Nano Research

Open Access

Research Article

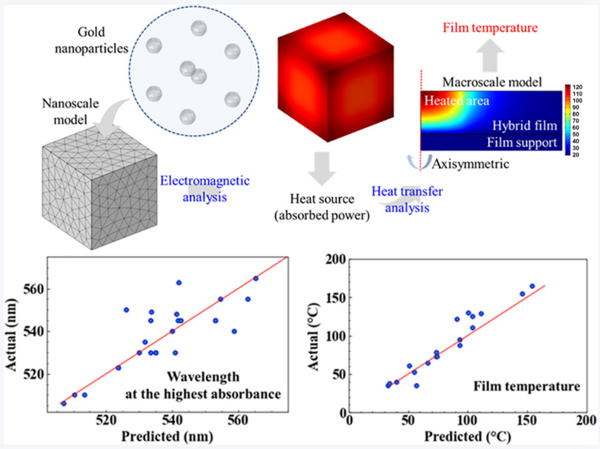
Machine learning approach for predicting optical and photothermal properties of gold nanoparticle/polymer hybrid films: Effect of synthetic data

Yi Je Cho[1,2], Harrison Chaney[1], and Kathy Lu[1,3] ✉

[1] Department of Materials Science and Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[2] Department of Advanced Materials Science and Engineering, Sunchon National University, Suncheon-si, Jeollanam-do 57922, Republic of Korea
[3] Department of Mechanical and Materials Engineering, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Cite this article: *Nano Research*, **2025**, *18*, 94907216. https://doi.org/10.26599/NR.2025.94907216

ABSTRACT: Emerging machine learning (ML) approaches have been adopted in various material systems to predict novel properties with the assistance of the corresponding large datasets. For new materials, however, collecting sufficient data points for model training is not feasible, which is the case for gold nanoparticle/polymer hybrid films. In this study, an ML approach coupled with finite element modeling was proposed for predicting the optical and photothermal properties of gold nanoparticle/polymer hybrid films. Experimental datasets of the optical and photothermal properties were built using results from the literature. Then, finite element analyses were conducted to generate synthetic data to satisfy the quality and quantity of the data required for training models. Correlation analysis and model training were performed using the datasets with and without synthetic data to evaluate their effects on predicting the performance of the ML models. The relative importance of features to targets (properties) was evaluated by correlation analysis. ML models with high accuracy were obtained by training various models from conventional to newly developed algorithms. Advantages, weaknesses, and improvement of the synthetic data addition were discussed. The proposed workflow and framework offer reliable prediction of optical and photothermal properties over different combinations of gold nanoparticles and polymer matrices, which can be extended to include more features related to processing parameters and microstructures.



KEYWORDS: machine learning, gold nanoparticle, polymer hybrid film, optical property, photothermal property, finite element modeling

1 Introduction

Metal and oxide nanoparticles (NPs) have been incorporated in polymer films, namely NP/polymer hybrids, to facilitate self-healing, sustain the performance of the hybrids, and extend the lifetime of devices used in electronic, optical, and/or magnetic applications [1–5]. Each NP with photothermal or superparamagnetic properties acts as an individual heat source under an optical or magnetic field to heat the polymer matrices to temperatures higher than glass transition temperatures. In particular, gold and silver NPs with photothermal properties have the benefit of healing local defects and cracks in devices simply by selective wavelength light irradiation, which is advantageous in preventing damage to other sensitive components [4–10].

The photothermal properties of the hybrid films are dependent on the energy absorption efficiency of the NPs and incident light, which allows tunable properties. However, the relationships of the optical and photothermal properties with incident light require more in-depth understanding. The photothermal efficiency is influenced by the size [9, 11], shape [11], dispersion [5, 12], and composition [5, 7, 13] of the NPs in the polymer matrices. For

Received: October 27, 2024; **Revised:** December 21, 2024
Accepted: December 25, 2024

✉Address correspondence to klu@uab.edu

instance, an NP diameter larger than 2 nm enables 100% light-to-heat conversion [9]. Spherical NPs show lower wavelengths at maximum extinction compared to rod-shaped ones [11]. Agglomeration of the NPs decreases the photothermal efficiency [5]. Higher NP contents in the hybrid films increase the maximum heating temperature of the films as the number of heat sources increases [5, 13]. Incident light can have different combinations of wavelength, irradiation time, intensity, and polarization. Long irradiation time at a wavelength of the maximum absorption leads to maximum temperature increase in the hybrid films [14–16].

To date, effects of microstructural characteristics (size, shape, dispersion, and composition of NPs and type of polymer matrices) and processing parameters (laser intensity and wavelength) on the optical and photothermal properties of the self-healing hybrid films have been investigated either experimentally, theoretically, or numerically. Although an understanding of the relative importance is crucial in designing new hybrid films, studies on the relative importance of these parameters to the properties are scarce. Analyzing the relative importance requires various combinations of experiments and simulations for synthesizing the materials and testing/predicting the properties, which is time-consuming and resource-intensive.

Recent developments in materials informatics and applications of correlation analysis and machine learning (ML) techniques have provided a new approach to understanding hidden relationships between material characteristics and processing conditions (features) and properties (targets) [17–19]. Correlation analysis gives rankings of the involved features and is highly correlated to a target [18–20]. For carefully curated data, the corresponding outputs can be generated, and realistic predictions can be made based on ML algorithms by finding hidden relationships. One of the challenges in this approach is the difficulty in obtaining reliable datasets with sufficient data points. Well-developed materials may have a reasonable amount of experimental data. However, for new materials, such as the NP/polymer hybrid films, collecting large enough experimental data points is impossible. To overcome this problem, various simulation techniques have been utilized to generate new synthetic data, which are mostly based on density functional theory [21–23] and molecular dynamics [24–26]. For the NP/polymer hybrid films, finite element modeling (FEM) has been adopted to estimate the optical (wavelength at the highest absorbance and absorption cross-section) and photothermal (temperature) properties [5, 27–30]. Therefore, this method can be potentially used for synthetic data generation required for the correlation analysis and ML.

In this study, a new material informatics framework associated with ML and FE modeling is proposed for the gold NP/polymer hybrid films to predict the optical and photothermal properties by correlating these properties with the microstructural characteristics and processing conditions. The datasets contain both experimental data collected from literature and synthetic data generated from the FE modeling to satisfy the quantity and quality required for ML. The correlation analysis was conducted between features (characteristics of gold NPs and polymer matrices and laser properties) and targets (optical and photothermal properties of the hybrid films), where the validity of the results was carefully checked with the reported results. Different ML models were trained with and without the addition of synthetic data to the experimental data collection, and the model performance was compared and analyzed. The workflow of the adopted approach is shown in Fig. 1.

2 Computational methods

2.1 Data collection from literature

To unearth valid relationships and build high-fidelity ML models between microstructural features of gold NPs, inherent matrix properties, laser properties, and resulting optical and photothermal properties of the hybrid films, reliable data points must be obtained. Therefore, experimental data was of initial interest. Our data was collected from peer-reviewed literature contained in the Clarivate database named Web of Science Science Citation Index (SCI) (updated by 20 August 2020). Searching terms were combinations of gold nanoparticles, polymer, film, hybrid, optical, and/or photothermal. Arranging and sorting processes of the collected literature were conducted in a commercial citation software Endnote. The literature examined bulk materials was ruled out. Then, the research papers covering spherical gold NPs were selected since the shape of the NPs was diverse, from sphere to rod, triangle, cube, shell, etc. The papers synthesized new matrices (mixtures of two or more polymers) were removed as finding their material information was impossible in the later collecting process. The final sets of literature contained papers dealing with polymer hybrid films with spherical gold NPs, different polymer matrices, and different gold NP contents and diameters. A comprehensive sampling process was conducted, encompassing a total of 40 distinct papers. These papers were then categorized based on their optical properties (wavelength of absorbance) and photothermal properties (film temperature) resulting from the incident light. The compilation of these papers represented the

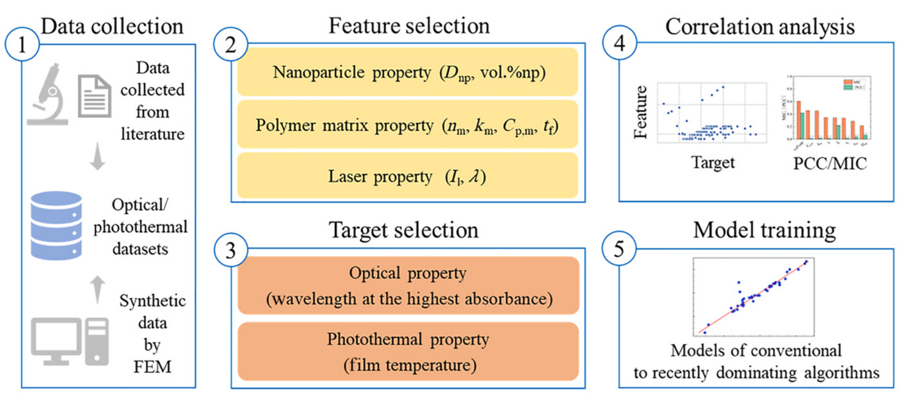


Figure 1 Schematic workflow of the machine learning approach coupled with FEM for the gold nanoparticle/polymer hybrid films.

most up-to-date and advanced data curation available for NP/polymer hybrid films.

