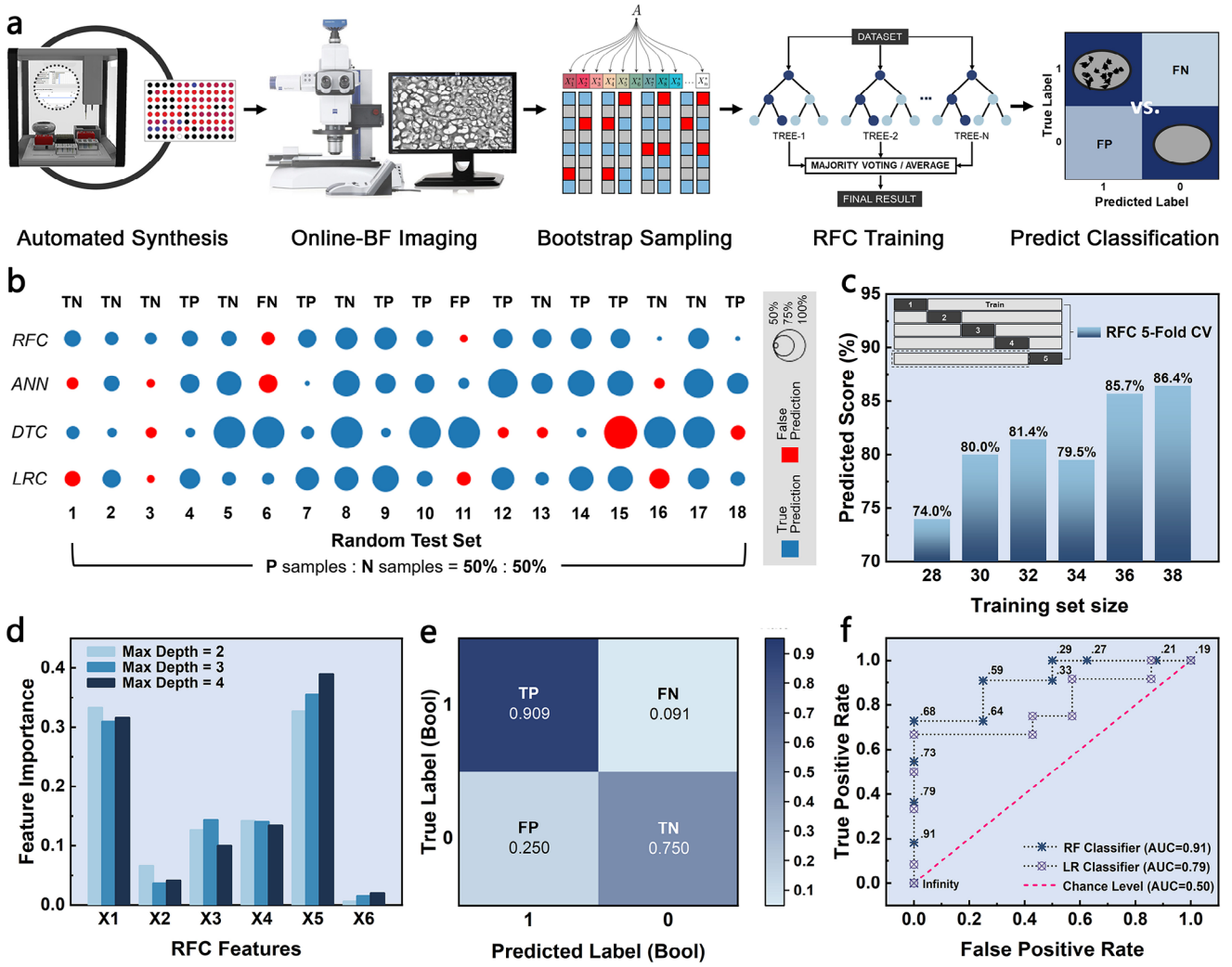

Figure 2. a) Working pipeline for training the Random Forest model with the automated platform for classified prediction of NP agglomerates. b) Visualized comparison of the prediction outcome and confidence for classification results of different models. c) Prediction score as a function of training set size. Each training set was tested by a five-fold cross-validation with outcome averaged from the five tasks in the respective subset. d) Feature importances obtained from the RFC with different depth thresholds. e) The confusion matrix outcome of the RFC, with predicted label versus true label for visualized comparison. f) The receiver operating characteristic curve comparison of RFC and LRC, with comparison to the chance level (dashed line).

in **Figure** 2a, with the robotic workstation enabling automated experimentation for AuNPs functionalization. Since the AuNPs solubility can be affected by many experimental factors, the design of experiment was applied to optimize the experimental budget, hence improving the fabrication efficiency of the LC-AuNPs composites as well. The online imaging system integrated with the temperature stage was controlled by a program that set a predefined temperature of 40 °C to get consistent results in the isotropic phase, a kinetic series of images was captured automatically via the sampling agent at a defined interval. It should be noted that the loading of AuNPs in the LC matrix was fixed (6 wt.%) to ensure the evaluation consistency. A comparison of AuNP aggregation in the LC matrix is demonstrated in Figure S3 (Supporting Information).

Data-driven approaches, especially machine learning, offer unique opportunities in the analysis of AuNPs functionalization. To make the working pipeline more efficient, sampling of the predicted classification for aggregates needs to be considered first. Herein, the expansive parameter space of 6 reaction conditions during functionalization were selected as the design region for the training input, where factors X1-X3 were ligands mole ratio (1-Hexanethiol to MDD-CBO), reaction rotation speed, and total moles of ligands; X4-X6 represented reaction time, reaction temperature, and AuNP concentration. Random Forest Classifier (RFC) was one of the ensembled learning models based on the bootstrap aggregating (bagging) algorithm, where bootstrap sampling was performed to ensure a number of equally sized subsets of a dataset are extracted with replacement. The generated subsets could then be used to train the parallel weak learners (Figure S2, Supporting Information). Due to the integration of multiple decision trees, its prediction result will be less prone to overfitting with better generalization ability. Therefore, RFC was found to be good choice for predicting the aggregation behavior of AuNPs helping to make

initial classification as well as giving screening analysis for the key factors.

To validate the baseline performance, the 6 reaction conditions were selected as input features for the initial training step. A benchmark test was performed among different classification models to see which was the most suitable, including logistic regression classifier (LRC), decision trees classifier (DTC), and artificial neuron network (ANN). The performance test was based on the same training set with a size of 40. To ensure the rationality of the validation, an external random test set size of 18 which contains equal positive (aggregated) and negative (no aggregation) samples was adopted (see Tables S1 and S2, Supporting Information, Test Evaluation). As shown in Figure 2b, this yielded a total of 160 runs of training and 72 runs of testing to complete the evaluation. To better assess the performance, not only the accuracy was considered but also the prediction confidence was assessed for each test.

All the models made false predictions in round 3 except the RFC. While the prediction confidence showed that the LRC seemed less likely to make mistakes, it still did not achieve as high accuracy as the RFC. The ANN model gave correct prediction in 77.8% of the cases with relatively small probability in false predictions. Higher accuracy was achieved by the RFC model with an overall score of 88.9%, leaving only two false prediction outcomes on round 6 and 11. Though RFC and ANN both made wrong classification on round 6, RFC demonstrated a lesser degree of incorrectness (Tables S3–S6, Supporting Information, Prediction probability). Furthermore, the overall score of the single decision tree model was lower than the ensembled random forest, again indicating a better generalization ability of the RFC. In summary, considering the training cost and dataset requirement, the RFC was selected for the initial screening and classified aggregation prediction. Apart from the classification results, it is important to know the true label of each prediction as it can be used to evaluate the algorithm better. The four characteristics that are used to define the measurement metrics of the RFC are summarized on the top row of the figure as a reference with a brief explanation listed below:

TP (True Positive) represents the LC-AuNPs samples which have been properly classified as exhibiting aggregation under the isotropic phase. TN (True Negative) represents the LC-AuNPs samples which have been properly classified as not exhibiting non-aggregation with AuNPs are soluble in the isotropic phase. FP (False Positive) represents the misclassified LC-AuNPs samples that were predicted to exhibit aggregation, but no agglomerates were observed. FN (False Negative) represents the misclassified LC-AuNPs samples that were not predicted to exhibit aggregation, but agglomerates were found.

