

Sustainable Dyeing Process Modeling for Recycled PET/PCT Microfibers via Gaussian Process Regression

Hyeokjun Cho and Seung Geol Lee*

Cite This: *ACS Omega* 2025, 10, 43150–43160

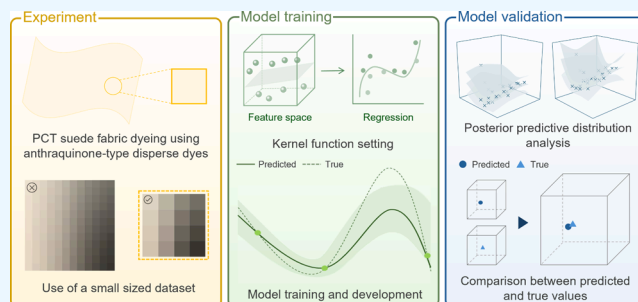
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: As the textile industry moves toward more sustainable and resource-efficient manufacturing, minimizing the experimental burden in dyeing processes has become increasingly critical. This study presents a Gaussian process regression (GPR)-based framework for predicting the colorimetric outcomes of dyeing processes involving ecofriendly fiber blends composed of recycled polyethylene terephthalate and polycyclohexylene dimethylene terephthalate. Using only 52 experimental data points, the model was trained to predict CIELAB color coordinates (L^* , a^* , b^*) as well as the K/S value based on dyeing variables such as temperature, time, and dye concentration. The GPR model achieved high prediction accuracy with coefficients of determination (R^2) of 0.96, 0.96, 0.73, and 0.95 for L^* , a^* , b^* , and K/S , respectively. Moreover, the probabilistic nature of GPR enables uncertainty quantification through posterior predictive distributions, offering both mean estimates and 95% confidence intervals. This capability supports robust decision-making in dyeing process design and quality control, especially in low-data regimes. The proposed approach demonstrates significant potential for reducing resource consumption and experimental iterations in fiber coloration, contributing to the development of data-efficient and environmentally sustainable dyeing systems.



1. INTRODUCTION

In the textile industry, there is a growing need for precise and efficient dyeing technologies that can simultaneously achieve quality control, process optimization, and reduction of environmental impact.^{1–5} In particular, dyeing is the most resource-intensive stage in textile processing, consuming large amounts of water, energy, and chemicals. As a result, there is growing interest in sustainable production methods that aim to reduce water usage, improve energy efficiency, and minimize chemical pollution and waste.^{6–10} In conventional dyeing processes, achieving consistent color reproduction requires repeated experiments and manual adjustment of process conditions, which significantly increases the required time as well as costs and waste generation.¹¹ Consequently, machine-learning prediction models have been explored as a way to reduce the number of repeated experiments and improve dyeing efficiency.^{12–16}

Machine learning-based prediction models offer the advantage of identifying dyeing conditions that yield colors close to the desired target without repeated experiments by learning from existing experimental data and predicting outcomes under new conditions.^{17,18} The L^* , a^* , and b^* in the widely used CIELAB color space are closely related to visual color perception. In addition, the K/S value, which represents the ratio of absorption to scattering, is another important indicator of color yield in dyed fabrics. By

incorporating this color strength metric into the prediction model, it is possible to more comprehensively evaluate dye uptake and color intensity, which are closely associated with fastness, a major indicator of fiber performance in textile applications. A model that accurately predicts these values can be highly useful for evaluating dyeing quality and designing dyeing processes.^{19–21} However, developing such a model requires sufficient amount of training data, which is often difficult to obtain due to the physical and economic limitations of actual textile dyeing experiments. During the early stages of research on a single color or a specific material, only a small sample size is typically available. In such small-sample settings, commonly used machine learning techniques such as support vector machine (SVM) and artificial neural network (ANN) tend to overfit, resulting in poor generalization performance, making it difficult to achieve accurate predictions.^{22–24}

In this study, a Gaussian process regression (GPR) model was introduced to address the challenge of a small sample size,

Received: July 2, 2025

Revised: August 7, 2025

Accepted: September 3, 2025

Published: September 13, 2025



aiming to achieve high predictive performance even with limited data.^{25–28} GPR is a probabilistic, Bayesian regression method that provides not only a single predicted value but also a confidence interval associated with the prediction.^{29–32} This characteristic provides a practical advantage over conventional linear regression models because it enables a quantitative assessment of prediction certainty. Such capability is particularly valuable for experimental design and dyeing process control. GPR is especially effective in environments with limited training data, as it helps prevent overfitting while maintaining strong generalization performance.^{33,34}

In this study, a GPR-based prediction model was developed using 52 experimental samples. The dyeing temperature, time, and dye concentration served as input variables, and the corresponding L^* , a^* , b^* , and K/S values of the dyed PCT fabric were the output variables to be predicted. To identify the optimal kernel function, cross-validation was performed on several possible kernels, and the model performance was evaluated using quantitative metrics such as mean squared error (MSE), mean absolute error (MAE), and R^2 . Furthermore, the posterior predictive distribution was analyzed to visualize predictive uncertainty, which is one of the core strengths of GPR. The agreement between predicted and actual values was also examined using several validation methods.

The approach proposed in this study serves as a foundation for future smart dyeing processes that enable color control without repeated experiments. By leveraging GPR's probabilistic predictions, the model can increase confidence in process design decisions and support decision-making while minimizing resource waste.^{35–38} It can also be applied to quality control and experimental design optimization, either as an alternative to conventional empirical approaches or as a complementary method.^{39–42} This study demonstrates that the GPR model can achieve high-confidence predictions even with a small sample size, highlighting its potential to contribute to the development of eco-friendly, sustainable dyeing processes.

2. EXPERIMENTAL SECTION

2.1. Materials. The fabric samples used in this study were sea-island type microfiber suede fabrics provided by Huvis Co., Ltd., South Korea and the characteristics of the dyed fabrics are shown in Table 1.

Table 1. Specifications of Suede Fabric

parameter	value
Yarn spec.	Recycled PET 50%: PCT 50%
Suede spec.	Yarn 70%: PU 30%
Type	Nonwoven fabric
Size (mm)	130
Thickness (mm)	1.18
Weight (g/m ²)	13.5

The suede was produced by blending recycled polyethylene terephthalate (PET) and polycyclohexylene dimethylene terephthalate (PCT) microfiber in a 50:50 ratio, followed by 30% impregnation with polyurethane (PU) resin. Figure 1 shows the chemical structures of the recycled PET and PCT polymers used.

Figure 1(a) shows the chemical structure of recycled PET, which was obtained by mechanically and chemically recycling waste PET bottles. Figure 1(b) presents the chemical structure

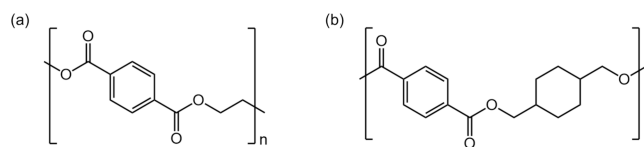


Figure 1. Chemical structure of fabric (a) recycled PET, and (b) PCT.

of PCT, a copolyester fiber material synthesized through an ester polycondensation reaction of terephthalic acid (TPA) or dimethyl terephthalate (DMT) with the precursor 1,4-cyclohexane dimethanol (CHDM). The disperse dye used for dyeing was Dorospers Dark Gray KKL, an anthraquinone-based dye (Archroma Co., Ltd., Switzerland). The dispersing agent used in the dye bath was Sunsoft RM-340S (NICCA KOREA Co., Ltd., Japan), and the pH of the dye bath was adjusted using acetic acid (CH₃COOH) (SAMIL Co., Ltd., South Korea). For the reduction cleaning process, the reducing agent Sera Con M-FAS (Dystar, Singapore) and NaOH (SAMIL Co., Ltd., Korea) were used.