From each source, the polymer matrix type, gold NP characteristics (diameter and content), laser information (laser intensity and wavelength), film thickness, and corresponding hybrid film properties (wavelength at the highest absorbance and film temperature) were extracted. Although there were attributes from the gold NPs, polymer matrix, and laser that could influence the hybrid film properties, they were neglected for the simplification of the datasets. For example, the dispersion nature of gold NPs in the polymer matrices largely impacts the optical and photothermal properties [5]. However, many papers omitted information about the NP dispersion, which led to our assumption that the NPs were uniformly distributed in the matrices. An increase in the irradiation time of the laser steadily ramps up the film temperature until it reaches a steady state. This was mostly missing in the literature. The type of polymer matrix, diameter and contents of gold NPs, and wavelength at which the highest absorbance occurred were collected from the optical property papers. Similarly, for the papers dealing with photothermal properties, relevant data on the diameter and content of gold NPs, the type of polymer matrix used, laser intensity and wavelength, and the film temperature during irradiation were meticulously gathered. Based on the extracted polymer information, the corresponding refractive index, thermal conductivity, and heat capacity were searched from handbooks and literature. Finally, 54 and 38 individual records were obtained for the optical and photothermal datasets, respectively. Table 1 shows a list of features and targets for the correlation analysis and model training.

2.2 Synthetic data generation from FE modeling

Although experimental data are considered the most reliable source of input for machine learning, the number of collected data points was too small to accomplish successful model training. In addition, the gold NP/polymer hybrid films are new materials, where the concept of self-healing by photothermal effects is also new. This led to scant data collection from the literature. To enrich the datasets, the FE modeling was adopted to generate new synthetic data points through various combinations of features applicable to model construction.

The FE modeling consisted of electromagnetic and heat transfer analyses in the nanoscale and another heat transfer analysis in the macroscale by using the commercial FE software COMSOL Multiphysics. The numerical methods were developed in our previous study [5]; nanoscale and macroscale models were constructed to estimate the optical and photothermal responses, respectively. When incident laser is irradiated on the gold NPs, the scattering field on one NP can interact with those on the other NPs

Table 1 List of features and targets for correlation analysis and model training

Classification	Subdivision	Parameters
Feature	Gold nanoparticle	Diameter (D_{np}), volume fraction (vol.%np)
	Polymer matrix	Refractive index (n_m), thermal conductivity (k_m), heat capacity ($C_{p,m}$), film thickness (t_f)
	Laser	Intensity (I_l), wavelength (λ)
Target	Optical property	Wavelength at the highest absorbance
	Photothermal property	Film temperature

in proximity. In this way, the electric field strength $\boldsymbol{E}$ is calculated by the sum of the incident electric field $\boldsymbol{E}_{inc}$ and the scattering field $\boldsymbol{E}_{sca}$ of the neighboring NP [31]

$$\boldsymbol{E} = \boldsymbol{E}_{inc} + \boldsymbol{E}_{sca} \tag{1}$$

The incident electric field $\boldsymbol{E}_{inc}$ can be expressed as follows when the incident laser travels along the positive x-direction with polarization in the z-direction

$$\boldsymbol{E}_{inc} = \overrightarrow{\boldsymbol{x}}\boldsymbol{E}_0 e^{-\mathrm{j}\cdot k_w \cdot \boldsymbol{z} \cdot n_m} \tag{2}$$

where $\overrightarrow{\boldsymbol{x}}$ is the direction vector, j is the imaginary unit where $\mathrm{j} = \sqrt{-1}$, $\boldsymbol{z}$ is the electric field polarized along the z axis. $\boldsymbol{E}_0$, e, k_w, and n_m are the incident electric field, electron charge, wave number, and refractive index of the polymer, respectively. The near-field in a region with a constant permittivity and no charge is obtained using Maxwell's equation as Helmholtz's equation [32]

$$\nabla \times \nabla \times \boldsymbol{E}_{sca} - \omega^2 \varepsilon_0 \mu_0 \varepsilon_r \boldsymbol{E}_{sca} = 0 \tag{3}$$

where the nabla symbol (∇) represents the gradient of the function. ε_r means relative permittivity. ω, ε_0, and μ_0 are angular frequency, vacuum permittivity, and magnetic permeability in a vacuum, respectively. The solution yields the electromagnetic behaviors of the NPs, such as electric field enhancement and absorption cross-section σ_{abs}.

The temperature evolution in the hybrid films due to the photothermal effect is estimated by solving a heat transfer equation with the addition of a heat source from the gold NPs. During laser irradiation, individual NPs absorb and dissipate energy, and act as individual heat sources that can be expressed as [27]

$$Q = \frac{\sigma_{abs} I_l}{V_{np}} \tag{4}$$

where I_l and V_{np} are the intensity of the incident field and volume of one NP, respectively. Therefore, the film temperature changes with the energy of the NPs. The local temperature field in the polymer matrix is generated by the heat dissipated from the NPs as a function of time, which is governed as

$$\rho_m C_{p,m} \frac{\partial T_m}{\partial t} = \nabla (k_m \nabla T_m) + Q \tag{5}$$

where ρ_m, $C_{p,m}$, T_m and k_m are the density, heat capacity, temperature, and thermal conductivity of the polymer matrix, respectively. The symbols in Eqs. (1)–(5), their values and material properties used for the simulations are explained in Table S1 and Fig. S1 in the Electronic Supplementary Material (ESM).

For the nanoscale model, eight spherical gold NPs of identical diameters were generated in a cube, as shown in Figs. 2(a) and 2(b). Four NPs were positioned at the top half of a cube, while the remaining four were placed on the bottom half to maximize their interactions. The edge length of the cube was determined by the diameter and volume fraction of the NPs. The interparticle distance was set to be identical to the distance of one particle to the closest wall of the cube. The diameter was varied at 2, 5, 10, 20, 30, and 50 nm, while the volume fraction was examined at 0.01%, 0.1%, 0.5%, 1%, 3%, and 10%. Also, five different polymer matrices, including polydimethylsiloxane (PDMS), polyvinyl alcohol (PVA), polyvinyl pyrrolidone (PVP), polystyrene (PS), and polyethyl-enimine (PEI), were selected to investigate the effect of the matrix refractive index on the properties of the hybrid films. The refractive index of the matrices used is shown in Fig. S1(b) in the ESM.

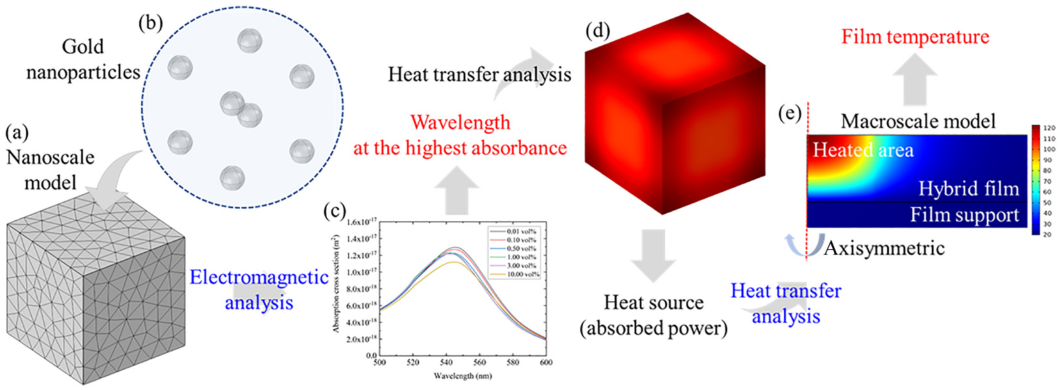


Figure 2 (a) A nanoscale model with meshes containing (b) eight gold nanoparticles, its typical results after (c) electromagnetic and (d) heat transfer analyses, and (e) a macroscale model after heat transfer analysis with a typical result of temperature estimation of hybrid films.

As for the electromagnetic analysis in nanoscale, the cubes were embedded in a controlled volume called a perfectly matched layer. The edge length of this layer was 20 times larger than the NP diameter to avoid any unnecessary reflection of light. The wavelength of the incident electric field from 500 to 600 nm was tested with an interval of 1 nm. The field propagated parallel to the positive *x*-axis and polarized along the *z*-axis. After the electromagnetic analysis, the absorption cross-section σ_{abs} was plotted as a function of wavelength (Fig. 2(c)). By analyzing the graph, the wavelength at the highest absorbance was determined, caused by the localized surface plasmon resonance (SPR). The absorption cross-section at this wavelength was then applied to Eq. (4) to calculate the heat source of the individual gold NPs, which was used as input for the heat transfer analysis to predict the volume-averaged heat source of the entire cube (Fig. 2(d)).