It was observed that the two false predictions of RFC were FN (round 6) and FP (round 11). Therefore, the RFC demonstrated the best prediction performance with a low false positive rate (FPR) as well as a high true positive rate (TPR) compared with prediction labels made by other models (Figure S4 and Table S7, Supporting Information, Characteristic Labels Calculation). To further validate the robustness of the RFC prediction score, a five-fold cross validation (CV) test was conducted. As shown in Figure 2c, after 5 runs of validation test on each model, an average score could then be obtained to represent its overall performance. With the increasing of training set size, the validation score was generally improved and approached a maximum of 86.4% with 38 samples. This was in line with the bootstrap aggregation algorithm, as only a subset of samples was used for growing the decision tree which led to more diverse subsets due to the increased training set size, providing a minimized out-of-bag (OOB) error rate in testing against the out-of-bag samples.[40,41] As a smaller OOB error rate was directly translated into a higher prediction score in validation runs with unknown samples,[17] which aligned with the RFC performance.

A built-in feature ranking method of the random forest model was adopted to further investigate the best features based on their contribution to the overall Gini coefficient decrease. Figure 2d summarized the ranking results based on different depths of the random forest after the grid search (Figure S5, Supporting Information), where important features could be directly translated into key factors related to the AuNPs aggregation. Since the optimized hyperparameters of the RFC were obtained under the depth of 3 after the random selected test set of 19 rounds (Table S8, Supporting Information), these feature values were taken as reference. Not surprisingly, the ligands ratio of 1-HT to MDD-CBO (X1) played a key role with an importance value of 0.31. However, the reaction temperature (X5) had a more significant contribution with a feature importance of 0.36 in the classified prediction of AuNPs aggregation behavior, indicating the temperature during the ligand exchange could result in different surface arrangements which finally affects the solubility of functionalized AuNPs. Interestingly, the relative amount of ligands (X3), as well as reaction time (X4), also have certain impacts on the prediction accuracy under different depths of the random forest model.

To better visualize the prediction performance of classified results, a confusion matrix with four basic characteristics mentioned above was summarized. As shown in Figure 2e, the optimized model had a high TP rate as well as a relative low FP rate, demonstrating its good performance in classifying the aggregation behavior of AuNPs. As an alternative metric, the F1 Score (also known as F Measure) was used to assess the predictive skill by evaluating its class-wise performance rather than an overall accuracy value.[42] This can be calculated by the following equation (see Supporting Information, Characteristic labels calculation):

$$F1\,Score = \frac{2*Precision*Recall}{Precision + Recall} \quad (1)$$

The precision measures is calculated by taking the TP values and dividing them among both TP and FP values, while the recall measure is calculated similarly by taking the TP values and dividing them among both TP and FN values.[43] The F1 Score of the RFC was 0.869, representing its good performance in both precision and recall, providing a reliable reference for the initial screening phase. By analyzing the prediction outcomes associated with the aggregation probability, a series of threshold values with TPR and FPR were calculated and presented via Receiver Operating Characteristic (ROC) curve. The ROC curve can not only be applied to assess the overall classification performance of one test but also to compare the performance of two or more.[44] For the two-class classification problem, a totally random classifier's ROC points tend toward the diagonal line and divide the ROC space, indicating a theoretical chance level of 50%. A brief

comparison between the LRC and RFC within the ROC space is summarized in Figure 2f. It should be noted that both models showed better prediction expectations as ROC curves were above the chance level under different cut-off values. However, the area under curve (AUC) for the RFC (0.91) was still 15.2% higher than that of the LRC, which was in line with a better prediction accuracy., The ROC curve for perfect classification lies in the upper-left corner or coordinate (0,1) with an AUC of 1.0, meaning no false negatives and no false positives. Compared with the LRC, the ROC curve of the RFC is closer to the upper-left corner, demonstrating better classification results. Further quantification via the Kolmogorov–Smirnov (KS) score was used to assess models; as a higher KS score represents better separation performance between the positive and negative samples.[45] The maximum difference of TPR and FPR at each threshold on the ROC curves was calculated to obtain the respective KS value. The KS score of LRC was 0.667 with a threshold of 70.4%, while a higher KS score of 0.727 for RFC was obtained at the threshold of 67.7% (Figure S6, Supporting Information), exhibiting a better classification performance of RFC. In summary, based on the multi-dimensional evaluation made above, the ensembled model of RFC demonstrated satisfactory performance in the initial classification of AuNPs aggregation behavior, consolidating a good baseline for the next step of investigation.

2.3. In-depth Optimization of Experimental Conditions with Generalized Additive Model

To efficiently control the experimental budgets within the AuNPs optimization, further analysis on experimental factors and their relationships during functionalization needs to be investigated. As an application-independent methodology with a significant reduction on trial numbers,[7–9] Design of Experiment (DoE) was used for guidance in the optimization. As shown in **Figure 3a**, all the factors were listed in a wind map, serving as the basis for screening sources of effects (Detailed upper and lower limits determined by DoE were summarized in Table S9, Supporting Information). Three total steps of DoE including 58 runs of the experiment were part of the iteration for factors screening (see Tables S10–S12, Supporting Information, DoE Analysis), namely the main effects, interaction effects, and high-order effects, respectively. With acceleration from automated synthesis and integrated devices, in situ UV–vis spectra could be obtained with online-BF images simultaneously. Auto processing of these spectroscopy greatly meets the purpose of feature reduction (see Supporting Information, Data Pre-Processing), as the efficiency of extracting half width half maximum (HWHM) and the peak location (λ_{SPR}) for each curve improved drastically compared with manual operation. Since the peak location is a critical feature for describing the aggregation behavior of AuNPs, extracting this value accurately is essential. Furthermore, automated pre-processing not only streamlines data formatting but also enhances model construction and training, ultimately facilitating the optimization of the AuNP synthesis process.

The initial step of DoE phase was to screen the main active effects during the AuNPs functionalization, which were considered to impact the response variables more than expected random variation. A list for the main single effect was evaluated by its significance of p-value to help distinguish the active ones among all factors (Figure S18, Supporting Information), the first round of screening analysis based on the multiple linear regression model (MLR) showed that the ligands ratio (A), rotation speed (B), and the reaction temperature (E) were significant. Herein, MLR was used as the baseline model (M0) serving as a reference for evaluation, which could be described as follows:

$$Y\ (x) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{2}$$

where β_i represents the coefficient of the respective factor x_i, β_0 is the constant which serves as the white noise (also known as the intercept).

Not surprisingly, the fitting performance needed to be improved by investigating other interaction effects. The second round of screening based on the fractional factorial design (FFD) gave a summary analysis of the complex interaction between main effects which were obtained through the initial step. Apart from the similar main effects distribution, Figure 3b showed that multi-interaction effects also have certain extent of impacts, for which large values correspond to greater significance. Since another key factor in the miscibility of the functionalized AuNPs with the LC was the reaction temperature (E), which may be linked to nanoscale phase separation. A random distribution of the two ligands in the ligand shell maximizes the free volume and accessibility to each interaction site, promoting miscibility. However, phase separation negatively impacts both free volume and ligand accessibility, ultimately reducing overall solubility. Experimentally, higher temperatures have been demonstrated to facilitate diffusion of ligands across the NP surface, potentially promoting this phase separation of binary ligand mixtures.[46,47] Besides, specific combination of interaction effects had different significance for responses, which should also be considered during the model construction. As illustrated in Figure 3c, considering the complexity of the main effects combined with multi-interaction effects in the first two steps, a general additive model (GAM) together with DoE screening was integrated to realize the iterated model construction from M0 to M1. One key advantage of adopting this GAM methodology is their ability to model nonlinear data while retaining interpretability, enabling simplified structure compared with the response surface model, therefore greatly improving its versatility.[23] Furthermore, high-order effects among selected factors were also studied with comparison to build the final model (Figures S19 and S20, Supporting Information), with detailed description on their relationships in the supporting information (GAM Construction). Therefore, a simplified format of the GAM model which contains all mentioned features could be written as:

$$Y = \sum\nolimits_{i=1}^{n} A_i + \sum\nolimits_{i,j=1}^{n} A_{ij(i\neq j)} + \sum\nolimits_{i=1}^{m} A_i^m + \varphi \tag{3}$$

where A_i represents the main effects of ligands ratio, reaction speed, and reaction temperature, respectively; A_{ij} are the binary multi-interaction among the active effects; A_i^m is the high-order effects interaction with a limited magnitude parameter to avoid over fitting, φ is a constant accounting for the white noise.