2.2. Dyeing. Figure 2 illustrates the dyeing and reduction cleaning process of suede fabric.

A dye bath was prepared using a liquor ratio of 1:20 for 13.5 g of suede fabric. Disperse dye was added at concentrations of 5% and 10% owf (on weight of fabric). Additionally, 1.89 mL of a 2 wt % dispersing agent and 5 wt % acetic acid were added to adjust the pH of the dye bath to 5. Dyeing was performed in an infrared dyeing machine (DL-6000, Daelim Starlet Co., South Korea) at 100–135 °C for 10–60 min. After dyeing, a reduction cleaning step was conducted to remove any unfixed dye from the fabric surface. A cleaning solution with a 1:20 liquor ratio, containing 6 g/L of the reducing agent (Sera Con M-FAS) and 6 mL/L of NaOH, was applied at 85 °C for 30 min to carry out the reduction cleaning process. The cleaned fabrics were rinsed with lukewarm water and dried, resulting in a total of 52 dyed samples under various process conditions.

2.3. Color Measurement. To evaluate the dye uptake of the dyed fabrics, a spectrophotometer (SCINCO Co., Ltd., South Korea) was used to measure surface reflectance at 20 nm intervals within the visible wavelength range of 360–740 nm under a 10° standard observer and D65 illumination conditions. From these measurements, the L^* , a^* , and b^* values were obtained. To minimize measurement error, each fabric sample was measured four times, and the average value was used as the final data.

2.4. Dye Uptake Characteristics. The adsorption and diffusion behavior of disperse dyes depends on the microstructure and chemical affinity of the fiber substrate. Polyester fibers such as rPET and PCT lack charged groups, so ionic bonding with dyes is negligible. Heating the fabric above its glass transition temperature expands the amorphous regions and allows anthraquinone dyes to diffuse into the polymer's free volume. The absorbed dye remains in the fiber chiefly through hydrophobic and π - π interactions between its aromatic rings and the polymer chains, assisted by weaker dipole and hydrogen-bond forces. In this study, the interaction strength between dye and fiber in rPET/PCT microfibers was quantified using the Kubelka–Munk K/S value. The Kubelka–Munk equation is presented in Eq. 1.

$$K/S = \frac{(1 - R)^2}{2R} \quad (1)$$

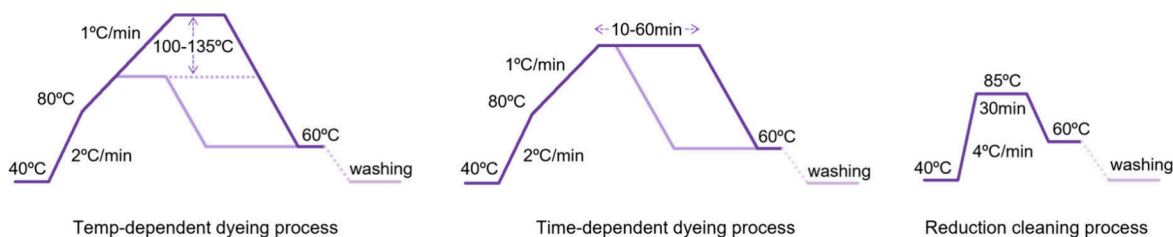


Figure 2. Schematic diagram of the dyeing and reduction cleaning process of the suede fabric.

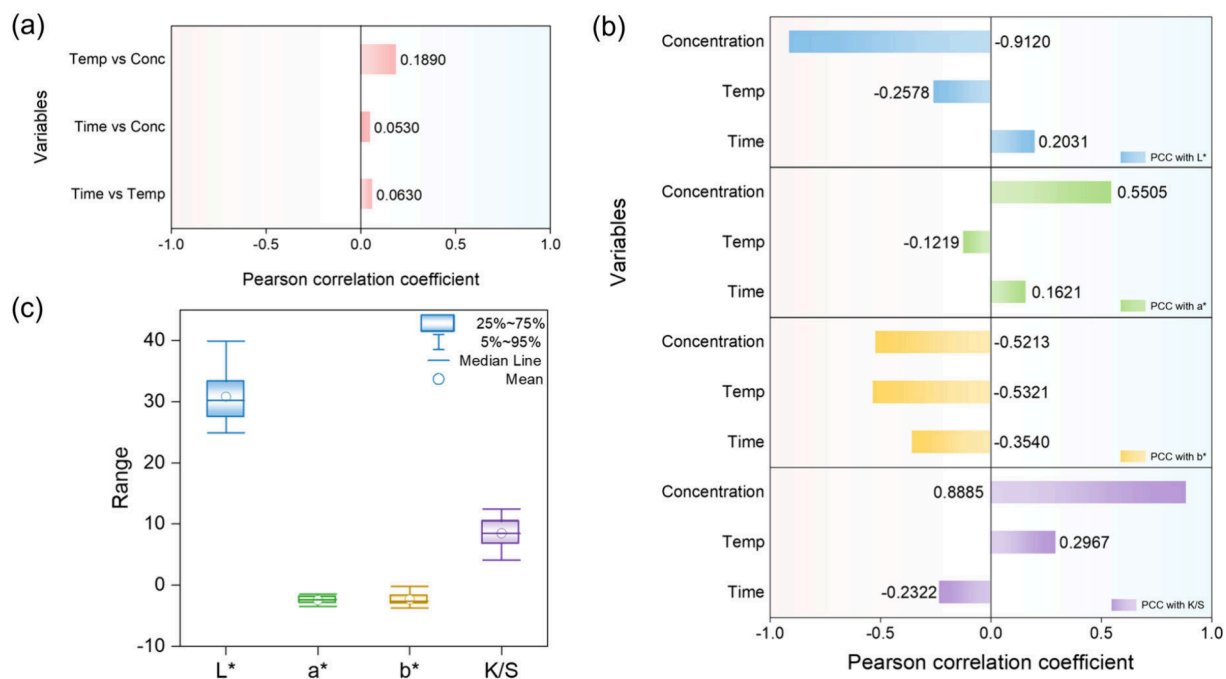


Figure 3. (a) Pearson correlation coefficient between input variables, (b) Pearson correlation coefficient between input and output variables, and (c) Box plots of the output variables.

where K is the absorption coefficient, S is the scattering coefficient, and R is the reflectance of the sample. Surface reflectance spectra were measured with a spectrophotometer, and the highest K/S value among all measured wavelengths was taken as the final result.

2.5. Model Selection. In this study, dyeing time (min), temperature ($^{\circ}\text{C}$), and dye concentration (% owf) were used as input variables, and the output variables were the L^* , a^* , b^* , and K/S values of the dyed fabric. The machine learning model used was GPR, a Bayesian regression method that estimates the distribution of functions based on given inputs. Given the limited data set (52 samples), GPR was selected for its ability to provide reliable predictions and strong generalization performance with small data. All modeling was implemented in Python 3.10.0 using scikit-learn 1.6.1.

2.6. Model Evaluation Criteria. To quantitatively evaluate the performance of the GPR model in predicting color data based on dyeing process variables, three representative regression metrics were used: MAE, MSE, and R^2 . MAE represents the mean absolute difference between predicted and actual values and provides an intuitive measure of prediction error. The MAE was calculated by eq 2.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

where y_i is the actual measured value from the experiment, and \hat{y}_i is the predicted value from the model. MSE is more sensitive to large errors and is useful for evaluating the robustness of the model. The MSE formula is presented as eq 3.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

The R^2 value indicates how well the model explains the actual data. A value closer to 1 suggests that the model has higher explanatory power. The formula for calculating R^2 is presented as eq 4.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

where \bar{y} denotes the mean of the actual experimental values. In addition, considering the small training data set in this study, a 5-fold cross-validation (CV) was performed to control overfitting and evaluate the model's generalization performance. The entire data set was randomly divided into five subsets, with one subset used for validation and the remaining subsets used for training. This process was repeated five times, and the average values of the evaluation metrics were used as the final performance results.