For the macroscale analysis, a two-dimensional axisymmetric model was created, comprising the hybrid film and its supporting material (Fig. 2(e)). The hybrid film was sectioned into two areas: one exposed to laser irradiation and the other remaining unirradiated. The volume-averaged heat sources obtained from the nanoscale analysis were used as input for predicting the temperature of the laser-irradiated area under steady-state conditions. The material properties (density, thermal conductivity, and heat capacity) of the hybrid films were calculated using the rule of mixture equation as a function of the volume fraction of NPs and polymer matrices. Three different laser intensities were chosen for the analysis (5000, 50,000, and 540,000 W/m^2).

A total of 180 simulations were carried out using the nanoscale models, comprising combinations of 5 volume fractions, 6 diameters, and 5 polymer matrices. For the macroscale analysis, 540 simulations were conducted (180 × 3 laser intensities = 540). The maximum absorption cross-section and the corresponding wavelength were extracted after each electromagnetic analysis of the nanoscale models. The film temperature was collected from the macroscale models after the heat transfer analysis. It is important to note that certain film temperatures were excluded from synthetic data collection due to their unrealistically high values, which could be attributed to either the use of high laser intensity or a high volume fraction of the NPs. The estimated temperatures higher than 250 °C were eliminated since burn-out of the films would occur in actual experiments. Moreover, a few nanoscale analyses resulted in divergent solutions, and consequently, these instances were omitted from the data collection process.

2.3 Data cleanup

With the addition of synthetic data to the experimental data, the optical and photothermal datasets possessed 213 and 220 individual records, respectively. From a data science point of view, outliers might be present in the datasets, which would cause problems in ML model training and should be eliminated. Possible occurrences of the outliers are miswritten values in literature, unrealistic simulation results, and human errors during data collection and data investigation of the experiments and simulations. Under this consideration, the isolation forest algorithm was adopted to find anomalies in the sub-datasets [33]. The isolation forest stands as an unsupervised learning algorithm that operates on the principles of the decision tree algorithm to pinpoint anomalies by effectively isolating outliers present within the dataset. Through a process of recursive and random partition generation, this algorithm establishes condensed paths within trees to identify and classify anomalous instances. The determination of each data point's classification is executed by means of an assigned anomaly score, where a score approximating 1 denotes an anomaly, while a significantly lower score (below 0.5) indicates a normal observation. Within the scope of our study, the isolation forest algorithm was optimized through the manipulation of the fraction of contaminated data, ranging from 0.2% to 20%, which followed the same procedure as in our previous research [19]. The optimized algorithm was applied to the optical and photothermal datasets with and without the synthetic data. Since the reliability of the synthetic data was not accessed, its addition to the experimental data could be contamination. Two outliers were detected from the experimental optical datasets regardless of the addition of synthetic data. In the photothermal dataset, two outliers were observed, which were synthetic data. The datasets, after removing the outliers, were used for the correlation analysis and model training. The final numbers of the data points, features, and targets in each dataset are shown in Table 2.

If a feature has a relationship (i.e., linear) with another or other features, this correlation can deteriorate the performance of the ML models by introducing redundancy. This increases computational complexity and difficulty in interpreting the individual effects of

Table 2 Number of data points, features, and targets in each dataset

Sub-dataset	No. of data points	No. of features	No. of targets
Optical	207	3	1
Photothermal	218	8	1

清华大学出版社 Tsinghua University Press | SciOpen

features on the targets. Figures S2 and S3 in the ESM show the Pearson correlation coefficient (PCC) of the matrices based on the optical and photothermal datasets that depict linear relationships between the features, where the strong positive and negative correlations indicate values close to +1 and −1, respectively. The optical datasets have no correlation between the features regardless of the presence of synthetic data, while strong correlations mostly diminish in the photothermal dataset after adding synthetic data. In Fig. S2(b) in the ESM, for example, PCC between the thermal conductivity of the polymer matrix and the NP volume fraction is −0.85. Also, the heat capacity and refractive index of the polymer matrix have a strong positive correlation (0.7). These strong correlations in the experimental datasets have no physical meaning, at which PCC becomes 0.17 and −0.28 with synthetic data, respectively. The small number of data points in the photothermal dataset could result in incorrect interpretation of correlations between the features. The increase in the number of data points by the addition of synthetic data has a significant effect on reducing the misleading correlations between the features, as shown in Fig. S3(b) in the ESM. However, the strong positive correlation between the laser wavelength and the film thickness was still observed even with the synthetic data. In the FE modeling, the film thickness was set to be identical in all the simulations, which would lead to the correlation being the same as the results without synthetic data. Nevertheless, more precise ML models are expected by using the photothermal dataset with the synthetic data, as the correlation coefficients were reduced.

2.4 Correlation analysis

Correlation analysis provides rankings of features based on the coefficient used. This ranking can be used for selecting the most important parameters in the fabrication process and for tuning microstructures and thus properties [19, 34]. Two different correlation coefficients were adopted: PCC and maximal information coefficient (MIC). MIC describes nonlinear relationships between features and outputs. For comparison of MIC and PCC, the absolute value of PCC (|PCC|) was used as an inverse correlation, which presents a negative PCC value. Also, the correlation analysis was conducted on the datasets with and without the synthetic data to investigate the corresponding effect on the ML results.

2.5 Model training

Seven different ML algorithms were examined for training the models: linear regression (LR), Bayesian ridge regression (BR), support vector regression (SVR), kernel ridge regression (KR), k-nearest neighbors regression (KNN), random forest regression (RF), and XGBoost regression (XGB). The algorithms utilized in this study encompassed a range of methods, from conventional to state-of-the-art, to evaluate their performance. Through these analyses, the most suitable algorithm for predicting the optical and photothermal properties of hybrid films was identified. The comprehensive information about each algorithm can be found in our previous study [19]. The performance of models was assessed using mean absolute error (MAE) expressed as,

$$\mathrm{MAE} = \frac{1}{N}\sum_{N} |y - \widehat{y}| \tag{6}$$

where y is the true value and $\widehat{y}$ is the predicted value over the total number of data N. Even though the R-squared (R^2) score is often utilized for comparing the model performance, it represents the goodness of fit measure for linear regression models, which would be inappropriate for nonlinear ones. Therefore, MAE is preferable in our study compared to R^2. Considering the size of the datasets, 5-fold cross-validation was used. All the data points were scaled before training. For outlier detection, correlation analysis, and machine learning, the adoption of scikit-learn [35], an open-source Python toolkit, in combination with XGBoost [36], was implemented. The default hyperparameters were used for each ML model in comparing the prediction performance. The effect of including synthetic data in the model training was analyzed.

3 Results and discussion

3.1 Analysis without synthetic data

The results of correlation analysis between the features and targets are shown in Fig. 3, which used the datasets without the synthetic data. For the optical dataset, the most positive correlation given by PCC is the NP diameter, while the most negative one is the refractive index of polymers (Fig. 3(a)). MIC shows a distinct order when compared to PCC (refractive index of polymer, NP diameter, and volume fraction of NPs in sequence). The correlation results for the NP diameter in Fig. 3(a) describe the experimental observation well, which is satisfactory from the data analytics viewpoint. The wavelength at the highest absorbance (SPR band) is known to shift to a longer wavelength for the NPs with a larger diameter [37–39]. However, the effect of the volume fraction of NPs on the wavelength at the highest absorbance is still unclear. With a simultaneous increase in the volume fraction and a decrease in the NP diameter, a shift to a shorter wavelength range was reported in one study [38]. However, no change in the wavelength

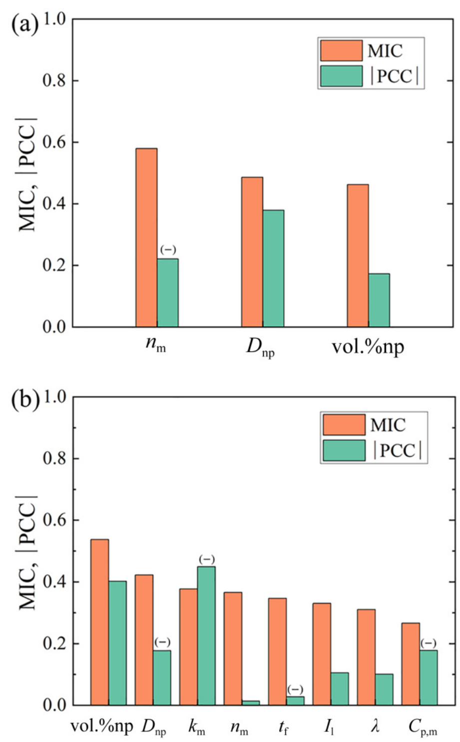


Figure 3 MIC and absolute PCC (|PCC|) of correlation analysis for (a) optical and (b) photothermal datasets without synthetic data. Minus (−) signs over the |PCC| bars indicate negative PCC values.

was found with increasing volume fraction in another study [40]. Also, volume fraction is highly related to the agglomeration of NPs. If NPs aggregate in the hybrid films, an additional peak should appear as a second plasmon band in the infrared region due to the coupling between the surface plasmon of the NPs in proximity. On the other hand, no shifting means the presence of individually separated NPs that have maintained their primary size [41–43]. The effect of the refractive index of polymers has been rarely investigated. Different polymer matrices with refractive index between 1.51 and 1.6 were tested with the same content of gold NPs, and no meaningful results about the wavelength were obtained [44].