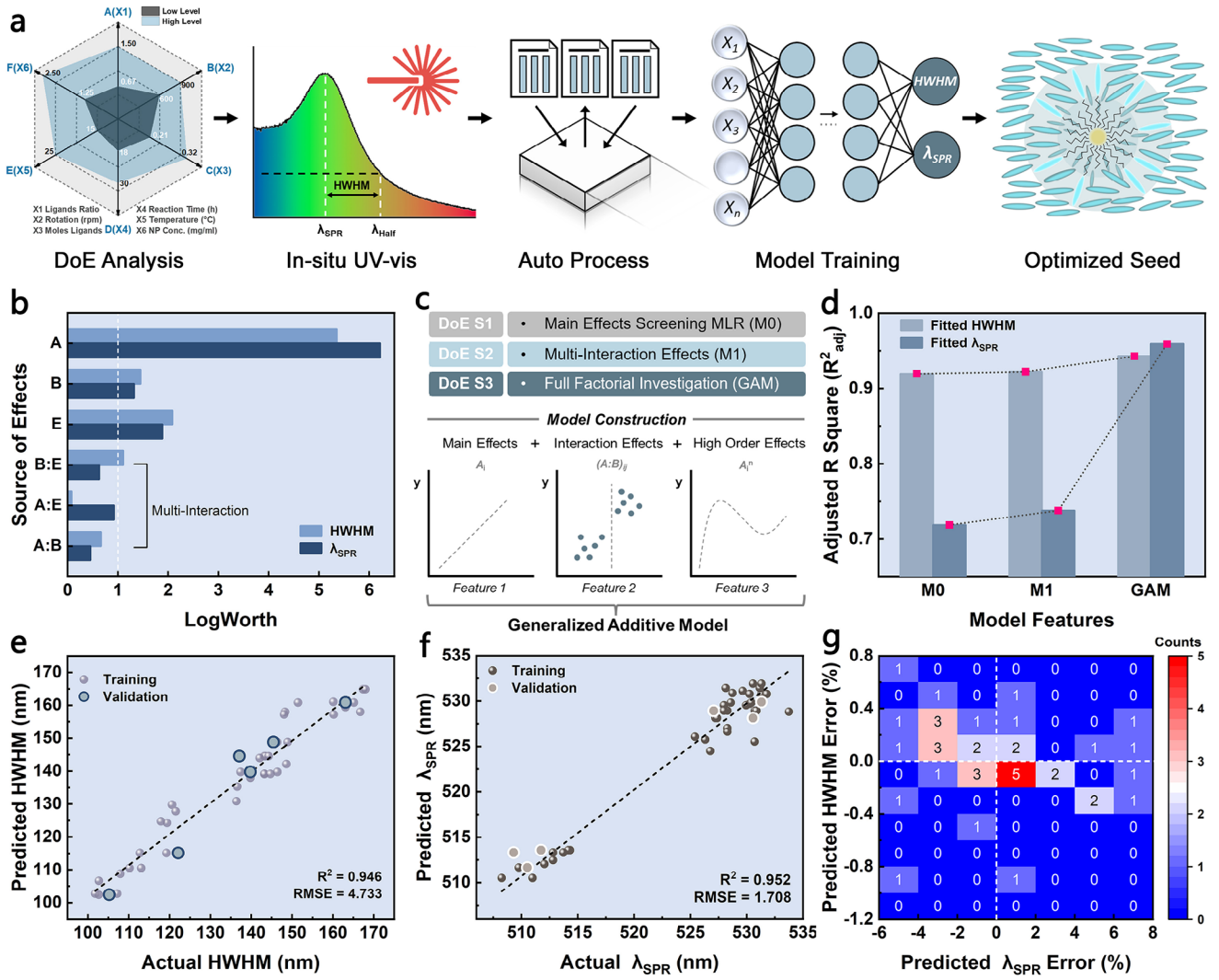

Figure 3. DoE analysis with generalized additive model (GAM) training from UV–vis spectra for optimization. a) Workflow for training the GAM with acceleration from the auto data process to get the optimized NP seed. b) Summarized source of effects from the second phase of DoE analysis, exhibiting the multi-interaction effects for different responses. c) Proposed iterations of the GAM, including DoE analysis, baseline choice, model construction based on screening features. d) Adjusted R square values among different features in the construction of the GAM. e) Predicted HWHM value versus actual HWHM value by training the GAM with the fitting of plots (dashed line) and validation. f) Predicted peak location (λ_{SPR}) versus actual peak location (λ_{SPR}) by training the GAM with the fitting of plots (dashed line) and validation. g) Percentage errors associated with the predicted HWHM and peak location (λ_{SPR}) in the heat map showing the error distribution.

Figure 3d presents the training results for different features based on the FFD datasets from rounds 2 and 3. It is important to note that the iterated model was trained separately for each spectral feature. The adjusted R^2 values for HWHM and peak location were calculated individually to assess the respective performance of the iterated models. The MLR baseline model (M0) could already reach a good prediction for HWHM, but it wasn't accurate enough to describe the response of peak location, as the adjusted R square was not satisfied due to high deviations from the fitting result. Iterated model M1 with multi-interaction effects which considered binary interaction of B and E showed improvement in the two responses compared with the baseline model. However, lack of high-order effects still restricts its interpretation ability. Besides, only one combination of multi-interaction effects may not cover enough features during the training process although their significance may be close to the threshold. Overall, after taking consideration of the potential binary combinations (A:B and A:E) and the second-order effects of A, the final GAM model sufficiently described the data compared with previous models, with adjusted R square for both responses increased, especially the peak location got significantly better fitting outcomes. These results highlight the non-linear dependence between the molar ratio of the two ligands (Factor A) and the characteristics of the SPR peak. Miscibility between the AuNPs and the LC depends on both enthalpic and entropic contributions. Soulé et al presented a model describing this balance wherein the enthalpic effects were directly proportional to the number of interaction sites between the LC and the mesogenic ligands.[48,49] In

contrast, the entropic contribution depended on the free volume available within the ligand shell; as a result, homoligand shells exhibited minimal entropy, with a maximum entropy observed at an intermediate mixture of the two ligands. This non-linear behavior was also confirmed experimentally in previous studies,[34] and is reflected in the quadratic term for the molar ratio of the ligands in Table S13 (Supporting Information) for GAM construction.

After the iterated process mentioned above, the constructed ensembled GAM model was then trained with the entire factorial dataset (excluding one center point) for further evaluation. In general, the prediction results of HWHM showed good fitness with the actual value (Figure 3e) based on the GAM model, with evaluation metrics including the mean square error (RMSE) as well as R square (0.946) confirmed its performance. External validation set also showed a good prediction score with a low RMSE of 4.63, exhibiting a consistent quantile distribution with the residual error (Figure S21, Supporting Information). Similarly, training results for the peak location were also summarized in Figure 3f. Although the predicted values showed a fluctuation tendency toward the higher wavenumber range, the overall performance was still better compared with previous models, with a high R square of 0.952 and low RMSE of 1.708 demonstrating its improved fitness and interpretation. The heat map shown in Figure 3g gave a visualization summary of the GAM model in the percentage prediction error for both responses, with dashed lines serving as the reference of zero. Both the percentage error of prediction followed a normal distribution, with similar quantile result summarized in Figure S22 (Supporting Information). The distribution of prediction errors in the heat map also showed that the GAM model could give more accurate results within the fourth quadrant where it tends to overpredict the peak location (λ_{SPR}) value when underpredicting the HWHM and vice versa. More discrete prediction errors were found for the peak location (λ_{SPR}) value compared with HWHM, especially in the third quadrant of the map. Nonetheless, the GAM model was able to provide prediction error for both responses within ±10%. Since the two-measurement metrics played an important role in describing the aggregation behavior of AuNPs within the LC matrix, optimal combination of experimental factors with description were summarized in Table S14 (Supporting Information) based on the predicted results via the GAM algorithm. Therefore, with the optimized reaction conditions for guidance of automated AuNPs functionalization, the preparation of soluble AuNP seed could be accelerated via the data-driven workflow, which then paved the way for the further exploration of its reversible self-assembly behavior within the LC matrix.