3. RESULTS AND DISCUSSION

3.1. Data Analysis. To quantitatively examine the degree of interaction among variables and to assess how each input variable contributes to predicting the output, the correlations were analyzed among the dyeing process parameters (dyeing temperature, time, and dye concentration) as well as between the input variables and the output color coordinates (L^* , a^* , and b^*) and color strength (K/S value). The analysis was conducted by calculating the Pearson correlation coefficient (PCC) based on the Pearson correlation formula, which is presented in eq 5.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

The PCC ranges from -1 to 1 , where a value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.^{43,44} A higher absolute PCC value indicates a stronger relationship between two variables, meaning that changes in one variable are closely associated with changes in the other. On the other hand, a lower absolute value indicates that the relationship is weak and difficult to interpret. The analysis revealed that the relationships among the input variables were generally weak. In contrast, certain input variables showed relatively strong relationships with specific output variables, suggesting that some dyeing conditions significantly affect particular color coordinates. The PCC results for both input–input and input–output variable pairs, along with the box plots illustrating the distribution of the output variables, are presented in Figure 3.

Figure 3(a) shows the correlations among the three input variables, which are dyeing temperature, time, and dye concentration. All correlation coefficients are below 0.2 , indicating very weak relationships and suggesting minimal influence among the input variables. Figure 3(b) presents the correlations between each input variable and the output variables, specifically the L^* , a^* , b^* , and K/S values that represent the fabric's color characteristics. The strongest correlation is observed between dye concentration and the L^* value, which represents the lightness of the fabric. The correlation coefficient is approximately -0.91 , indicating that the lightness significantly decreases as the dye concentration increases. Dye concentration is also strongly and positively correlated with the K/S value, reflecting the intuitive increase in overall color strength as the amount of dye increases. Similarly, the a^* value, which represents the red-green axis of the color space, tends to increase with higher dye concentrations, shifting the fabric color toward red. Conversely, the b^* value, reflecting the yellow-blue axis, tends to decrease, shifting the color toward blue. Temperature shows a moderate positive correlation with the K/S value, while dyeing time exhibits a weak negative correlation, suggesting that prolonged time alone does not necessarily enhance color yield under the studied conditions. As shown in Figure 3(c), the L^* value exhibited a low correlation with dyeing temperature and time. Similarly, the a^* and b^* values showed weaker correlations with temperature and time than with dye concentration, partly due to their relatively narrow data distributions, which made such relationships more difficult to identify. The K/S value follows the same trend, where dye concentration is the dominant factor, while temperature and

dyeing time have only a minor impact. These results suggest that the input variables can be considered largely independent, with dye concentration being a more critical factor for color change compared to temperature and time. The strong correlations between dye concentration and the L^* , a^* , b^* , and K/S values confirm that the fabrics in the data set were properly dyed under the given experimental conditions, accurately representing both color depth and chromatic characteristics.

3.2. Model Configuration. The data set consisted of 52 experimental samples, of which 80% were randomly selected as training data and the remaining 20% were used as test data. Since the input variables had different units, standardization was performed by adjusting the mean to 0 and the standard deviation to 1 to ensure that all variables contributed equally to model training. The same standardization was applied to the test set. The kernel function is a core component of the GPR model, enabling it to learn nonlinear relationships between input variables and directly influencing its predictive accuracy and generalization performance. Therefore, the choice of kernel is critically important.^{45–47} In this study, various kernel function combinations were constructed by incorporating the White kernel into existing kernels to model measurement noise and observation errors more accurately and to better reflect the uncertainty present in actual data.^{48,49} Among these combinations, the kernel function combination with the lowest MSE value from the CV was selected for the final model. The CV MSE results for each kernel function combination are presented in Table 2.

Table 2. CV MSE Results for Different Kernel Combinations

kernel function	CV MSE
Matern + White	0.926
Rational quadratic + White	0.949
RBF + Matern + White	0.971
RBF + White	1.043

Among all the combinations tested, the combination of the Matern kernel and White kernel showed the lowest MSE and was selected as the final model. The length-scale parameter (l), which controls the range of influence over input correlations in the kernel function, was set to 1 . The smoothness parameter (ν) of the Matern kernel, which determines the complexity and variability of the function to be learned, was set to 1.5 . This kernel combination achieved the lowest MSE and effectively captured underlying data patterns even with the small sample size, which justifies its selection as the final model. The final model was trained using the entire data set with standardized input values. Hyperparameters such as the length scale and noise level were optimized by repeating the training process up to 10 times. This multi-iteration optimization helped the model approach a global optimum and improved its generalization performance. Since the GPR model predicts four outputs, the model was designed to predict them in parallel while sharing the same kernel across all outputs.

3.3. Model Performance Evaluation. The predictive performance of the GPR model was evaluated using 5 -fold CV on the data set of 52 dyeing experiments. The MAE, MSE, and R^2 values for each output variable of the GPR model are presented in Table 3.

Table 3. Predictive Performance Metrics of the GPR Model Based on 5-fold CV

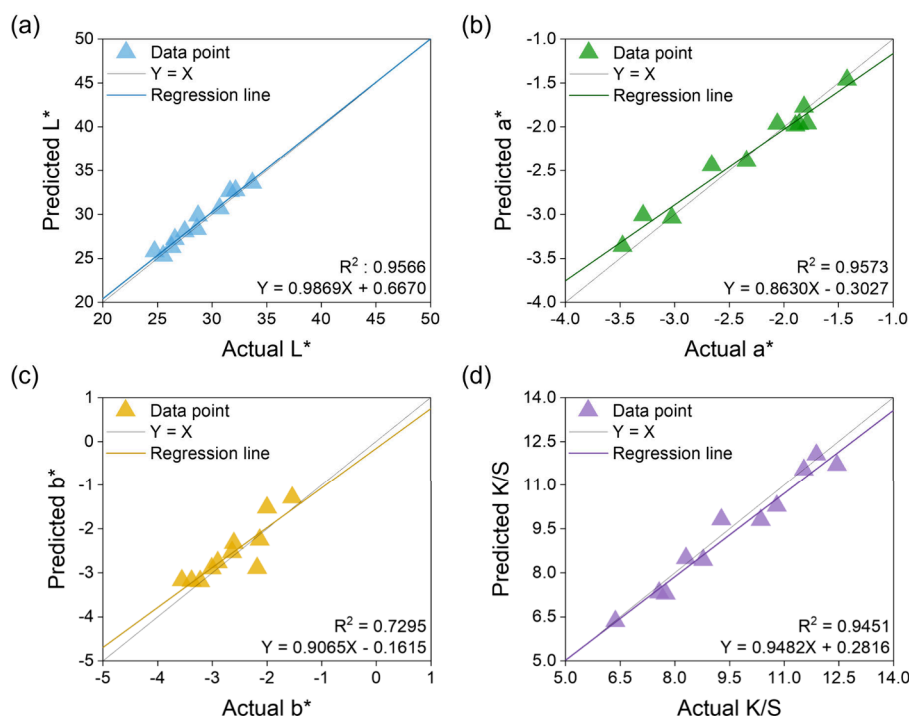
metrics	L^*	a^*	b^*	K/S
MAE	0.4950	0.1129	0.2441	0.3608
MSE	0.3444	0.0183	0.0993	0.1857
R^2	0.9566	0.9573	0.7295	0.9451

The prediction results for each output variable demonstrated consistently high accuracy. As shown in Table 3, the predicted MAE and MSE for the L^* were 0.4950 and 0.3444, respectively, which are very low given that the lightness scale ranges from 0 to 100. Similarly, the MAE values for a^* and b^* were 0.1129 and 0.2441, and the MSE values were 0.0183 and 0.0993, indicating minimal differences between the predicted and actual values. For the K/S output, the MAE and MSE were 0.3608 and 0.1857, respectively, showing that the predicted color strength values closely matched the experimental measurements as well. The R^2 values confirm the model's strong predictive capability: 0.9566 for L^* and 0.9573 for a^* (indicating that about 95% of the variance is explained in each case), 0.7295 for b^* , and 0.9451 for K/S . The somewhat lower R^2 for b^* suggests that b^* is harder to predict, possibly due to its narrower range or higher noise, but it still indicates a reasonable fit. These results suggest that, despite the small sample size, the GPR model successfully captured the complex nonlinear relationships in the dyeing process, resulting in reliable predictions for all color components and the K/S value confirming the model's suitability for this application.