The correlation analysis of the photothermal dataset effectively illustrates the experimental observations about the features (Fig. 3(b)). The most positively and negatively correlated features obtained with PCC are the volume fraction of NPs and the thermal conductivity of polymers, respectively. For MIC, the volume fraction of NPs holds the highest rank among the features, with the coefficient showing a declining trend in the following sequence: volume fraction of NPs, NP diameter, and thermal conductivity of the polymers. As each NP in the hybrid films serves as an individual heat source, the films with a higher volume fraction of NPs will produce more heat. Simultaneously, the superposed heating by heat fluxes diffused from the adjacent NPs also contributes to the rise in temperature [41]. Although a decrease in the radiation efficiency and an increase in the thermal conductivity of the films take place with more NPs, their effects are insignificant [40, 45]. The effectiveness of the heat isolation generated within NPs is enhanced when using less conductive polymer matrices, leading to a temperature increase in the films. This inversely proportional relationship is captured by the most negative PCC value. The NP diameter shows another negative PCC. The number of heat sources increases with decreasing NP diameter if the volume fraction is constant. Therefore, small NPs generate more heat compared to large NPs [46]. The heat capacity of the polymer matrix exhibits a negative correlation with the film temperature according to PCC. This implies that a polymer matrix with a high heat capacity would demand more energy to raise temperature.

For the refractive index of polymers, film thickness, and laser intensity and wavelength, PCC and MIC present completely different behaviors. No remarkable relationship between these features and film temperature is observed from PCC, while strong correlations are seen from MIC. Similar to the correlation results from the optical dataset, the relationships between the refractive index of the polymers and the film temperature are difficult to interpret because the laser penetrates the film for the given film thickness (1.3 mm) covered in the photothermal dataset, although the optical transmittance of the films decreases with increasing thickness. Further, a laser wavelength at the local surface plasmon resonance yields the highest absorption cross-section of NPs, consequently leading to the greatest temperature increase. In addition, the laser intensity is strongly proportional to the film temperature [47].

It should be noted that PCC and MIC consider different statistical relationships between features and targets; the correlation analysis is not to determine which coefficient is preferable over the other. It is possible that PCC better captures a linear relationship between one of the features and the target, while MIC is favorable to present a nonlinear one. The limited size of the datasets poses challenges in establishing correlations between the wavelength at the highest absorbance and the refractive index of the polymers, between the wavelength at the highest absorbance and the NP volume fraction (in the optical dataset), and between the film thickness and the film temperature (in the photothermal dataset), using both PCC and MIC. If more data related to these features are available, their effects on the targets can be assessed through correlation analysis.

The training of seven ML models was conducted without synthetic data. The model performance was compared by MAE, as shown in Fig. 4. The MAE units in Fig. 4 vary with the targets. For predicting the wavelength at the highest absorbance (Fig. 4(a)), MAE is the highest in KR, shows a similar level in LR, BR, and SVR, becomes smaller in KNN, and is further smaller in RF and XGB. The highest MAE for the film temperature is observed with the LR model, followed by BR, SVR, and KR models, while the error decreases in the sequence of KNN, RF, and XGB models (Fig. 4(b)). The KNN, RF, and XGB models demonstrate relatively accurate performance when using either the optical or photothermal dataset. Especially, the decision-tree based models (RF and XGB) show the lowest error. The linear regression models (LR and BR) train the data, assuming linear relationships between the features and targets. In the scatter plots of the features vs. target in Fig. S4 in the ESM, only several features present a linearity to the target (such as the NP diameter in the optical dataset and the laser intensity in the photothermal dataset), which would lead to large errors. Since the support vector machine-based models (SVR and KR) utilize linear kernels, the high MAE would be attributed to the linear functions. On the other hand, the models considering nonlinear relationships (KNN, RF, and XGB) seem preferable for predicting the optical and photothermal properties.

The shifting of wavelength with changes in the NP characteristics is reported in a range from a few to 50 nanometers. However, the

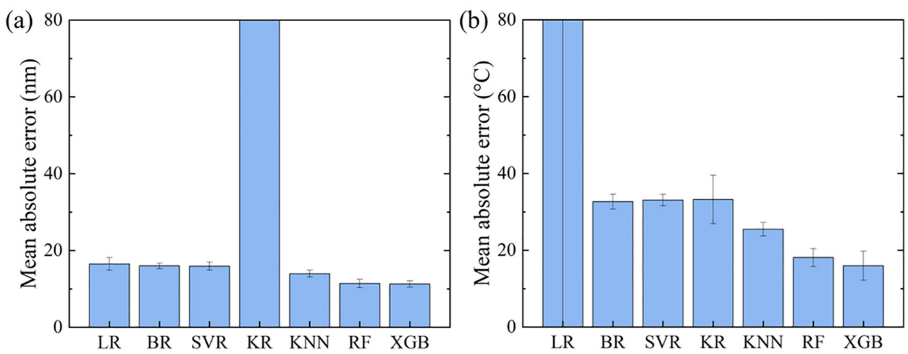


Figure 4 Prediction performance of the different ML models (LR, BR, SVR, KR, KNN, RF, and XGB) for (a) wavelength at the highest absorbance and (b) film temperature using the datasets without synthetic data.

MAE values (11.2–122.0 nm) are observed to exceed the shifted range, which means that the obtained ML models are incapable of accurately estimating the optical properties of the hybrid films. The error for the film temperature, as depicted in Fig. 4(b), is unrealistically substantial, spanning from 15.9 to 53,198 °C. Given the glass transition and melting temperatures of the polymer matrices, the high MAE values for the film temperature would potentially lead to false predictions.

3.2 Analysis with synthetic data

Correlation analysis was also performed using the datasets with synthetic data, and its results are shown in Fig. 5. There are changes in the coefficients and rankings after adding the synthetic data. For the MIC of the optical dataset (Fig. 5(a)), the refractive index of the polymers has a value almost close to 1, while the coefficients of other features decrease. Since the effects of the refractive index of the polymers on the wavelength at the highest absorbance have been unclear, the results in Fig. 5(a) are interesting in that a strong correlation is present. For PCC, the coefficients and rankings remain nearly identical to those without the synthetic data, except for the refractive index of the polymers, which shifts from positive to negative.

For the correlation analysis of the photothermal dataset with the synthetic data (Fig. 5(b)), MIC exhibits that the most correlated feature is the volume fraction of NPs, consistent with the results obtained without the synthetic data (Fig. 3(b)). However, the rankings of the other features change. The second-ranked feature, NP diameter, which previously held a strong correlation, becomes the least correlated when the synthetic data are included. Conversely, the heat capacity of the polymers, which was the least correlated feature without the synthetic data, now takes the second-ranked position with the synthetic data. In the results by PCC, the volume fraction of NPs retains its position as the top-ranked feature, the same as that from MIC. The most negatively correlated feature (thermal conductivity of the polymers) becomes non-correlated with the addition of the synthetic data. The PCC values of the other features are close to 0 after the inclusion of the synthetic data, except for the laser intensity, which exhibits an increase in correlation.

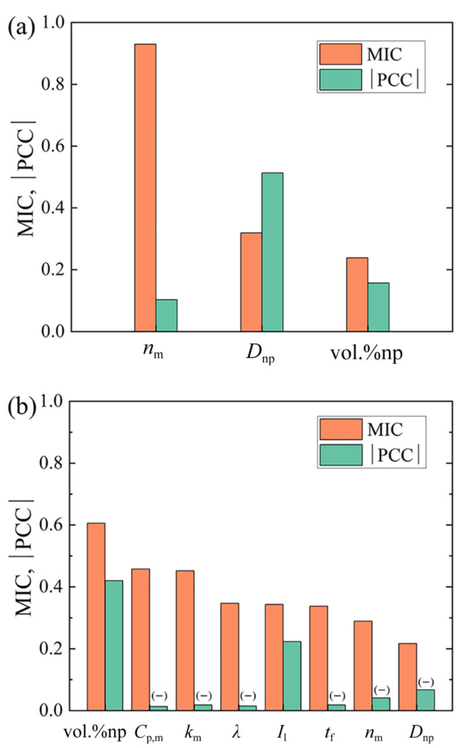


Figure 5 MIC and absolute PCC (|PCC|) of correlation analysis for (a) optical and (b) photothermal datasets with synthetic data. Minus (–) signs over the |PCC| bars indicate negative PCC values.