2.4. Validation and Thermocycling of the Optimized AuNPs-LC System

Based on the screening result via the data-driven workflow, AuNPs optimized for solubility in LC matrix (LC-AuNPs) were fabricated for thermocycling validation. As exhibited in **Figure 4a**, the typical cooling dynamics of the LC phase transition and the subsequent diffusion of AuNPs into the nematic phase were recorded through bright field microscopy images with in situ UV–vis spectroscopy. Herein, the LC-AuNPs composite was first heated to 50 °C for 60 min and then cooled to 25 °C at the rate of 1 °C min^{-1}. During this process, the phase transition of the LC-AuNPs composite was observed ≈32.8 °C, which is in line with previous studies that reported on the decrease of phase transition temperature due to the existence of AuNPs.[36,50]

The temperature-induced self-assembly exhibited two stages. In the first 20 s, compartmentalization occurred, creating AuNP-enriched and AuNP-depleted regions, with the enriched areas displaying spinodal-type decomposition (Figure S23, Supporting Information). By 30 s, these patterns evolved into distinct AuNP agglomerates, marking the second stage. As randomly aligned domains from the phase transition interacted with the LC cell margins, uniformly aligned nematic domains began forming post-isotropic-nematic transition. A zoomed-in view of a specific area (dashed box) is shown at the 30s. The growth of these uniformly aligned nematic domains, facilitated by LC cell alignment, led to AuNP segregation in gap regions, resulting in increased AuNP concentration over time (Figure S24, Supporting Information). These findings are consistent with our previous studies on the microscale assembly of AuNPs.[34] In situ UV–vis spectroscopy effectively monitored AuNP behavior during the process. As shown in Figure S25 (Supporting Information), triggering the phase transition process ≈32.5 °C led to a sharp decrease in Abs_{max}, aligning with pattern changes in newly formed nematic domains and indicating rapid partitioning of AuNP agglomerates in less ordered regions. Utilizing an automated analytic toolbox, absorbance values were quantified from numerous spectra during the dynamic process. (Figure S26, Supporting Information).

The reversibility of the assembly process was further explored by the automated toolbox, where spectra features including HWHM and peak intensity were extracted for evaluation. Herein, the LC-AuNPs material was heated up from the room temperature (1 °C min^{-1}), further maintained at the terminal temperature of 40 °C to validate the dissociation of the agglomerates when heated from the nematic to the isotropic phase.

As exhibited in Figure 4b, this reversible self-assembly during heating process had a significant transformation during the first 30 s, as the patterns changed from isolated AuNP aggregations to continuous merged margins during the phase transition period. With increasing temperature (Figure 4c), the AuNP agglomerates gradually redissolved into the LC matrix. Similarly, the spectra results did not exhibit any difference under the nematic phase until the phase transition process was triggered ≈33 °C (Figure 4d). As the absorbance intensity increased drastically then gradually dropped down until reached a steady state, with detailed results in Figure S27 (Supporting Information). HWHM was then extracted from the UV–vis spectra via our automated analytic toolbox to accelerate the solubility evaluation of the thermocycling process. Figure 4e shows the variation of HWHM values during a typical cycle of reversible test. The margin of clusters could be classified as different phase regions where the value of dots tended to accumulate in the higher space for nematic phase and vice versa. Besides, a broadening ≈10–15 nm of HWHM values were observed from the isotropic region to nematic region, indicating the occurrence of AuNPs agglomerates after the phase transition period which was in line with previous studies which reported that the aggregation would result in a broadening of the SPR peak in the absorbance spectra.[51,52]

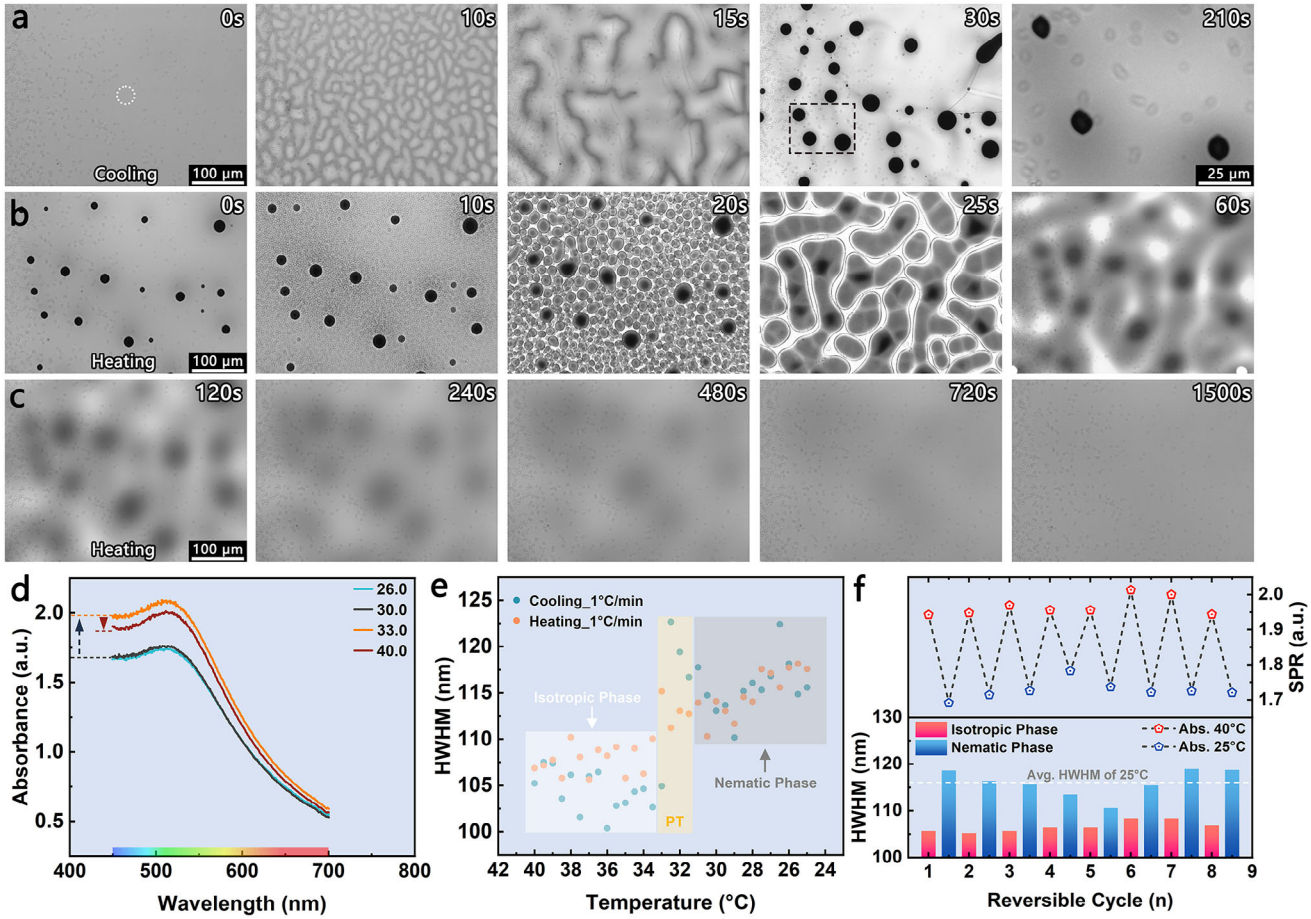

Figure 4. The reversibility of the AuNPs agglomerates with self-assembly features during thermocycling. a) A typical kinetic series of bright field microscopy images of the LC-AuNPs composite during the cooling process, the images were captured from 33.4 °C (1 °C min^{-1}). b,c) BF images recorded the reversible behavior of LC-AuNPs composite heated from 32.4 °C (1 °C min^{-1}), a terminal temperature of 40.0 °C was maintained for an additional 20 min to exhibit its reversibility to the initial state. d) In situ UV–vis absorbance spectra were recorded during the heating process, exhibiting the typical increase of absorbance value. e) Half width at half maximum (HWHM) was extracted during the cooling-heating cycle between the temperature range of 25 to 40 °C, including the isotropic and nematic phase. f) Variation of SPR peak value and HWHM as a function of reversible cooling-heating cycle.