3.4. Model Validation. To validate the trained GPR model, the relationship between the predicted and actual values was examined using linear regression analysis. The results of the regression analysis, including the regression lines representing the relationship between the predicted and actual values, are presented in Figure 4.

Figure 4 presents a linear regression analysis with the actual values on the x -axis and the predicted values on y -axis, illustrating the relationship between the known measurements and the predictions generated by the model. A high prediction accuracy is indicated when the data points lie close to the regression line. The results show that all output variables exhibit a strong linear relationship between actual and predicted values, with most points distributed near the ideal $y = x$ line. The predicted L^* values are tightly clustered along the line, demonstrating a high degree of agreement with the actual values. The a^* and b^* values also show strong correlations, despite some minor deviations. Likewise, the K/S predictions align closely with the ideal $y = x$ line, indicating accurate modeling of color strength. The equations of the regression lines indicate that the slopes are close to 1 and the intercepts are near 0 for all output variables, suggesting that the predictions are unbiased and consistent with the actual values. In addition to the regression analysis, this study also examined the quantification of predictive uncertainty, one of the main advantages of the GPR model. Rather than producing a single point estimate, GPR provides a distribution of possible outcomes for each input within a Bayesian regression framework. This allows the model to quantify the confidence of each prediction. Figure 5 visualizes the posterior predictive distribution for each output variable, including the predicted mean and the 95% confidence intervals ($\pm 1.96\sigma$).

The analysis showed that the confidence intervals were narrow for most data points, indicating that the model generated predictions with a high level of certainty. The centers of the predictive distributions were also closely aligned with the actual measured values, visually confirming that the predictions were generally unbiased and reasonably accurate. To further evaluate the reliability of the predictive distributions, the absolute errors between the actual and predicted values were compared against the 95% confidence intervals for

**Figure 4.** Result of linear regression analysis with (a) L^* values, (b) a^* values, (c) b^* values, and (d) K/S values.

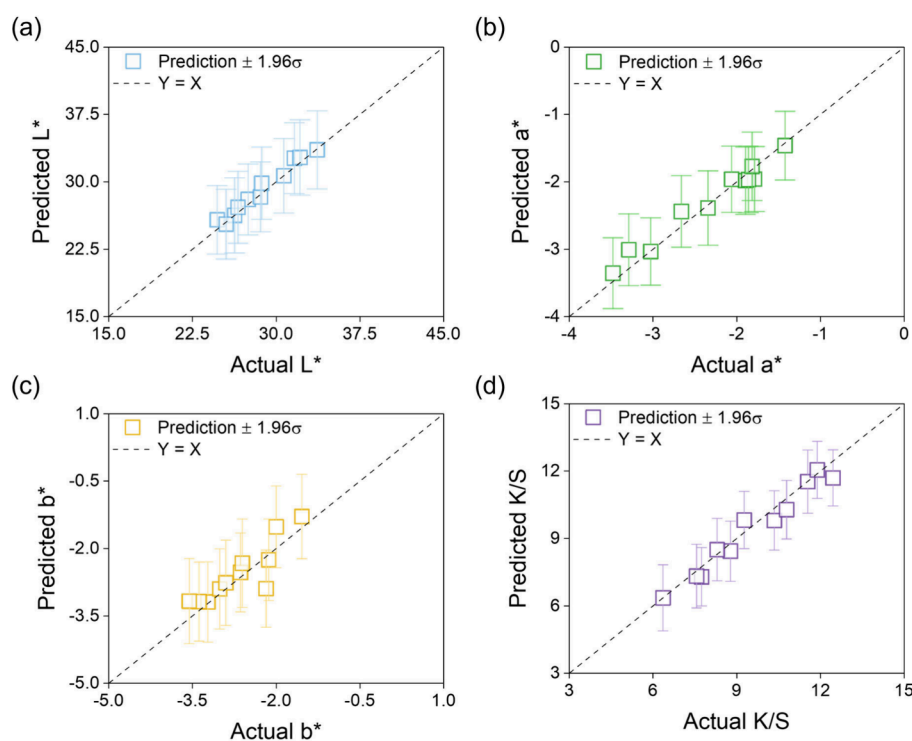


Figure 5. GPR-based posterior predictions of each output variable (a) L^* values, (b) a^* values, (c) b^* values, and (d) K/S values.

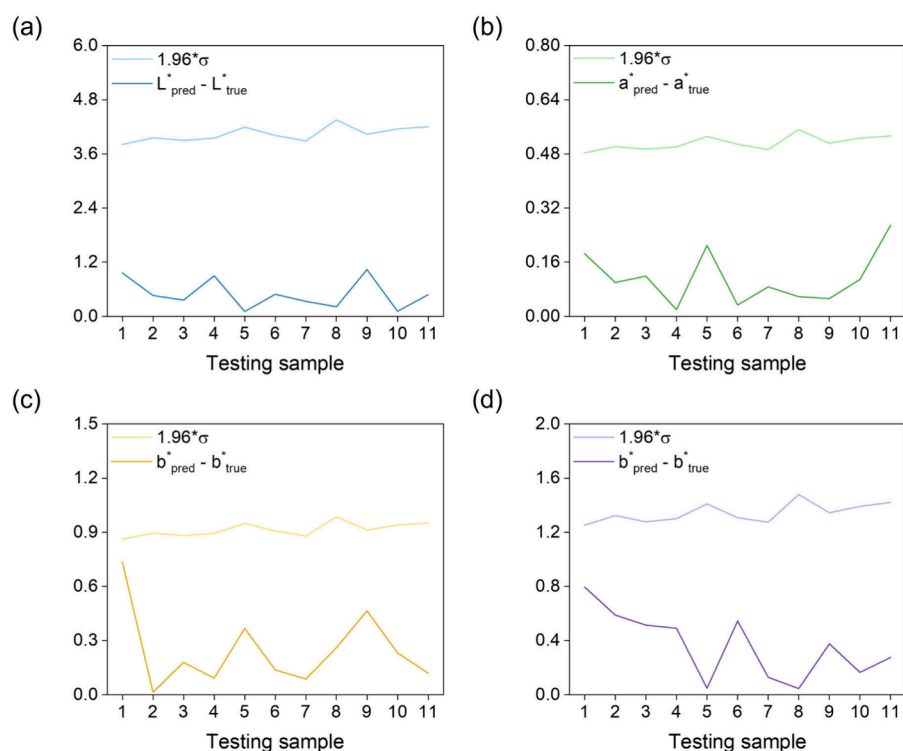


Figure 6. Comparison between prediction errors and 95% confidence intervals for (a) L^* values, (b) a^* values, (c) b^* values, and (d) K/S values.

each output variable. The results of this comparison are presented in Figure 6.

For all samples of each output variable, the absolute errors were smaller than the corresponding confidence intervals, indicating that the actual values were well contained within the confidence bounds provided by the GPR model. These results confirm that the GPR model demonstrated consistent

performance in both prediction accuracy and uncertainty estimation, suggesting its potential as a reliable predictive tool in dyeing processes where repeated experiments are challenging.

3.5. Posterior Prediction Surface Analysis. One of the main advantages of the GPR model is its ability to generate and analyze posterior predictive distributions for new inputs.