PCC for the optical and photothermal datasets changes significantly after adding the synthetic data. The presence of randomly distributed synthetic data points might weaken the linearity observed in the experimental datasets. Only the features known to have direct relationships with the targets, including the NP diameter in the optical dataset and the volume fraction of NPs in the photothermal dataset, have unremarkable impacts from the synthetic data.

Since the reliability of the synthetic data was not assessed, it is difficult to accept the accuracy of the predicted properties by the trained ML models. While the synthetic data could offer insights into the trends of the estimated properties based on feature variations, experimental validation of the synthetic data remains essential to evaluate its reliability properly. Therefore, two different routes of making the training and testing datasets are adopted to determine the effects of the synthetic data on the performance of the models. The first route is using the combined datasets of the experimental and synthetic data to separate them into training and test sets without distinguishing the sources of each data point. The second route is that 20% of the experimental data was left out for the test set before combining the experimental and synthetic data. Then, the rest of the experimental data and synthetic data are put together and used as the training set. We assumed that if the synthetic data is reliable, the model performance would be identical regardless of the route.

After the model training via the first route, MAEs as a function of the addition ratio of the synthetic data are compared in Fig. 6. For the optical dataset, KR exhibits the highest MAE, surpassing the y-axis range of Fig. 6(a). MAE reduces in the following order: LR, BR, SVR, KNN, RF, and XGB. This would be attributed to the scattered nature of the data points after the synthetic data addition, as the linearity of the features to targets was diluted due to the inclusion of the synthetic data (Fig. S5 in the ESM). It is encouraging that the accuracy of the models increases with the addition of the synthetic data (except for LR at 1.0 ratio). Figure 6(a) clearly shows this enhancement when compared with the MAE only with the experimental data (MAE_{exp}, Fig. 4(a)). The accuracy of the models with synthetic data is about 1.2–3.3 times better than that without. The KR model shows the most improvement with the synthetic data addition (Fig. 6(b)), although its MAE is still higher than that of the other models (beyond the y-axis range of Fig. 6(a)). The next enhanced model is XGB, the same as the model without the synthetic data.

For predicting the film temperature with the synthetic data, KR has the highest MAE among the models, where the error becomes smaller in SVR, BR, LR, and KNN, and further smaller in RF and XGB (Fig. 6(c)). With the addition of the synthetic data, only RF and XGB follow the same decreasing trend as that of the optical dataset (Fig. 6(a)). The rest of the models remain almost constant with small variations. Even from the 0.8 ratio, MAE slightly increases. Except for RF and XGB, the addition of the synthetic data is likely to contaminate the experimental data. This is demonstrated in Fig. 6(d), which compares the MAE values before and after the

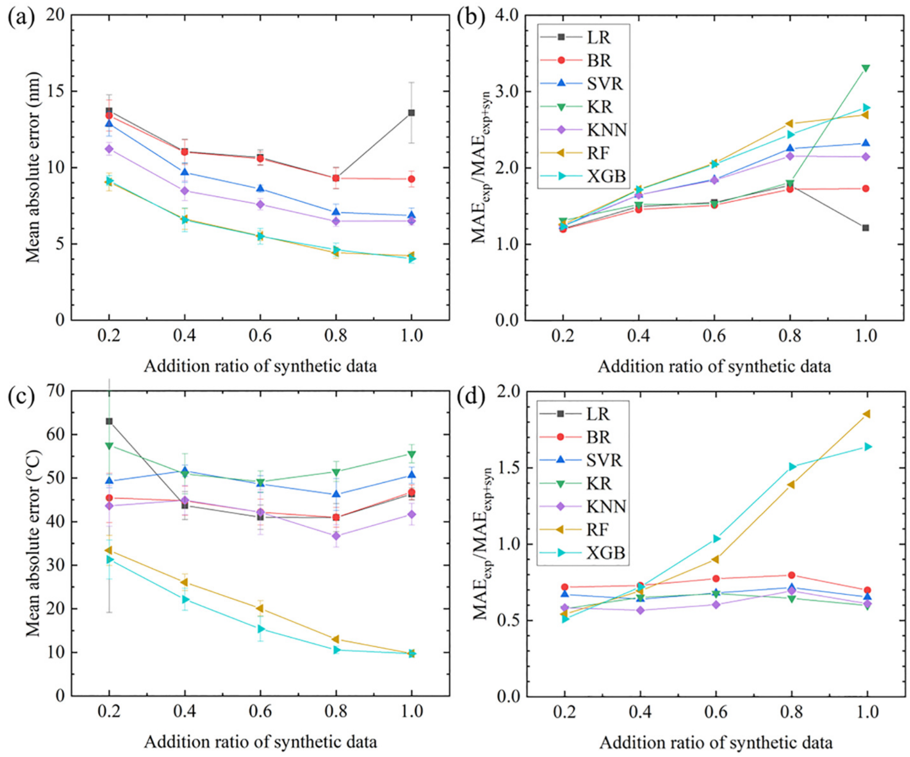


Figure 6 Prediction performance of different ML models (LR, BR, SVR, KR, KNN, RF, and XGB) for (a) wavelength at the highest absorbance and (c) film temperature as a function of the addition ratio of synthetic data where the training and test sets are split from the combined experimental and synthetic datasets. ((b) and (d)) Changes in the mean absolute error ratios.

addition. The MAE ratio of LR, BR, SVR, KR, and KNN is less than 1 regardless of the addition ratio, which means that the prediction performance is degraded due to the synthetic data. Small amounts of synthetic data addition (less than 0.6 ratio) deteriorate the accuracy of RF and XGB, while the ratio over 0.8 enhances the performance about 1.85 times compared to the models without the synthetic data.

The addition of the synthetic data to the experimental data leads to different outcomes (increase, decrease, or remain constant) depending on the type of ML models and properties estimated. This is caused by the reliability of the synthetic data, which is related to how precisely the FE models simulate the experimental phenomena. The declining MAE across the models for the optical property indicates that the synthetic data correctly describes the actual experiments, while the synthetic data in the photothermal dataset have the opposite effect.

The results of the model training through the second route are shown in Fig. 7. The trend of decreasing MAE with the addition of the synthetic data for the optical property follows the same pattern as observed in the first route (Fig. 7(a)). A comparison of the MAE values in Figs. 6(a) and 7(a) reveal that LR and KNN remain stable, SVR and KR increase, and BR, RF, and XGB decrease.

In the first route, as the synthetic data are included in the test dataset, the MAE of the trained models is anticipated to decrease with the addition of synthetic data. The FE analysis is implemented by solving the constitutive equations applied to the FE models. Therefore, the synthetic data, both in the training and test sets, reflect specific correlations of the features to the target underlying the FE analysis. On the other hand, in the second route, the errors of the ML models will only decrease if the synthetic data contain the correlations between the features and the target observed in the experimental data and if the ML models can interpret these correlations correctly. This is the case for the SVR and KR models (Fig. 7(a)). Although the model showing the most increased accuracy after the addition of the synthetic data is KR (Fig. 7(b)), its MAE is too high to ensure reliability. The MAE ratios of all the models are higher than 1, meaning enhanced performance with synthetic data.

The addition of the synthetic data to the photothermal dataset gives almost constant MAEs for SVR and KNN at 40–50 °C, as shown in Fig. 7(c), while LR, BR, and KR become less accurate. For instance, the MAE of LR rapidly increases up to 60% addition of synthetic data, and at 80% addition the MAE is about 10 times higher than the initial value. The most precise models are RF and XGB among the models. The addition of the synthetic data deteriorates the accuracy of all the models except for RF and XGB (Fig. 7(d)). When incorporating more than 80% of the synthetic data into the experimental data, RF and XGB demonstrate improved performance compared to the models without the synthetic data. The RF model experiences a reduction in MAE by up to 1.34 times.