A series of cycling aging tests were conducted to test the robustness of reversibility. Typically, the composite was first heated to 40 °C and maintained for 15 min and then cooled to 25 °C (1 °C min^{-1}), an extra duration of 15 min was maintained once it reached the terminal point. Based on the fast feature extraction toolbox, the maximum absorbance (SPR peak value) under different phases as well as the HWHM values were captured efficiently and then summarized. As shown in Figure 4f, the typical responses from the absorbance spectra didn't exhibit any significant difference after 8 rounds of reversible cycling test. A slight increase of SPR peak value was observed during the 6th cycling test but soon recovered to the initial state. Meanwhile, the HWHM values exhibited robust performance without becoming broader even after the whole 8 cycling rounds, especially under the isotropic phase. It should be noted that the final responses even had the same value as the initial state, indicating the good reversibility and high robustness for the temperature-induced self-assembly behavior, which could be important in the periodicity application of microstructure features. Thus, compared with manually locating the data point within the curve for calculating spectral features, the downstream analytic toolbox enabled high-throughput analysis (HTA) based on large amounts of spectra curves became feasible and efficient during the thermocycling of the phase transition process.

2.5. CV-Based Methodology for Quantified Analysis of AuNPs-LC System

In order to quantify the reversibility of agglomerate formation during the self-assembly, i.e., whether the assembly would finally fully dissolve upon heating to the isotropic state at the required temperature, referenced information entropy was adopted to assess the LC-AuNPs system. The information entropy (also known as Shannon's entropy) could serve as the basis for evaluating the uniformity of a system,[53,54] thus it could be used here for statistical analysis of pixel values in the images during the self-assembly period. A previous study introduced this approach successfully.[55] However, it required manual conversion of each image into grids and lacked the capability for batch processing

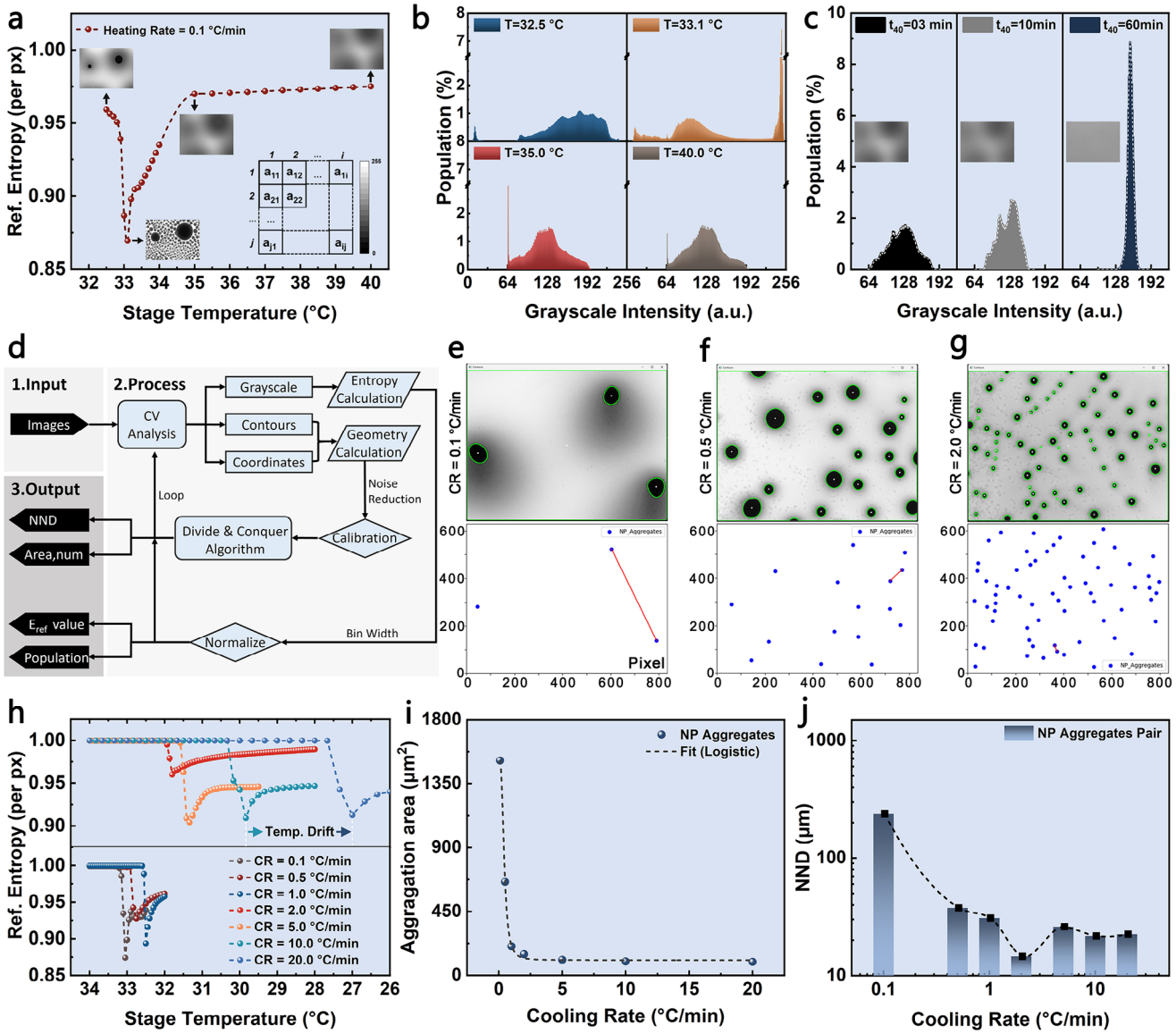

Figure 5. Computer vision based (CV) methodology for the image analysis of self-assembly behavior. a) Information entropy of the LC-AuNPs composite during its heating from 32.5 °C with key frames extracted. b) Grayscale intensity distribution analysis at different temperatures. c) Grayscale intensity distribution at the terminal temperature (40.0 °C), exhibiting its fully reversible features with a tendency to the normal distribution. d) IPO structure of the analysis toolbox based on the CV methodology. e–g) Image recognition for NP aggregates under different cooling rates, with visualization results of the closet pair distance shown by Matplotlib. h) Application of our toolbox for efficient analysis of information entropy at different cooling rates under high resolution. i) Aggregation area results obtained from the toolbox. j) NND results obtained from the Divide & Conquer algorithm exhibiting the reversible aggregate behavior of AuNPs.

of high-resolution files. Additionally, due to manual limitations, it couldn't display the dynamic range distribution of grayscale intensity within the images. To tackle the automated batch processing and further reveal the bottom-level reasons for information entropy variation, the high-throughput analysis (HTA) of the greyscale intensity in pixel level within images is required though challenging. Regarding its analytic unit was the pixel value, computer vision (CV) methodology with imaging libraries of *Pillow* and *cv2* was used for the algorithm development.

For a perfectly homogenous system, each pixel should have an equal grayscale value, resulting in a entropy per pixel getting close to 1. As the variation in absorbance for each pixel increases, grayscale values show variation (from 0 to 255), thus the referenced entropy will reduce. As shown in **Figure** 5a, the quantified referenced entropy (E_{Ref}) based on pixel-level HTA during the dynamic phase transition process was calculated via the automated toolbox, along with fast extraction of key frames among GB-level files. Detailed description can be found in the supporting information (CV-based Algorithm for Entropy Calculation).