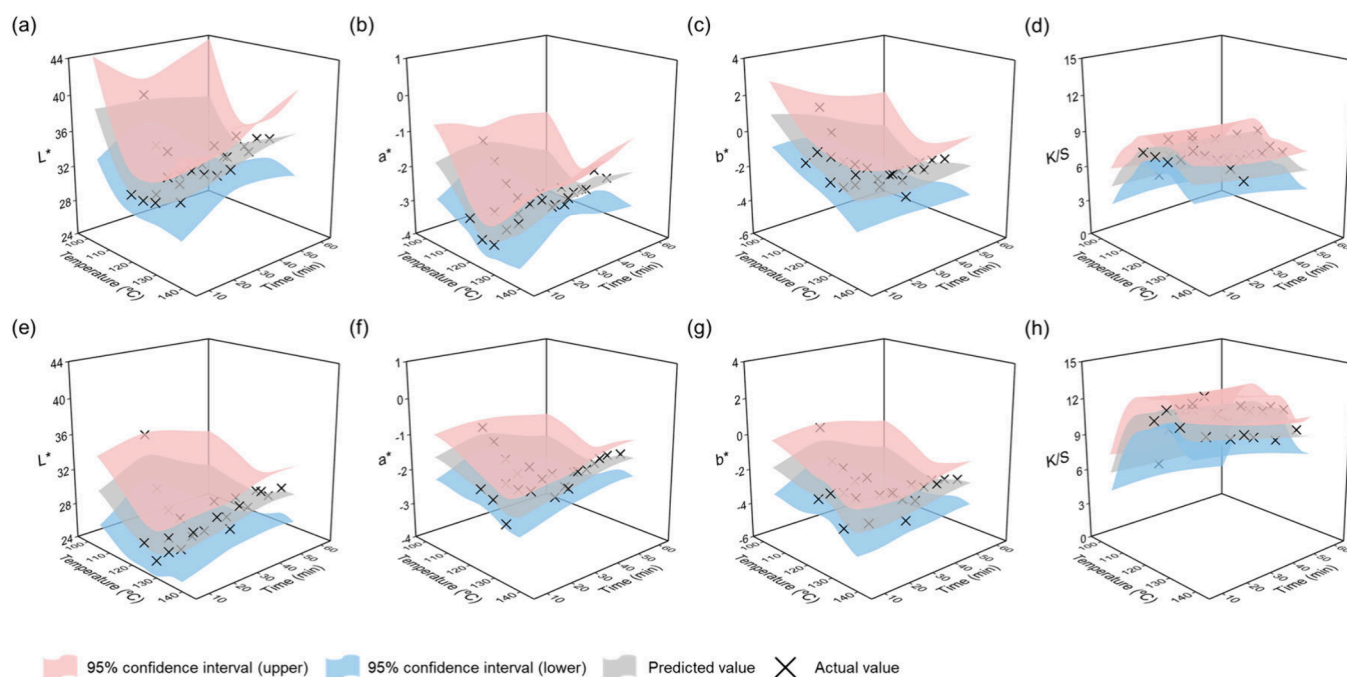


Figure 7. Posterior predictive surfaces of (a) L^* value, (b) a^* value, (c) b^* value, (d) K/S value at 5% dye concentration, and (e) L^* value, (f) a^* value, (g) b^* value, (h) K/S value at 10% dye concentration, plotted against dyeing temperature and time with 95% confidence intervals and actual values.

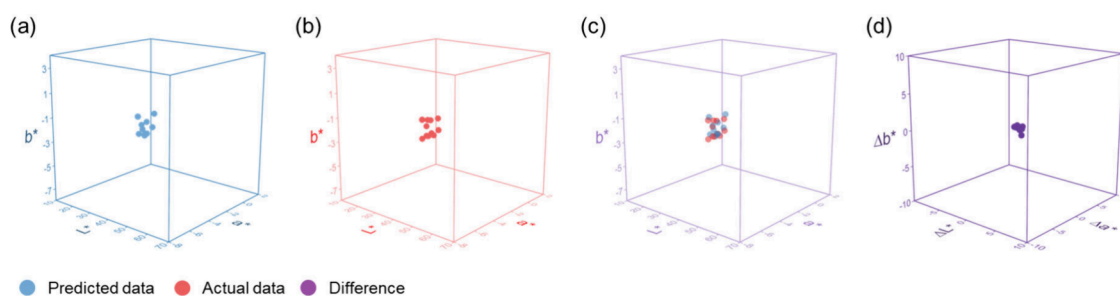


Figure 8. Comparison of color values in 3D space (a) predicted values by GPR model, (b) actual values by experiment, (c) combined scatter plot of predicted and actual values, (d) absolute error between predicted and actual values.

Based on this property, prediction surfaces were constructed for each output variable by setting dyeing time and temperature as independent variables, while fixing dye concentration at 5% and 10%. The actual measured values corresponding to each condition were also visualized alongside the predictions. The posterior prediction results are presented in Figure 7.

For a dye concentration of 5%, Figure 7(a–d) presents the predicted mean surfaces for L^* , a^* , b^* , and K/S , respectively. Similarly, Figure 7(e–h) shows the corresponding surfaces for L^* , a^* , b^* , and K/S at a dye concentration of 10%. Within the temperature and time ranges where training data were concentrated, the predicted surfaces closely matched the actual measured values, and most data points fell within the 95% confidence intervals provided by the GPR model. This indicates that the model maintained high predictive accuracy in these regions and provided reliable estimates of output variability. In areas where the confidence intervals appeared relatively narrow, the model demonstrated high certainty in its predictions under those specific conditions. In contrast, regions with limited or no training data showed more abrupt changes in the surface slope and broader confidence intervals, reflecting

increased uncertainty in those extrapolated areas. Predictions that extrapolate beyond the original data range carry increased risk and require cautious interpretation, underscoring an inherent limitation in the model's ability to generalize. These results highlight the relationship between data density in the input space and predictive uncertainty, offering a practical means of identifying less stable prediction zones. Such visual analysis based on posterior distributions underscores the value of GPR not only in generating point predictions but also in quantifying the level of certainty associated with each prediction. This capability is particularly useful for designing dyeing process conditions with greater confidence and identifying potential risks in experimental settings where repeated trials are limited. The current data set comprises dye concentrations of 5% and 10% o/w, which are sufficient for the present analysis. Incorporating intermediate concentrations and additional types of disperse dyes in future work would enable the model to encompass a broader concentration range and enhance its generalizability.

3.6. Comparative Analysis. *3.6.1. Comparison with Actual Data.* To visually compare the performance of the GPR model, the predicted color values were plotted alongside the

Table 4. Comparison of Actual and Predicted Values with Absolute Errors for Each Test Sample

sample	actual values				predicted values				absolute error			
	L^*	a^*	b^*	K/S	L^*	a^*	b^*	K/S	L^*	a^*	b^*	K/S
1	24.76	-1.78	-2.17	12.48	25.72	-1.97	-2.90	11.68	0.96	0.19	0.73	0.80
2	27.53	-1.89	-3.21	10.37	28.00	-1.99	-3.20	9.78	0.47	0.10	0.01	0.59
3	28.64	-1.85	-3.37	9.29	28.28	-1.97	-3.19	9.80	0.36	0.12	0.18	0.51
4	31.69	-3.02	-3.00	7.76	32.59	-3.04	-2.91	7.27	0.90	0.02	0.09	0.49
5	26.33	-2.66	-3.55	11.57	26.22	-2.45	-3.18	11.52	0.11	0.21	0.37	0.05
6	26.63	-1.81	-2.12	10.81	27.12	-1.78	-2.26	10.27	0.49	0.03	0.14	0.54
7	25.56	-2.05	-2.63	11.91	25.23	-1.97	-2.54	12.04	0.33	0.08	0.09	0.13
8	33.73	-2.34	-2.60	6.38	33.52	-2.40	-2.34	6.33	0.21	0.06	0.26	0.05
9	28.76	-1.42	-1.99	8.79	29.79	-1.47	-1.53	8.42	1.03	0.05	0.46	0.37
10	30.74	-3.47	-1.53	8.32	30.63	-3.37	-1.30	8.48	0.11	0.10	0.23	0.16
11	32.19	-3.28	-2.89	7.58	32.67	-3.02	-2.77	7.30	0.48	0.26	0.12	0.28