From the data analytics perspective, the second route is more appropriate than the first one. The objective of model training is to estimate the properties based on actual experiments rather than simulation results. Therefore, it is reasonable to leave out a part of the experimental data as a test set, then the remaining data are combined with synthetic data as the training set. This method helps to compensate for the sparse data points. Moreover, it facilitates the determination of the reliability of the synthetic data. For example, if the MAE of a model decreases (even with a small amount of synthetic data) compared to using only the experimental data, it suggests that the simulations used to generate the synthetic data are

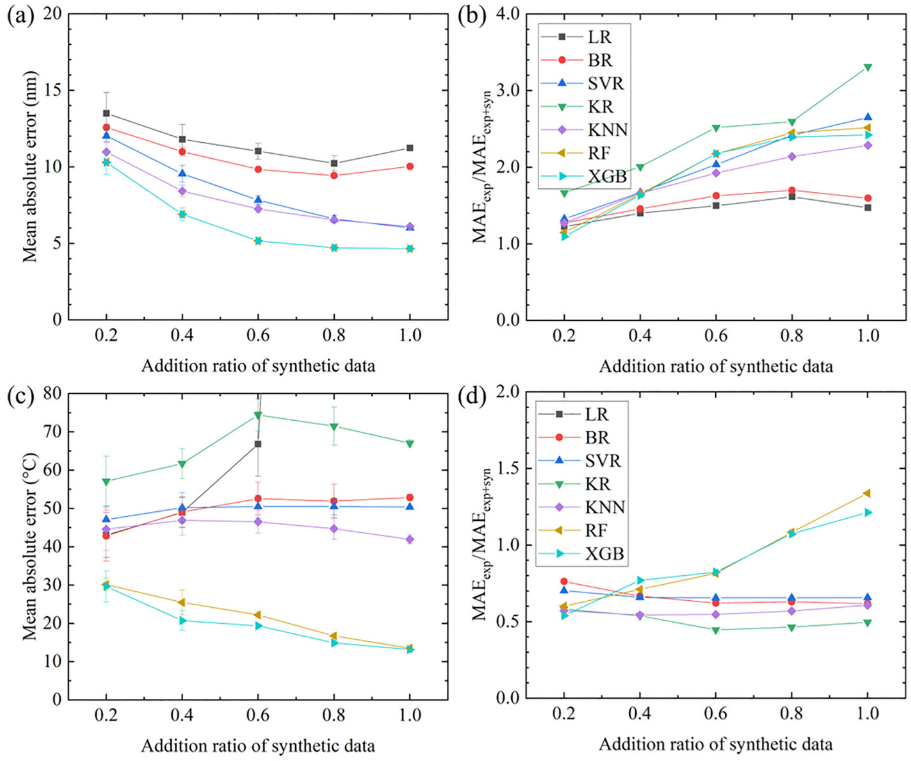


Figure 7 Prediction performance of the different ML models (LR, BR, SVR, KR, KNN, RF, and XGB) for (a) wavelength at the highest absorbance and (c) film temperature as a function of the addition ratio of the synthetic data where the test set is split from the experimental dataset (20%). ((b) and (d)) Changes in the mean absolute error ratios.

accurate. Conversely, if MAE remains lower than when using only the experimental data, regardless of the quantity of the synthetic data included, it becomes necessary to consider excluding the simulation inputs.

Before training models, datasets are generally split into training and test sets by a ratio of 8:2 during data pre-processing. However, in the second route, as 20% of the experimental data were reserved as the test set, the ratios of the test set to the total data points in the optical and photothermal datasets are only 3.8% and 4.6%, respectively. The number of data points in each test set is insufficient to examine the prediction performance of the trained models. Therefore, 50% of the experimental data were used as test sets, which were 13% and 8.7% of the optical and photothermal datasets, respectively. Then, the model performance is analyzed using the second route, where the results are shown in Fig. S6 in the ESM. Regardless of the amount of data points allocated for the test sets, MAE declines with the addition of the synthetic data in both the optical and photothermal datasets. With an increase of the experimental data added from 20% to 50% as the test sets, the accuracy of the ML models improves. For example, with a synthetic data addition ratio of 1, the RF model of the optical dataset exhibits an error value of 4.53 nm when 20% of the experimental data serves as the test set (Fig. 7(a)). However, when the test set is comprised of 50% of the experimental data, MAE decreases to 3.16 nm (Fig. S6(a) in the ESM). In the same manner, the MAE of the XGB model for the photothermal dataset decreases from 13.1 to 10.7 °C with increasing experimental data in the test set (Fig. 7(c) and Fig. S6(c) in the ESM).

It is interesting that the amount of synthetic data required to outperform the models solely trained with the experimental data changes with the size of the test set split from the experimental datasets. Only 60% of the synthetic data is necessary when 50% of the experimental data are used for the test set, as shown in Fig. 7(d). This means that fewer simulations are needed for generating the synthetic data. The more accurate ML models are obtained even with less amount of experimental data trained, which demonstrates

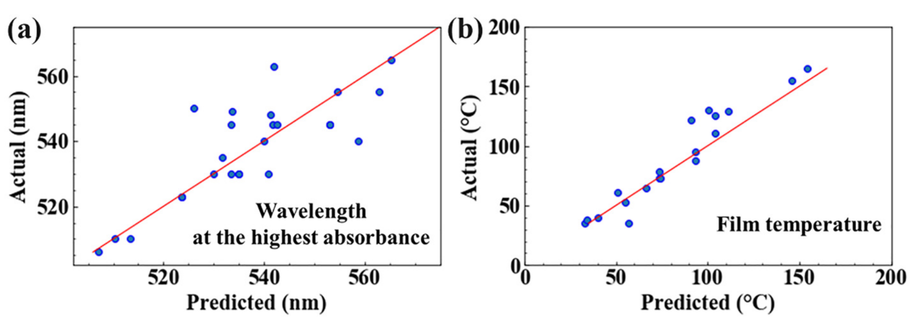


Figure 8 Actual versus predicted wavelength at the highest absorbance by the RF model and film temperature by the XGB model when the test sets are split from the experimental dataset (50%) and the addition ratio of the synthetic data is 1. (a) Wavelength at the highest absorbance, (b) film temperature.

the significance of optimizing the test set size. The size of the test set used in Fig. 7 is too small and inadequate for examining the ML models. Figure 8 shows the scatter plots between the actual and predicted optical and photothermal properties when 50% of the experimental data were used as the test sets and the addition ratio of the synthetic data was 1. The models predict the properties well, where the MAE of the RF and XGB models is 6.89 nm and 9.5 °C, respectively. However, larger scatters indicate the necessity of model improvement for better prediction.

3.3 Discussion

The experimental data used in this study were collected up to a specific time. As new data points are being continuously produced and reported, the accuracy of the ML models will increase with an accumulation of data. Also, the use of simulations to generate synthetic data can improve model performance. The proposed ML framework coupled with the FE modeling can be a supportive approach to developing predictive models for the properties of the NP/polymer hybrid films, where difficulty in collecting data is present. The developed FE models improve the accuracy of estimating the wavelength at the highest absorbance without experiments by variations of polymer types, NP diameters, and NP volume fractions as inputs. Once the ML models for the photothermal property become more precise, designing new hybrid films in conjunction with the models for the desired optical properties can be facilitated. For instance, hybrid films can be tuned to tailor local surface plasmonic resonance at a specific wavelength, then the film thickness and laser intensity can be chosen to yield maximum self-healing capacity at a desired temperature. Two examples of investigating potential research areas of hybrid films with the developed ML models are presented in the ESM (Figs. S7 and S8 and Tables S2 and S3 in the ESM). For the optical property, the influence of polymer refractive index on the wavelength at maximum absorbance was examined using both real and hypothetical polymers with the refractive index ranging from 1.40 to 2.30 (Table S2 in the ESM). The predicted wavelength changes, resulting from variations in the refractive index, were unique and previously unreported (Fig. S7 in the ESM), which emphasizes a need to consider the polymer effects on optical properties. In addition, the effects of thermal conductivity and heat capacity of polymer on the photothermal property were evaluated (Table S3 in the ESM), revealing that thermal conductivity should be prioritized over heat capacity for future studies (Fig. S8 in the ESM). These examples also demonstrate that the proposed ML framework has the potential to substantially decrease the experimental efforts, including human labor and time required for investigating new materials.

There are still some weaknesses to overcome in the proposed framework, including the FEM and ML methods. Although state-of-the-art experimental data were collected, the number of data points usable in the ML training was far less compared to the general ML approaches, as the concept of the NP hybrid films was relatively new. In the conceptualization step of this study, collections of the electrical properties (conductivity or resistivity) were carried out. However, only a little over 10 data points were obtained, which were dropped for study. A small-sized dataset is likely to generate a biased ML model. For example, in the photothermal experiments considered, the wavelength is in the range of 500–600 nm (Fig. S4 in the ESM). A few investigated the wavelength of over 800 nm. If experimental observations regarding the wavelength in the 600–800 nm, below 500 nm, and over 800 nm are available, the effects of the wavelength on the photothermal property can be more precisely interpreted by the ML models.

The research papers concerning the optical and photothermal properties were sorted manually, and the features were selected based on their relevance to the targets and feasibility of data extraction. Throughout this process, certain factors known to impact the optical and photothermal properties were deliberately excluded, which included the distribution and shape of NPs as well as the pulse duration and polarization of incident lights. Incorporating such omitted data would undoubtedly enhance the accuracy of the ML models.