It can be found that during the typical reversibility test (heating rate: 0.1 °C min^{-1}) the referenced entropy of LC-AuNPs system

was ≈0.959 before the phase transition period as the existing NP agglomerates formed during the nematic phase affected the uniformity of the grayscale distribution. However, the entropy value dropped drastically to 0.867 per pixel at the onset of the transition ≈33 °C, which was in line with the key frame as numerous small aggregations could be observed during this period. After that, the referenced entropy showed a continuous logistic increasing tendency with the diffusion of AuNPs into the LC matrix at the isotropic state, thus demonstrating homogeneity.

To further reveal the bottom-level reasons for the variation, a detailed statistical HTA of grayscale information for all pixels (resolution: 1936 × 1456p, N = 2.82 × 10^6) during the dynamic process was summarized in Figure 5b, which showed the underlying grayscale intensity distribution for those key frames. The distribution of the grayscale intensity before the phase transition period showed a broad trend across the range from 0 to 255, especially at the onset ≈33 °C, as both the population in the high-intensity range (224 to 255) and low-intensity range (0 to 64) increased significantly, indicating a higher degree of disorder for the grayscale intensity of all pixels, which was not only consistent with the related keyframe but also the drastic decrease of the referenced entropy.

Although further distribution after 35 °C showed little difference as the diffusion of AuNPs carried on continuously, the grayscale intensity tended to converge into a narrower range from 64 to 192. It should be noted that the final distribution of grayscale intensity(t_{40} = 60 min) exhibited in Figure 5c even converged within a smaller range of 32 (from 128 to 160) and followed a Gaussian distribution after an extra period of 60 min at the terminal temperature of 40 °C, indicating a quite uniform state compared with the initial phase transition period, also in line with the homogenous BF image. With the help of CV-based methodology, the automated batch processing of quantification entropy could not only greatly improve the working efficiency by analyzing GB-level images but also demonstrate the underlying causes from the statistical view in analyzing the reversible self-assembly behavior of AuNPs via the HTA method in pixel level.

Apart from the global metrics of reversibility based on the referenced entropy, further investigation via pattern recognition was also conducted to give a specific analysis of individual agglomerate behavior. Since the size and distribution of AuNPs aggregation would also be affected by the temperate rate during the initial phase decomposition (Figure S28, Supporting Information), the distance between each agglomerate is important in describing its assembly behavior. Therefore, geometry features calculation, as well as Nearest Neighbor Distance (NND) value were included in the evaluation metrics for the image recognition algorithm. Conventional analytic methods often require manual marking within each image for calibration of length scale, this time-consuming and repetitive approach greatly restricts the investigation among agglomerate systems. To tackle this challenging problem, pattern recognition was also required in the downstream analytic process for image files.

The summarized Input-Process-Output (IPO) model in Figure 5d exhibits the workflow of our CV-based toolbox, the algorithm could automatically realize the pattern recognition of AuNPs agglomerates on CV analysis with batch of images as the input source. Contours and coordinates of agglomerates were then calculated and plotted for geometric features, along with the Divide & Conquer algorithm served as the solver for the NND value (Detailed description can be found in Supporting Information, Image Recognition for NND algorithm). A series of different cooling rates were employed during the analysis of temperature-induced self-assembly behavior of NP arrays. Figure 5e showed the recognized contours (green margins) of the BF image captured under the nematic phase at 26 °C with a cooling rate (CR) of 0.1 °C min^{-1}, with related geometric centers extracted into the coordinate system scaled from the input image. Under the low CR value (0.1 °C min^{-1}), NP agglomerate tends to disperse separately and are further spaced from each other, with a relatively high NND value of 237.3 μm as well as a bigger aggregation area of 1512.4 $μm^2$ obtained from the CV analysis. The closest pair of NP agglomerates was marked with a red line in the pattern recognition figure for guidance. With the increase of the cooling rate (Figure 5f), the size of AuNP agglomerates tends to reduce while the number of aggregates increases drastically, especially in Figure 5g, as analyzing such a typical image is tough and time-consuming for manual processing let alone numerous BF images during the dynamic process. Accelerated by the analytic toolbox, solving such kind of problems became feasible as a clear comparison could be made between Figure 5g,e, where the marked closest pair was quite close to each other at the cooling rate of 2.0 °C min^{-1}, indicating a significant lower NND value along with more aggregates under the same scope.

Taking advantage of the batch processing performance, thousands of BF images (10 GB-level) under different cooling rates via high-throughput analysis were realized, with summarized curves in Figure 5h for comparison. Similar to the reversibility test mentioned above, the referenced entropy showed a converse variation tendency. However, a temperature drift of the minimum E_{ref} was observed (especially under higher cooling rates), this was mainly caused by the interfacial thermal resistance (ITR) between the LC cell and the temperature control stage, as a higher cooling rate means more heat flux among the through-plane direction among the matting surfaces where interfacial micro-gaps are filled with nonflowing air with very low thermal conductivity (≈0.026 W m^{-1} K^{-1}),[56] a thermal interface material (TIM) might be considered. Herein, considering the application scenario including the optical path, a transparent TIM along with matched refractivity might be needed for future investigation. Although a lower CR value means more accurate results, significant time consumption was also generated. Thus, to reach a balance between the accuracy and efficiency, the optimal cooling rate of 1.0 °C min^{-1} was selected during the previous experimentation. Besides, the average aggregation area under different cooling rates were also obtained via the CV-based toolbox. As shown in Figure 5i, the average area values fitted well with the logistic trend decrease, remaining a relative stable level ≈100 $μm^2$. Similar trend of results for the NND values was also found in Figure 5j, where the distance for the closest pair of agglomerates dropped drastically with the increase of cooling rate over 0.5 °C min^{-1}, further exhibiting a relative low level under 40 μm. Overall, with the help of CV-based toolbox, quantified evaluation for the LC-AuNPs was realized with the high-throughput analysis in the downstream workflow where the aggregation behavior of NP agglomerates showed some regular patterns, as a higher cooling rate (5 to 20 °C min^{-1}) would cause more aggregates along with smaller size and spacing, while the average

area and NND value would follow a logistic variation trend with the cooling rate increase. Thanks to the image recognition algorithm based on computer vision methodology, large number of imaging files could be processed efficiently with our CV-based toolbox while generating useful data library for quantification analysis to better understand the aggregation behavior of AuNPs during the self-assembly process. More importantly, the outlined workflow with acceleration from CV-based approach has the potential to be implemented in other nanomaterial systems (e.g., flow chemistry for iron oxide NPs, IONP), where continuous production was generated for rapid analysis about its nucleation or growth.

3. Conclusion

In this study, automated functionalization of AuNPs via the OT-2 platform, combined with a downstream analytic toolbox provided the basis for an efficient data-driven workflow. A random forest model was used to give classified prediction on AuNP aggregates behavior. The ensembled RFC model could improve the feature identification of agglomerates to facilitate the initial screening of given factors. The implementation of the RFC workflow revealed that on relatively small training sets of 40 groups, high accuracy with prediction confidence could be achieved. More importantly, feature importance during the construction of the random forest also helps explain the experimental factors which can be of great value in determining the aggregation behavior of AuNPs. Besides, with the help of DoE screening, the total experimental budget could be controlled under a relatively small scale, which accelerated the AuNP optimization procedure with iterated construction of the GAM algorithm. The combined workflow implied that the analysis of AuNPs aggregation behavior does not require an extensive set of synthetic factors to cover the full optimization space. Instead, a smaller and rational experimental budget via DoE can be used to train the model for the prediction of AuNP aggregates behavior along with its spectral characteristics. The optimized soluble AuNPs were efficiently evaluated under thermocycling by the extracted spectral features from the automated analytic toolbox. Combined with CV-based methodology, batch processing of GB-level images was realized for analyzing referenced entropy of the LC-AuNPs system during the self-assembly process. Further revealing the underlying causes of the variation from the greyscale intensity distribution on pixel-level via high-throughput analysis. Pattern recognition algorithms were integrated in analyzing AuNPs agglomerates for quantified NND values, which followed a logistic variation trend with the cooling rate increase. Accelerated by the downstream analytic toolbox, the automated data-driven workflow can aid in the design and efficacy improvement of soft nanomaterial systems that target functionalized NPs for future applications toward programmable and reconfigurable materials.