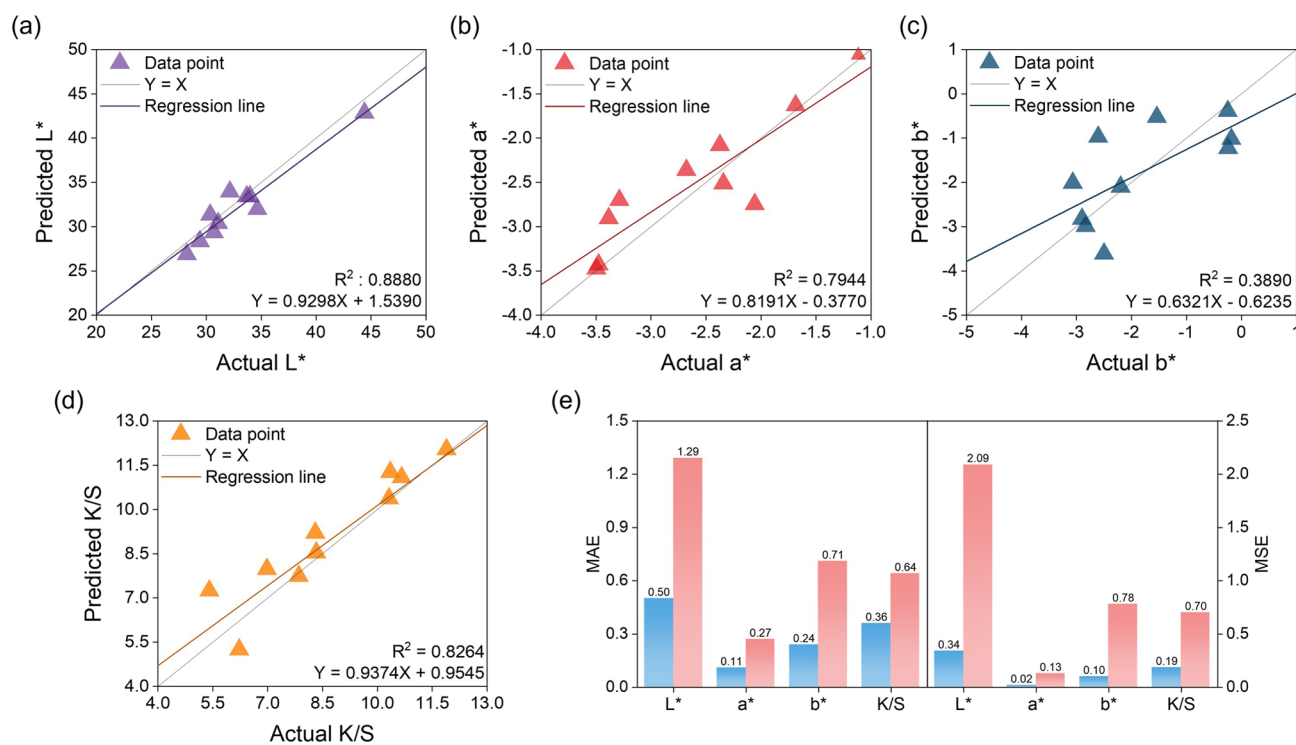


Figure 9. Linear regression analysis of ANN predictions with (a) L^* values, (b) a^* values, (c) b^* values, (d) K/S values, and (e) comparison of MAE and MSE values between ANN and GPR models.

actual experimental values in a three-dimensional space. The comparison results are presented in Figure 8.

Figure 8(a) shows the GPR model's predicted color coordinates for the test samples in a 3D Lab space, while Figure 8(b) plots the actual experimental color values for those samples. The points in (a) and (b) are clustered in similar regions, indicating that the GPR model effectively captures the real color distribution. In Figure 8(c), the predicted and actual points are superimposed, and their close proximity suggests that the differences are minimal. Figure 8(d) visualizes the absolute errors of each test sample, showing that most points are concentrated near the origin. This result confirms that the differences between predicted and actual values are small, demonstrating the model's overall high accuracy. Table 4 summarizes the experimental values, predicted values, and corresponding absolute errors for each test sample.

For the L^* , the mean absolute error was approximately 0.50. The a^* , b^* , and K/S also showed low average errors of around

0.11, 0.24, and 0.36, respectively. These values indicate color differences that are difficult to distinguish visually, confirming the high predictive performance of the model.

3.6.2. Comparison with the ANN Model. To compare the predictive performance of the GPR model with that of another model using a limited data set, an ANN model was constructed based on the same experimental data. The ANN model was trained on 80% of the data and tested on the remaining 20%. The network consisted of an input layer, two hidden layers (with 64 and 32 neurons), and an output layer with 4 neurons to predict L^* , a^* , b^* , and K/S values simultaneously. Rectified Linear Unit (ReLU) was applied to the hidden layers, and the network was trained using the Adam optimizer. Figure 9 summarizes the ANN's performance.

Figure 9(a), (b), and (d) show the linear regression analysis results for the ANN's predictions of L^* , a^* , and K/S values, respectively. While the predicted values generally align with the actual values, the deviations from the regression line are larger

compared to those of the GPR model, and several outliers can be observed. Figure 9(c) shows the ANN's predictions for b^* values, which appear less accurate than the other three outputs and exhibit the greatest dispersion among the data points. Figure 9(e) compares the MAE and MSE of the ANN and GPR models. The GPR model achieved lower error values on all metrics compared to the ANN. These results suggest that GPR is more robust and achieves better generalization performance than ANN for regression tasks with a small sample size. As a Bayesian regression model, GPR naturally guards against overfitting and provides an uncertainty estimate with each prediction, quantifying the confidence in the output values. This capability makes GPR a practical choice for predictive modeling on experimental data, especially in cases of limited sample sizes. Similar trends have been reported in previous studies, and the results of this study are consistent with such findings.^{50–53} The ANN model underperformed compared to the GPR model in this study. This was mainly due to the limited size of the data set. However, with a larger amount of data, ANN may offer advantages in terms of scalability and training efficiency. Therefore, it is important to select an appropriate model depending on the size of the data set and the specific application context.

4. CONCLUSION

In this study, a GPR model was constructed using 52 experimental samples to predict the relationship between dyeing process variables (dyeing temperature, time, and concentration) and the resulting color coordinates (L^* , a^* , and b^*) and K/S value. To improve predictive performance, multiple kernel functions were compared via cross-validation, and the best-performing kernel combination was selected. The model achieved R^2 values of 0.96, 0.96, 0.73, and 0.95 for L^* , a^* , b^* , and K/S respectively, demonstrating strong explanatory power. Furthermore, for all four outputs the MAE and MSE were both below 0.5, confirming the model's high predictive performance on quantitative metrics.

One main advantage of GPR is its ability to generate posterior predictive distributions, which provide both point estimates and a measure of uncertainty. Using this feature, the model's confidence was evaluated both visually (via 95% confidence intervals) and quantitatively. Most predictions closely matched the experimental data, and the narrow 95% confidence intervals indicated a high level of certainty in those predictions. Additionally, the prediction errors for all test samples were smaller than the model's predicted standard deviations, confirming that the GPR model can reliably reproduce the experimental outcomes within its uncertainty bounds.