For the FE modeling, although many attempts have been carried out to estimate the absorption cross-section of gold NPs, the geometry of the models used has been different from study to study and there is no standard model for the optical analysis. For example, only one sphere was used to model a single particle [48–51]. To understand the effects of the distance between two NPs in proximity on the local surface plasmonic resonance, two NPs were only positioned along or normal to the light direction [27]. Some studies dealt with three or more NPs in a row or multilayers [27–30, 52]. However, the FE models used in these previous studies cannot reflect the actual random, inhomogeneous, and/or aggregated distribution of NPs observed in the microstructures of the hybrid films. It was demonstrated that the NP distribution strikingly affected the absorption cross-section [5]. In this study, evenly distributed eight NPs with an equal interparticle distance were modeled. The effects of the NP number and distribution in the FE model on the estimated optical property in consideration of the features used are ambiguous. In this regard, several additional simulations were conducted to analyze the NP shape and spatial distribution effects on the optical and photothermal properties. To account for the NP shape, new FE models with nanocubes and nanorods were generated. Each nanocube and nanorod possessed a volume equivalent to that of a spherical NP with a diameter of 10 nm (Fig. S9(a) in the ESM), and the volume fraction was maintained at 0.5%. Also, a model with randomly distributed spherical NPs was constructed to introduce a spatial distribution different from the periodic arrangement (Fig. S9(a) in the ESM). The absorption cross-sections of the NPs with different shapes and orientations, embedded in a PDMS film, are presented in Fig. S9(b) in the ESM. The cross-sections differed depending on the NP shape, decreased in the order of nanocube, randomly distributed spherical NP, spherical NP, and nanorod. The wavelengths at maximum absorbance were 525 (spherical NP), 570 (nanocube), 520 (nanorod), and 529 nm (randomly distributed spherical NP). Using the obtained absorption cross-sections, the maximum temperature of the hybrid film under an incident laser intensity of 50,000 W/m^2 was estimated. The films containing spherical NPs, nanocubes, nanorods, and randomly distributed spherical NPs were heated to maximum temperatures of 112.7, 119.0, 60.3, and 116.9 °C, respectively. The NP shape and spatial orientation clearly impact the optical and photothermal properties, which indicates the need for further investigation. It may be necessary to develop new FE models that are independent of these effects, where experimental validation of the obtained synthetic data would support the accuracy of the new models.

Future efforts will focus on new data collection, improvement of the FE models, and experimental validation of the trained ML models. Manual collection of the data in the literature can be substituted with natural language processing to easily extract the experimental values. The FE models that well describe the actual microstructure of NPs and optimization of the boundary conditions in the photothermal analysis will improve the accuracy of the ML models even with less synthetic data. Possibly, new features based on thermodynamics and/or at the molecule levels can be utilized for correlation analysis and model training. Complex but powerful ML models, such as neural network, can be used if a large dataset is provided. Experimental validation of the predicted property and inverse prediction will maximize the usability of the ML models.

4 Conclusions

An ML approach coupled with finite element modeling is proposed for predicting the optical and photothermal properties of gold nanoparticle/polymer hybrid films. To ensure enough data points for reliable model training, both literature sources and finite element modeling are used to gather the optical and photothermal datasets. The datasets are comprised of features such as characteristics of the gold nanoparticles, polymer matrix, laser, and film, and targets including film properties (wavelength at the highest absorbance and film temperature). Correlation analysis and ML model training are conducted using datasets with and without synthetic data to assess the reliability of the synthetic data. From the correlation analysis, the rankings of the features describing the importance of the features in relation to targets are obtained, from which the similarities and differences to the experimental observation are discussed. Changes in correlations between the features and the targets, caused by the addition of synthetic data, bring attention to often-overlooked features with high ranks, which could be essential for manipulating the optical and photothermal properties of the hybrid films. The addition of synthetic data substantially enhances the accuracy of the ML models, although a threshold for the synthetic data included is present. In addition, the effects of the test set size split from the experimental data are analyzed, indicating that the performance of the ML models is highly influenced by the test set size. Overall, the approach demonstrated in this study can offer reliable prediction of optical and photothermal properties across a range of compositions and suggest valuable guidance for designing novel self-healing hybrid films. Moreover, incorporating relevant features and data with more precise finite element models will further improve the ML models.

Electronic Supplementary Material: Supplementary material (symbols and values for finite element modeling, PCC matrices, scatter plots, and examples of investigating potential areas of research using the developed ML models) is available in the online version of this article at https://doi.org/10.26599/NR.2025.94907216.

Data availability

All data needed to support the conclusions in the paper are presented in the manuscript and the Electronic Supplementary Material. Additional data related to this paper may be requested from the corresponding author upon request.

Acknowledgements

The authors would like to acknowledge the financial support from National Science Foundation under grant (No. CBET-2024546) and the School of Engineering at University of Alabama at Birmingham.

Declaration of competing interest

All the contributing authors report no conflict of interests in this work.

Author contribution statement

Y. J. C.: Conceptualization, data curation, formal analysis, investigation, methodology software, validation, visualization, writing – original draft writing. H. C.: Data curation, investigation. K. L.: Conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing – review & editing. All the authors have approved the final manuscript.

Use of AI statement

None.

References

[1] Hanemann, T.; Szabó, D. V. Polymer-nanoparticle composites: From synthesis to modern applications. *Materials* **2010**, *3*, 3468–3517.

[2] Dang, Z. M.; Yuan, J. K.; Zha, J. W.; Zhou, T.; Li, S. T.; Hu, G. H. Fundamentals, processes and applications of high-permittivity polymer-matrix composites. *Prog. Mater. Sci.* **2012**, *57*, 660–723.

[3] Otaegui, J. R.; Ruiz-Molina, D.; Hernando, J.; Roscini, C. Multistimuli-responsive smart windows based on paraffin-polymer composites. *Chem. Eng. J.* **2023**, *463*, 142390.

[4] Chen, L.; Si, L. P.; Wu, F.; Chan, S. Y.; Yu, P. Y.; Fei, B. Electrical and mechanical self-healing membrane using gold nanoparticles as localized "nano-heaters". *J. Mater. Chem. C* **2016**, *4*, 10018–10025.

[5] Cho, Y. J.; Kong, L. C.; Islam, R.; Nie, M. T.; Zhou, W.; Lu, K. Photothermal self-healing of gold nanoparticle-polystyrene hybrids. *Nanoscale* **2020**, *12*, 20726–20736.

[6] Yang, Y.; He, J. L.; Li, Q.; Gao, L.; Hu, J.; Zeng, R.; Qin, J.; Wang, S. X.; Wang, Q. Self-healing of electrical damage in polymers using superparamagnetic nanoparticles. *Nat. Nanotechnol.* **2019**, *14*, 151–155.

[7] Yang, Y.; Yang, M. H.; Zhang, S. M.; Lin, X.; Dang, Z. M.; Wang, D. R. Photoinduced healing of polyolefin dielectrics enabled by surface plasmon resonance of gold nanoparticles. *J. Appl. Polym. Sci.* **2019**, *136*, 47158–47167.

[8] Prezgot, D.; Bottomley, A.; Jorgenson, E.; Ianoul, A. Thermoplasmonic patterning of silver nanocrystal/polymer composite thin films. *Adv. Mater. Interface.* **2021**, *8*, 2100738.

[9] Amendola, V.; Pilot, R.; Frasconi, M.; Maragò, O. M.; Iatì, M. A. Surface plasmon resonance in gold nanoparticles: A review. *J. Phys.: Condens. Matter* **2017**, *29*, 203002.

[10] Cao, Z. X.; Wang, R. G.; Yang, F.; Hao, L. F.; Jiao, W. C.; Liu, W. B.; Wang, Q.; Zhang, B. Y. Photothermal healing of a glass fiber reinforced composite interface by gold nanoparticles. *RSC Adv.* **2015**, *5*, 102167–102172.

[11] Jain, P. K.; Lee, K. S.; El-Sayed, I. H.; El-Sayed, M. A. Calculated absorption and scattering properties of gold nanoparticles of different size, shape, and composition: Applications in biological imaging and biomedicine. *J. Phys. Chem. B* **2006**, *110*, 7238–7248.