4. Experimental Section

Materials: The gold salt, hydrogen tetrachloroaurate (III) hydrate ($HAuCl_4 \cdot 3H_2O$, 95%) was purchased from Fisher scientific (UK). The dichloromethane (DCM), t-butylamine-borane complex (tBAB, 97%), and oleylamine (OAm, with C18 content: 80%–90%) were supplied by Sigma-Aldrich. The absolute ethyl alcohol (EtOH) was obtained from Fisher scientific (UK). The n-Octane (99%) and 1-hexanethiol ligand (1-HT, 96%) were purchased from thermos scientific. The long mesogenic ligand, 4′-(12-Mercaptododecyloxy) Biphenyl-4-Carbonitrile (MDD-CBO) was synthesized following the adapted methodology of Milette et al.[34] Oleylamine-capped gold nanoparticles were obtained via a modular process, detailed synthetic and analytic procedures could be found in the previous study.[57] The liquid crystal of 4-cyano-4′-pentylbiphenyl (5-CB, 99.8%) came from Synthon chemicals with a phase transition temperature of 34 °C. All reagents were commercially purchased and used as received without further purification.

Hardware Platform: Opentrons (OT-2) pipetting robot together with the P1000 Gen2 and P50 Gen1 were used for liquid handling. An INHECO thermoshake AC (TSAC) with single temperature control unit (STC) was deployed to realize the automated reaction with rotary mixing and temperature control. The program for controlling the home-made smart centrifuge was written in C++ via Arduino IDE. Customized labware and accessories were constructed from a range of 3D printed (SLA), CNC, and laser cut components. Detailed information and the assembly instruction were summarized in the Supporting Information.

Automated AuNPs Functionalization: The ligand exchange (LE) process for automatic protocols involves four steps: 1) solution mix, 2) reaction quench, 3) solvent removal, and 4) centrifugation with purification. The entire process was realized through the combined application of the TSAC module and customized labware within the OT-2 platform. During the LE protocols, a customized well plate was used for the temporary placement of reaction tubes. Specific reaction conditions were determined by the DoE and subsequent experimentation. A detailed description for the automated protocols along with proposed procedures can be found in the Supporting Information.

Computational Module: For model training, classification of the AuNP aggregation behavior was achieved through the random forest model, which was trained by the five-fold cross-validation approach. The RF classification model was implemented using the open sourced *scikit-learn* library,[58] with version of the CART algorithm for machine learning.[59] Besides, for classification models working on correlated neighbor factors, the optimal feature-to-sample ratio could be approximated by $k = \sqrt{m}$ selected in the final model, where m represented the total numbers of features.[60,61] Therefore, at every node level in the RF, optimized features via the bootstrap sampling is considered for partitioning with an ensembled result for each decision tree summarized.[61,62] The criterion for partitioning the sample space (splitting principle) at each node level was evaluated according to the Gini coefficient. Detailed descriptions of the algorithm about the feature selection, model training, and validation can be found in the Supporting Information, along with the summary of training parameters of the classifier with evaluation metrics (Tables S7 and S8, Figures S2 and S3, Supporting Information). Automated processing of the UV–vis data focused on extracting key features, such as the half width half maximum (HWHM) and the peak location (λ_{SPR}) from the spectra information. The background signal was removed from the original spectra using the proprietary OceanView software provided with the UV–vis module (Ocean Optics, QE Pro-65000). Since the in situ UV–vis measurements were conducted using a liquid crystal cell, signal calibration was performed for both light-field and dark-field backgrounds to ensure consistency and minimize artefacts. Manually obtaining the peak location and width from the spectra was time-consuming and could result in deviations due to human error or bias. The spectra of the LC-AuNP composite were obtained from the visible range to near IR, with a sampling interval of 0.78 nm, representing over 2000 data points. Therefore, the auto process and analysis of the UV–vis data could not only simplify the spectra with key features (see Supporting Information, Data Pre-Processing) but also greatly help format the data library for machine learning. The regression model for predicting spectral features of LC-AuNPs composites was based on the multivariate linear regression (MLR) model, with iterated construction of generalized additive model (GAM) algorithm. Three rounds of DoE analysis were carried out with iteration to build the GAM model including features of the main effects, multi-interaction effects, and high-order effects to obtain an optimal fitting performance. Detailed description of the model

construction can be found in the Supporting Information, together with the summary for the residual quantile distribution of different responses (Figure S22, Supporting Information).

Image Recognition: Computer vision (CV) analysis for entropy was mainly based on the *Pillow* imaging library (PIL), where the BF images were first converted into the grayscale format for array analysis with statistic counts. The sampling step was set as 1 s, as shorter imaging intervals (means high frames per second) resulted in undue computational costs due to the high resolution of the pictures. The referenced entropy calculation algorithm for NP aggregations reversibility was based on the gray value distribution from each pixel of the input image, with further intensity calculation for summarized information entropy as well as normalization to get the final referenced entropy. Detailed information can be found in the Section S13 (Supporting Information). Besides, *cv2* from the *OpenCV* library was used for the pattern recognition during the processing of input images,[63] aiming to get contours information of agglomerates with a general step of grayscale conversion followed by a Gaussian smoothing. Thresholding was then applied to create a binary-based library of the pixel information as well as the geometric features of each agglomerate. Nearest Neighbor Distance (NND) calculation was based on the Divide & Conquer algorithm to balance accuracy and efficiency, with a detailed description along with comparison of brute-force search approach summarized in Supporting Information (Image Recognition for NND algorithm).

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Yueyang Gao acknowledges the joint research funding provided by the UCL Dean's Prize and the China Scholarship Council for the postgraduate research scholarship in the Faculty of Engineering Sciences. The authors thank the UCL Design Workshop for part of the manufacturing work. Furthermore, Yueyang Gao wants to thank Zeyu Liao and Tingchen Zhao's friendship and support during the research and growth. Yueyang Gao is especially grateful to his parents, Yuan Gao and Dongxian Gu, who have always given encouragement (like red bean buns) and companionship in everyday life.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

automated workflow, computer vision analysis, gold nanoparticles, machine learning, reversible self-assembly

Received: January 17, 2025
Revised: March 4, 2025
Published online:

[1] Y. Ouyang, M. Fadeev, P. Zhang, R. Carmieli, J. Li, Y. S. Sohn, O. Karmi, R. Nechushtai, E. Pikarsky, C. Fan, I. Willner, *ACS Nano.* **2022**, *16*, 18232.
[2] Z. Guo, G. Yu, Z. Zhang, Y. Han, G. Guan, W. Yang, M.-Y. Han, *Adv. Mater.* **2023**, *35*, 2206700.
[3] G. A. Vinnacombe-Willson, Y. Conti, S. J. Jonas, P. S. Weiss, A. Mihi, L. Scarabelli, *Adv. Mater.* **2022**, *34*, 2205330.
[4] P. D. Howes, R. Chandrawati, M. M. Stevens, *Science.* **2014**, *346*, 1247390.
[5] J. Guo, L. Liu, B. Bian, J. Wang, X. Zhao, Y. Zhang, Y. Yan, *Adv. Funct. Mater.* **2023**, *33*, 2212666.
[6] K. Higgins, S. M. Valleti, M. Ziatdinov, S. V. Kalinin, M. Ahmadi, *ACS Energy Lett.* **2020**, *5*, 3426.
[7] O. Pereira, M. Ruth, D. Gerbig, R. C. Wende, P. R. Schreiner, *J. Am. Chem. Soc.* **2024**, *146*, 14576.
[8] S. Alonso-Gil, A. Males, P. Z. Fernandes, S. J. Williams, G. J. Davies, C. Rovira, *J. Am. Chem. Soc.* **2017**, *139*, 1085.
[9] V. Nori, A. Sinibaldi, F. Pesciaioli, A. Carlone, *Synthesis.* **2022**, *54*, 4246.
[10] V. Czitrom, *Am. Stat.* **1999**, *53*, 126.
[11] F. Yates, *Biometrics.* **1964**, *20*, 307.
[12] R. Kaushik, P. Desreumaux, J.-B. Mouret, *Front. Rob. AI.* **2020**, *6*, 151.
[13] A. D. Clayton, J. A. Manson, C. J. Taylor, T. W. Chamberlain, B. A. Taylor, G. Clemens, R. A. Bourne, *React. Chem. Eng.* **2019**, *4*, 1545.
[14] J. P. Janet, H. J. Kulik, *Machine Learning in Chemistry*, American Chemical Society, Washington DC **2020**.
[15] M. S. Brieuc, C. D. Waters, D. P. Drinan, K. A. Naish, *Mol. Ecol. Resour.* **2018**, *18*, 755.
[16] J. Riedel, M. Lettow, M. Grabarics, M. Götze, R. L. Miller, G.-J. Boons, G. Meijer, G. von Helden, G. P. Szekeres, K. Pagel, *J. Am. Chem. Soc.* **2023**, *145*, 7859.
[17] L. Breiman, *Mach. Learn.* **2001**, *45*, 5.
[18] V. Zivanovic, S. Seifert, D. Drescher, P. Schrade, S. Werner, P. Guttmann, G. P. Szekeres, S. Bachmann, G. Schneider, C. Arenz, *ACS Nano.* **2019**, *13*, 9363.
[19] B. R. Smith, K. M. Ashton, A. Brodbelt, T. Dawson, M. D. Jenkinson, N. T. Hunt, D. S. Palmer, M. J. Baker, *Analyst.* **2016**, *141*, 3668.
[20] S. Link, M. A. El-Sayed, *J. Phys. Chem. B.* **1999**, *103*, 4212.
[21] S. K. Ghosh, T. Pal, *Chem. Rev.* **2007**, *107*, 4797.
[22] T. J. Hastie, *Statistical Models*, Routledge, Milton Park **2017**, p. 249.
[23] A. M. Zuranski, S. S. Gandhi, A. G. Doyle, *J. Am. Chem. Soc.* **2023**, *145*, 7898.
[24] T. Kato, J. Uchida, T. Ichikawa, T. Sakamoto, *Angew. Chem., Int. Ed.* **2018**, *57*, 4355.
[25] L. Wang, Q. Li, *Adv. Funct. Mater.* **2016**, *26*, 10.
[26] J. P. F. Lagerwall, G. Scalia, *Curr. Appl. Phys.* **2012**, *12*, 1387.
[27] C. Blanc, D. Coursault, E. Lacaze, *Liq. Cryst. Rev.* **2013**, *1*, 83.
[28] I. I. Smalyukh, *Annu. Rev. Condens. Matter Phys.* **2018**, *9*, 207.
[29] M. A. Gharbi, S. Manet, J. Lhermitte, S. Brown, J. Milette, V. Toader, M. Sutton, L. Reven, *ACS Nano.* **2016**, *10*, 3410.
[30] K. Araya, K. Iwasaki, *Mol. Cryst. Liq. Cryst.* **2003**, *392*, 49.
[31] M. Mitov, C. Portet, C. Bourgerette, E. Snoeck, M. Verelst, *Nat. Mater.* **2002**, *1*, 229.
[32] E. Lee, Y. Xia, R. C. Ferrier Jr, H.-N. Kim, M. A. Gharbi, K. J. Stebe, R. D. Kamien, R. J. Composto, S. Yang, *Adv. Mater.* **2016**, *28*, 2731.
[33] H. Qi, B. Kinkead, V. M. Marx, H. R. Zhang, T. Hegmann, *ChemPhysChem.* **2009**, *10*, 1211.
[34] J. Milette, V. Toader, L. Reven, R. B. Lennox, *J. Mater. Chem.* **2011**, *21*, 9043.
[35] J. Milette, S. J. Cowling, V. Toader, C. Lavigne, I. M. Saez, R. Bruce Lennox, J. W. Goodby, L. Reven, *Soft Matter.* **2012**, *8*, 173.
[36] J. Milette, V. Toader, E. R. Soulé, R. B. Lennox, A. D. Rey, L. Reven, *Langmuir.* **2013**, *29*, 1258.
[37] P. Dutta, T.-Y. Su, A.-Y. Fu, M.-C. Chang, Y.-J. Guo, I. J. Tsai, P.-K. Wei, Y.-S. Chang, C.-Y. Lin, Y.-J. Fan, *Chem. Eng. J.* **2022**, *434*, 133864.
[38] Y. Gao, A. Redfearn, S. Dawes, J. Shi, A. Taylor, H. Wurdemann, S. Guldin, *J. Open Hardware.* **2024**, *8*, 14.
[39] Y. Yang, L. A. Serrano, S. Guldin, *Langmuir.* **2018**, *34*, 6820.

[40] A. Géron, *Hands-On Machine Learning with Scikit-Learn Keras, and TensorFlow*, 3rd ed., O'Reilly Media, Inc., Sebastopol, CA **2022**.
[41] C. Zhang, Y. Ma, *Ensemble Machine Learning*, Springer, New York **2012**.
[42] P. Singh, N. Singh, K. K. Singh, A. Singh, in *Machine Learning and the Internet of Medical Things in Healthcare*, (Eds: K. K. Singh, M. Elhoseny, A. Singh, A. A. Elngar), Academic Press, Cambridge **2021**, p. 89.
[43] S. Uddin, I. Haque, H. Lu, M. A. Moni, E. Gide, *Sci. Rep.* **2022**, *12*, 6256.
[44] F. S. Nahm, *Korean J. Anesthesiol.* **2022**, *75*, 25.
[45] G. Zeng, *Commun. Stat. – Simulat. Comput.* **2017**, *46*, 7744.
[46] P. Ionita, A. Volkov, G. Jeschke, V. Chechik, *Anal. Chem.* **2008**, *80*, 95.
[47] M. Schmitt, J. Zhang, J. Lee, B. Lee, X. Ning, R. Zhang, A. Karim, R. F. Davis, K. Matyjaszewski, M. R. Bockstaller, *Sci. Adv.* **2016**, *2*, 1601484.
[48] E. R. Soulé, J. Milette, L. Reven, A. D. Rey, *Soft Matter.* **2012**, *8*, 2860.
[49] E. R. Soulé, C. E. Hoppe, J. Borrajo, R. J. J. Williams, *Ind. Eng. Chem. Res.* **2010**, *49*, 7008.
[50] M. V. Gorkunov, G. A. Shandryuk, A. M. Shatalova, I. Y. Kutergina, A. S. Merekalov, Y. V. Kudryavtsev, R. V. Talroze, M. A. Osipov, *Soft Matter.* **2013**, *9*, 3578.
[51] J. Zhang, H. Liu, Z. Wang, N. Ming, *Adv. Funct. Mater.* **2007**, *17*, 3295.
[52] V. Amendola, M. Meneghetti, *J. Phys. Chem. C.* **2009**, *113*, 4277.
[53] D. Y. Tsai, Y. Lee, E. Matsuyama, *J. Digit. Imag.* **2008**, *21*, 338.
[54] C. E. Shannon, *The Bell Syst. Techn. J.* **1948**, *27*, 379.
[55] N. M. Fhionnlaoich, S. Guldin, *Chem. Mater.* **2020**, *32*, 3701.
[56] Y. Gao, D. Bao, M. Zhang, Y. Cui, F. Xu, X. Shen, Y. Zhu, H. Wang, *Small.* **2021**, *18*, 2105567.
[57] N. M. Fhionnlaoich, S. Schrettl, N. B. Tito, Y. Yang, M. Nair, L. A. Serrano, K. Harkness, P. J. Silva, H. Frauenrath, F. Serra, W. C. Carter, F. Stellacci, S. Guldin, *ACS Nano.* **2023**, *17*, 9906.
[58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, RandomForestClassifier, **2020**.
[60] J. Hua, Z. Xiong, J. Lowey, E. Suh, E. R. Dougherty, *Bioinformatics.* **2005**, *21*, 1509.
[61] A. Liaw, M. Wiener, *R. News.* **2002**, *2*, 18.
[62] T. Hastie, R. Tibshirani, J. H. Friedman, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York **2009**.
[63] K. Pulli, A. Baksheev, K. Kornyakov, V. Eruhimov, *Commun. ACM.* **2012**, *55*, 61.