This study experimentally demonstrated that the GPR model can be effectively applied to predict the color coordinates of dyed fabrics based on dyeing process variables. In summary, the results of this study establish GPR as a promising approach for achieving high predictive accuracy even with very limited data. By reducing the need for iterative dyeing experiments, this approach can potentially cut down wastewater generation and the consumption of water, energy, and chemicals, contributing to more environmentally sustainable dyeing practices. It also provides a practical foundation for efficient quality control and process design, with potential applications in smart dyeing systems, real-time control in manufacturing, digital twin-based process control, and eco-friendly textile product development. Applying the GPR model

directly on the factory floor can facilitate adaptive optimization, thereby enhancing sustainability and productivity in large-scale dyeing operations.

AUTHOR INFORMATION

Corresponding Author

Seung Geol Lee – Department of Materials Science and Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea; Graduate School of Semiconductor Materials and Devices Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea; orcid.org/0000-0001-7965-7387; Email: seungeol.lee@unist.ac.kr

Author

Hyeokjun Cho – Department of Materials Science and Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsomega.5c06345>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the Material and Component Development (No. 20024150) funded by the Korea Evaluation Institute of Industrial Technology. This study was supported by the Korea Research Institute of Chemical Technology (KK2551-30).

REFERENCES

- (1) Cui, H.; Zhu, Z.; Chen, F.; Han, Y.; Li, R. Performance on the Forward Osmosis Membrane with Polyethersulfone Supported Layer Modified by Sulfonated Carbon Nanotubes for Dyeing Wastewater Treatment. *J. Environ. Chem. Eng.* **2025**, *13*, 117303.
- (2) Chowdhury, M.; Babu, M. S.; Hossain, S.; Mia, R.; Kabir, S. M. M. Optimizing textile dyeing and finishing for improved energy efficiency and sustainability in fleece knitted fabrics. *Cleaner Energy Syst.* **2024**, *9*, 100154.
- (3) Wang, K.; Luo, Q.; Ding, W.; Zhang, Q.; Ji, D.; Wang, R.; Qin, X. Nitrogen-mediated dyeing of denim with natural indigo: Towards sustainable and efficient coloration. *Dyes Pigm.* **2025**, *239*, 112778.
- (4) Sahin, S.; Demir, C.; Güçer, S. Simultaneous UV–vis spectrophotometric determination of disperse dyes in textile wastewater by partial least squares and principal component regression. *Dyes Pigm.* **2007**, *73* (3), 368–376.
- (5) Villar, L.; Pita, M.; González, B.; Sánchez, P. B. Hydrolytic-Assisted Fractionation of Textile Waste Containing Cotton and Polyester. *Fiber Polym.* **2024**, *25* (7), 2763–2772.
- (6) Selvaraj, V.; Karthika, T. S.; Mansiya, C.; Alagar, M. An over review on recently developed techniques, mechanisms and intermediate involved in the advanced azo dye degradation for industrial applications. *J. Mol. Struct.* **2021**, *1224*, 129195.
- (7) Ghaffar, A.; Mehdi, M.; Otho, A. A.; Tagar, U.; Hakro, R. A.; Hussain, S. Electrospun silk nanofibers for numerous adsorption-desorption cycles on Reactive Black 5 and reuse dye for textile coloration. *J. Environ. Chem. Eng.* **2023**, *11* (6), 111188.
- (8) Orkhonbaatar, Z.; Lee, D. W.; Prabhakar, M. N.; Song, J. I. Upcycling Recycled PET Fibers for High-Performance Polypropylene Composites: Innovations in Sustainable Material Design. *Fiber Polym.* **2025**, *26*, 4471–4486, DOI: [10.1007/s12221-025-01086-w](https://doi.org/10.1007/s12221-025-01086-w).