[12] Govorov, A. O.; Richardson, H. H. Generating heat with metal nanoparticles. *Nano Today* **2007**, *2*, 30–38.
[13] Peng, P.; Zhang, B. Y.; Cao, Z. X.; Hao, L. F.; Yang, F.; Jiao, W. C.; Liu, W. B.; Wang, R. G. Photothermally induced scratch healing effects of thermoplastic nanocomposites with gold nanoparticles. *Compos. Sci. Technol.* **2016**, *133*, 165–172.
[14] Hashimoto, S.; Werner, D.; Uwada, T. Studies on the interaction of pulsed lasers with plasmonic gold nanoparticles toward light manipulation, heat management, and nanofabrication. *J. Photochem. Photobiol. C: Photochem. Rev.* **2012**, *13*, 28–54.
[15] Su, Q.; King, S.; Li, L. Y.; Wang, T. Y.; Gigax, J.; Shao, L.; Lanford, W. A.; Nastasi, M. Microstructure-mechanical properties correlation in irradiated amorphous SiOC. *Scr. Mater.* **2018**, *146*, 316–320.
[16] Halas, N. J.; Lal, S.; Link, S.; Chang, W. S.; Natelson, D.; Hafner, J. H.; Nordlander, P. A Plethora of plasmonics from the Laboratory for nanophotonics at rice university. *Adv. Mater.* **2012**, *24*, 4842–4877.
[17] Sharma, A.; Mukhopadhyay, T.; Rangappa, S. M.; Siengchin, S.; Kushvaha, V. Advances in computational intelligence of polymer composite materials: Machine learning assisted modeling, analysis and design. *Arch. Comput. Methods Eng.* **2022**, *29*, 3341–3385.
[18] Shin, D.; Yamamoto, Y.; Brady, M. P.; Lee, S.; Haynes, J. A. Modern data analytics approach to predict creep of high-temperature alloys. *Acta Mater.* **2019**, *168*, 321–330.
[19] Cho, Y. J.; Lu, K. A materials informatics approach for composition and property prediction of polymer-derived silicon oxycarbides. *Mater. Today Adv.* **2023**, *18*, 100384.
[20] Chong, X. Y.; Shang, S. L.; Krajewski, A. M.; Shimanek, J. D.; Du, W. H.; Wang, Y.; Feng, J.; Shin, D.; Beese, A. M.; Liu, Z. K. Correlation analysis of materials properties by machine learning: Illustrated with stacking fault energy from first-principles calculations in dilute fcc-based alloys. *J. Phys.: Condens. Matter* **2021**, *33*, 295702.
[21] Yi, Y.; Wang, L. M.; Chen, Z. Y. Adaptive global kernel interval SVR-based machine learning for accelerated dielectric constant prediction of polymer-based dielectric energy storage. *Renew. Energy* **2021**, *176*, 81–88.
[22] Shayeganfar, F.; Shahsavari, R. Deep learning method to accelerate discovery of hybrid polymer-graphene composites. *Sci. Rep.* **2021**, *11*, 15111.
[23] Fiedler, L.; Shah, K.; Bussmann, M.; Cangi, A. Deep dive into machine learning density functional theory for materials science and chemistry. *Phys. Rev. Mater.* **2022**, *6*, 040301.
[24] Jin, K.; Luo, H.; Wang, Z. Y.; Wang, H.; Tao, J. Composition optimization of a high-performance epoxy resin based on molecular dynamics and machine learning. *Mater. Des.* **2020**, *194*, 108932.
[25] Ma, R. M.; Zhang, H. F.; Xu, J. X.; Sun, L. N.; Hayashi, Y.; Yoshida, R.; Shiomi, J.; Wang, J. X.; Luo, T. F. Machine learning-assisted exploration of thermally conductive polymers based on high-throughput molecular dynamics simulations. *Mater. Today Phys.* **2022**, *28*, 100850.
[26] Li, H. X.; Tian, H.; Chen, Y. W.; Xiao, S. A.; Zhao, X. Y.; Gao, Y. Y.; Zhang, L. Q. Analyzing and predicting the viscosity of polymer nanocomposites in the conditions of temperature, shear rate, and nanoparticle loading with molecular dynamics simulations and machine learning. *J. Phys. Chem. B* **2023**, *127*, 3596–3605.
[27] Zeng, Z. C.; Wang, H.; Johns, P.; Hartland, G. V.; Schultz, Z. D. Photothermal microscopy of coupled nanostructures and the impact of nanoscale heating in surface-enhanced Raman spectroscopy. *J. Phys. Chem. C* **2017**, *121*, 11623–11631.
[28] Yanagiya, S. I.; Sekimoto, N.; Furube, A. Photothermal dynamics of micro-glass beads coated with gold nanoparticles in water: Fine bubble generation and fluid-induced laser trapping. *Jpn. J. Appl. Phys.* **2018**, *57*, 115001.
[29] Palermo, G.; Cataldi, U.; Condello, A.; Caputo, R.; Bürgi, T.; Umeton, C.; De Luca, A. Flexible thermo-plasmonics: An opto-mechanical control of the heat generated at the nanoscale. *Nanoscale* **2018**, *10*, 16556–16561.
[30] Hasegawa, M.; Watanabe, K.; Namigata, H.; Welling, T. A. J.; Suga, K.; Nagao, D. Surface lattice resonance in three-dimensional plasmonic arrays fabricated via self-assembly of silica-coated gold nanoparticles. *J. Colloid Interface Sci.* **2023**, *633*, 226–232.
[31] Jain, P. K.; El-Sayed, M. A. Plasmonic coupling in noble metal nanostructures. *Chem. Phys. Lett.* **2010**, *487*, 153–164.
[32] Jin, J. M. *The Finite Element Method in Electromagnetics*; 3rd ed. John Wiley & Sons: Hoboken, 2014.
[33] Liu, F. T.; Ting, K. M.; Zhou, Z. H. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp 413–422.
[34] Reshef, D. N.; Reshef, Y. A.; Finucane, H. K.; Grossman, S. R.; McVean, G.; Turnbaugh, P. J.; Lander, E. S.; Mitzenmacher, M.; Sabeti, P. C. Detecting novel associations in large data sets. *Science* **2011**, *334*, 1518–1524.
[35] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
[36] Chen, T. Q.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016, pp 785–794.
[37] Kelly, K. L.; Coronado, E.; Zhao, L. L.; Schatz, G. C. The optical properties of metal nanoparticles: The influence of size, shape, and dielectric environment. *J. Phys. Chem. B* **2003**, *107*, 668–677.
[38] Abyaneh, M. K.; Paramanik, D.; Varma, S.; Gosavi, S. W.; Kulkarni, S. K. Formation of gold nanoparticles in polymethylmethacrylate by UV irradiation. *J. Phys. D: Appl. Phys.* **2007**, *40*, 3771–3779.
[39] Nadal, E.; Barros, N.; Peres, L.; Goetz, V.; Respaud, M.; Soulantica, K.; Kachachi, H. *In situ* synthesis of gold nanoparticles in polymer films under concentrated sunlight: Control of nanoparticle size and shape with solar flux. *React. Chem. Eng.* **2020**, *5*, 330–341.
[40] Dunklin, J. R.; Forcherio, G. T.; Berry, K. R.; Roper, D. K. Asymmetric reduction of gold nanoparticles into thermoplasmonic polydimethylsiloxane thin films. *ACS Appl. Mater. Interfaces* **2013**, *5*, 8457–8466.
[41] Hong, J.; Hwang, D. K.; Park, C.; Kim, Y. Photothermal performance of plasmonic patch with gold nanoparticles embedded on polymer matrix. *Korean J. Chem. Eng.* **2019**, *36*, 1746–1751.
[42] Janovák, L.; Dékány, I. Optical properties and electric conductivity of gold nanoparticle-containing, hydrogel-based thin layer composite films obtained by photopolymerization. *Appl. Surf. Sci.* **2010**, *256*, 2809–2817.
[43] Sheeney-Haj-Ichia, L.; Sharabi, G.; Willner, I. Control of the electronic properties of thermosensitive poly(*N*-isopropylacrylamide) and Au-nano-particle/poly(*N*-isopropylacrylamide) composite hydrogels upon phase transition. *Adv. Funct. Mater.* **2002**, *12*, 27–32.
[44] Abargues, R.; Abderrafi, K.; Pedrueza, E.; Gradess, R.; Marques-Hueso, J.; Valdés, J. L.; Martínez-Pastor, J. Optical properties of different polymer thin films containing *in situ* synthesized Ag and Au nanoparticles. *New J. Chem.* **2009**, *33*, 1720–1725.
[45] Berry, K. R. Jr. ; Russell, A. G.; Blake, P. A.; Roper, D. K. Gold nanoparticles reduced in situ and dispersed in polymer thin films: Optical and thermal properties. *Nanotechnology* **2012**, *23*, 375703.
[46] Jiang, K.; Smith, D. A.; Pinchuk, A. Size-dependent photothermal conversion efficiencies of plasmonically heated gold nanoparticles. *J Phys. Chem. C* **2013**, *117*, 27073–27080.

[47] Maity, S.; Bochinski, J. R.; Clarke, L. I. Metal nanoparticles acting as light-activated heating elements within composite materials. *Adv. Funct. Mater.* **2012**, *22*, 5259–5270.

[48] Grosges, T.; Barchiesi, D. Gold nanoparticles as a photothermal agent in cancer therapy: The thermal ablation characteristic length. *Molecules* **2018**, *23*, 1316.

[49] Aibara, I.; Mukai, S.; Hashimoto, S. Plasmonic-heating-induced nanoscale phase separation of free poly(*N*-isopropylacrylamide) molecules. *J. Phys. Chem. C* **2016**, *120*, 17745–17752.

[50] Setoura, K.; Okada, Y.; Werner, D.; Hashimoto, S. Observation of nanoscale cooling effects by substrates and the surrounding media for single gold nanoparticles under CW-laser illumination. *ACS Nano* **2013**, *7*, 7874–7885.

[51] Enders, M.; Mukai, S.; Uwada, T.; Hashimoto, S. Plasmonic nanofabrication through optical heating. *J. Phys. Chem. C* **2016**, *120*, 6723–6732.

[52] Davletshin, Y. R.; Kumaradas, J. C. The role of morphology and coupling of gold nanoparticles in optical breakdown during picosecond pulse exposures. *Beilstein J. Nanotechnol.* **2016**, *7*, 869–880.