- (9) Adeyemo, A. A.; Adeoye, I. O.; Bello, O. S. Adsorption of dyes using different types of clay: a review. *Appl. Water Sci.* **2017**, *7*, 543–568.
- (10) Dhage, N. D.; Landage, S. M.; Mali, A. R. Eco-Friendly Textile Dyeing: Employing Dendrobium Sonia and Senna Auriculata Extract as Sustainable Alternative to Synthetic Dyes and Mordants. *Fiber Polym.* **2025**, *26*, 4333–4344, DOI: 10.1007/s12221-025-01082-0.
- (11) Uddin, F. Environmental hazard in textile dyeing wastewater from local textile industry. *Cellulose* **2021**, *28* (17), 10715–10739.
- (12) Pervez, M. N.; Yeo, W. S.; Lin, L.; Xiong, X.; Naddeo, V.; Cai, Y. Optimization and prediction of the cotton fabric dyeing process using Taguchi design-integrated machine learning approach. *Sci. Rep.* **2023**, *13* (1), 12363.
- (13) Yu, C.; Xi, Z.; Lu, Y.; Tao, K.; Yi, Z. LSSVM-based color prediction for cotton fabrics with reactive pad-dry-pad-steam dyeing. *Chemom. Intell. Lab. Syst.* **2020**, *199*, 103956.
- (14) Cho, H.; Lee, J. E.; Kim, A. R.; Kang, Y. J.; Song, S. H.; Sim, J. H.; Lee, S. G. Predicting Dyeing Properties and Light Fastness Rating of Recycled PET by Artificial Neural Network. *Fiber Polym.* **2024**, *25* (9), 3493–3502.
- (15) Vadood, M.; Haji, A. A hybrid artificial intelligence model to predict the color coordinates of polyester fabric dyed with madder natural dye. *Expert Syst. Appl.* **2022**, *193*, 116514.
- (16) Xie, Y. T.; Zhang, H.; Zhang, S. J.; Xiao, S. L.; Li, Q.; Qin, X. N. A Data-Driven Approach for Predicting Industrial Dyeing Recipes of Polyester Fabrics. *Fiber Polym.* **2024**, *25* (8), 2985–2991.
- (17) Mahato, K. D.; Kumar, U. Optimized Machine learning techniques Enable prediction of organic dyes photophysical Properties: Absorption Wavelengths, emission Wavelengths, and quantum yields. *Spectrochim. Acta, Part A* **2024**, *308*, 123768.
- (18) Chen, J.; Lin, Y.; Liu, Y. Fast prediction of optimal reaction conditions and dyeing effects of natural dyes on silk fabrics by lightweight integrated learning (XGBoost) models. *Color. Technol.* **2025**, *141* (2), 223–237.
- (19) Yu, C.; Cao, W.; Liu, Y.; Shi, K.; Ning, J. Evaluation of a novel computer dye recipe prediction method based on the pso-lssvm models and single reactive dye database. *Chemom. Intell. Lab. Syst.* **2021**, *218*, 104430. DOI: 10.1016/j.chemolab.2021.104430.
- (20) Senthilkumar, M. Modelling of CIELAB values in vinyl Sulphone dye application using feed-forward neural networks. *Dyes Pigment.* **2007**, *75* (2), 356–361.
- (21) Li, F.; Chen, C.; Mao, Z. A novel approach for recipe prediction of fabric dyeing based on feature-weighted support vector regression and particle swarm optimization. *Color. Technol.* **2022**, *138* (5), 495–508.
- (22) Yildirim, P.; Birant, D.; Alpyildiz, T. Data mining and machine learning in textile industry. *WIREs Data Min. Knowl. Discovery* **2018**, *8* (1), No. e1228.
- (23) Ingle, N.; Jasper, W. J. A review of deep learning and artificial intelligence in dyeing, printing and finishing. *Text. Res. J.* **2025**, *95* (5–6), 625–657.
- (24) An, F.; Cao, S.; Ma, S.; Shu, D.; Han, B.; Li, W.; Liu, R. Causality-inspired surface defect detection by transferring knowledge from natural images. *Eng. Appl. Artif. Intell.* **2025**, *142*, 109984.
- (25) Xu, P.; Ji, X.; Li, M.; Lu, W. Small data machine learning in materials science. *npj Comput. Mater.* **2023**, *9* (1), 42.
- (26) Zhang, J.; Xiao, C.; Yang, W.; Liang, X.; Zhang, L.; Wang, X.; Dai, R. Improving prediction of groundwater quality in situations of limited monitoring data based on virtual sample generation and Gaussian process regression. *Water Res.* **2024**, *267*, 122498.
- (27) Vuckovic, K.; Prakash, S. Remaining useful life prediction using gaussian process regression model. In *Annual Conference of the PHM Society*; PHN Society, 2022; Vol. 14.
- (28) Liu, C.; Duan, Z.; Zhang, B.; Zhao, Y.; Yuan, Z.; Zhang, Y.; Wu, Y.; Jiang, Y.; Tai, H. Local Gaussian process regression with small sample data for temperature and humidity compensation of polyaniline-cerium dioxide NH₃ sensor. *Sens. Actuators, B* **2023**, *378*, 133113.
- (29) Korolev, V.; Nevolin, I.; Protsenko, P. A universal similarity based approach for predictive uncertainty quantification in materials science. *Sci. Rep.* **2022**, *12* (1), 14931.
- (30) Li, J.; Wang, H. Gaussian Processes Regression for Uncertainty Quantification: An Introductory Tutorial. *arXiv preprint arXiv:2502.03090* **2025**, DOI: 10.48550/arXiv.2502.03090.
- (31) Fiedler, C.; Scherer, C. W.; Trimpe, S. Practical and rigorous uncertainty bounds for Gaussian process regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*; Association for the Advancement of Artificial Intelligence, 2021; Vol. 35, pp 7439–7447, DOI: 10.1609/aaai.v35i8.16912.
- (32) Chang, C.; Zeng, T. A hybrid data-driven-physics-constrained Gaussian process regression framework with deep kernel for uncertainty quantification. *J. Comput. Phys.* **2023**, *486*, 112129.
- (33) Shen, Y.; Zhang, W.; Wang, J.; Feng, C.; Qiao, Y.; Sun, C. A boom damage prediction framework of wheeled cranes combining hybrid features of acceleration and Gaussian process regression. *Measurement* **2023**, *221*, 113401.
- (34) Chen, Y.; Dong, Z.; Su, J.; Wang, Y.; Han, Z.; Zhou, D.; Zhao, Y.; Bao, Y. Framework of airfoil max lift-to-drag ratio prediction using hybrid feature mining and Gaussian process regression. *Energy Convers. Manage.* **2021**, *243*, 114339.
- (35) Morita, Y.; Rezaeiravesh, S.; Tabatabaei, N.; Vinuesa, R.; Fukagata, K.; Schlatter, P. Applying Bayesian optimization with Gaussian process regression to computational fluid dynamics problems. *J. Comput. Phys.* **2022**, *449*, 110788.
- (36) Maier, M.; Rupenyan, A.; Bobst, C.; Wegener, K. Self-optimizing grinding machines using Gaussian process models and constrained Bayesian optimization. *Int. J. Adv. Manuf. Technol.* **2020**, *108*, 539–552.
- (37) Zagorowska, M.; Degner, M.; Ortmann, L.; Ahmed, A.; Bolognani, S.; del Rio Chanona, E. A.; Mercangöz, M. Online Feedback Optimization of compressor stations with model adaptation using Gaussian process regression. *J. Process Control* **2023**, *121*, 119–133.
- (38) Niu, T.; Xu, Z.; Luo, H.; Zhou, Z. Hybrid Gaussian process regression with temporal feature extraction for partially interpretable remaining useful life interval prediction in Aeroengine prognostics. *Sci. Rep.* **2025**, *15* (1), 11057.
- (39) Bomberna, T.; Maleux, G.; Debbaut, C. Adaptive design of experiments to fit surrogate Gaussian process regression models allows fast sensitivity analysis of the input waveform for patient-specific 3D CFD models of liver radioembolization. *Comput. Methods Programs Biomed.* **2024**, *252*, DOI: 10.1016/j.cmpb.2024.108234.
- (40) Fang, W.; Zhu, Y.-C.; Cheng, Y.; Hao, Y.-P.; Richardson, J. O. Robust Gaussian Process Regression method for efficient tunneling pathway optimization: Application to surface processes. *J. Chem. Theory Comput.* **2024**, *20* (9), 3766–3778.
- (41) Luo, W.; Shen, Y.; Fu, C.; Feng, X.; Huang, Q. Exploring the CO₂ conversion activated by the dielectric barrier discharge plasma assisted with photocatalyst via machine learning. *J. Environ. Chem. Eng.* **2024**, *12* (6), 114428.
- (42) Han, Y.; Dai, W.; Zhou, L.; Guo, L.; Liu, M.; Wang, D.; Ju, Y. Predicting the adsorption capacity of geopolymers for heavy metals in solution based on machine learning. *J. Environ. Chem. Eng.* **2025**, *13* (2), 115978.
- (43) Benesty, J.; Chen, J.; Huang, Y. On the importance of the Pearson correlation coefficient in noise reduction. *IEEE Transactions on Audio, Speech, and Language Processing* **2008**, *16* (4), 757–765.
- (44) Adler, J.; Parmryd, I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry, Part A* **2010**, *77* (8), 733–742.
- (45) Akian, J.-L.; Bonnet, L.; Owhadi, H.; Savin, É. Learning “best” kernels from data in Gaussian process regression. With application to aerodynamics. *J. Comput. Phys.* **2022**, *470*, 111595.
- (46) Pan, Y.; Zeng, X.; Xu, H.; Sun, Y.; Wang, D.; Wu, J. Evaluation of Gaussian process regression kernel functions for improving groundwater prediction. *J. Hydrol.* **2021**, *603*, 126960.

(47) Shi, X.; Jiang, D.; Qian, W.; Liang, Y. Application of the Gaussian process regression method based on a combined kernel function in engine performance prediction. *ACS Omega* **2022**, *7* (45), 41732–41743.

(48) Albertani, F. E.; Thom, A. J. Modified noise kernels in Gaussian process modelling of energy surfaces. *arXiv preprint arXiv:2212.14699*, 2022, DOI: 10.48550/arXiv.2212.14699.

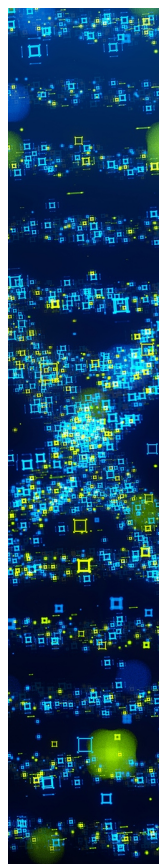
(49) Barile, C.; Carone, S.; Casavola, C.; Pappaletta, G. Implementation of Gaussian Process Regression to strain data in residual stress measurements by hole drilling. *Measurement* **2023**, *211*, 112590.

(50) Dkhili, N.; Thil, S.; Eynard, J.; Grieu, S. Comparative study between Gaussian process regression and long short-term memory neural networks for intraday grid load forecasting. In *2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*; IEEE, 2020; pp 1–6, DOI: 10.1109/EEEIC/ICPSEurope49358.2020.9160537.

(51) Zhou, Y.; Liu, Y.; Wang, D.; Liu, X. Comparison of machine-learning models for predicting short-term building heating load using operational parameters. *Energy Build.* **2021**, *253*, 111505.

(52) Jankovič, D.; Šimic, M.; Herakovič, N. A data-driven simulation and Gaussian process regression model for hydraulic press condition diagnosis. *Adv. Eng. Inf.* **2024**, *59*, 102276.

(53) Myren, S.; Lawrence, E. A comparison of Gaussian processes and neural networks for computer model emulation and calibration. *Stat. Anal. Data Min.* **2021**, *14* (6), 606–623.



CAS BIOFINDER DISCOVERY PLATFORM™

STOP DIGGING THROUGH DATA —START MAKING DISCOVERIES

CAS BioFinder helps you find the
right biological insights in seconds

Start your search

