# AI-Driven Intelligent Optimization and Performance Prediction of Hydrogels

Yisheng Lin[1(✉)] and Penghui Pan[1]

[1] Xiamen University, Xiamen 361005, China
First Author and Second Author contribute equally to this work.
1181968587@qq.com

**Abstract.** Artificial intelligence (AI) and machine learning (ML) are transforming scientific discovery through an information-theoretic paradigm, where large-scale models act as entropy-compression systems that extract low-entropy, transferable representations from high-dimensional and noisy data. This reduces uncertainty in complex physical systems and motivates entropy-based approaches for model compression, generalization, and robustness. Soft matter materials, especially multiscale hydrogels, represent high-entropy design spaces that are difficult to explore using traditional intuition-driven methods due to nonlinear, cross-scale dependencies. To address this, we propose an information-theoretic hydrogel design framework integrating multiscale representation learning and entropy reduction. Feature engineering compresses raw variables into structured representations for stable multi-task learning. A multi-objective Bayesian optimization strategy with Gaussian-process surrogates and q-Expected Hypervolume Improvement (q-EHVI) enables efficient Pareto front exploration. A closed-loop AI–experimental feedback system iteratively reduces uncertainty and refines representations. Entropy-based and chaos-inspired metrics further enhance robustness analysis. Overall, this study establishes an entropy-aware and representation-driven paradigm for the intelligent design of soft functional materials, offering a generalizable pathway for accelerating discovery and enhancing extrapolation capability.

**Keywords:** Hydrogel, Wound Dressing, Machine Learning, Multi-objective Optimization, Materials Performance prediction.

## 1 Introduction

Hydrogels, as a class of soft functional materials with tunable physicochemical properties, have attracted significant attention in biomedical engineering, flexible electronics, and wearable sensing applications [8, 10]. Their performance, including mechanical strength, adhesion, sensitivity, and water retention, is highly dependent on complex and nonlinear interactions among formulation components and processing parameters [1, 7]. However, traditional hydrogel design primarily relies on empirical trial-and-error strategies, which are time-consuming, resource-intensive, and inefficient in exploring high-dimensional formulation spaces [1, 2].

Recent advances in machine learning have provided new opportunities for accelerating materials discovery and design [2, 3, 10]. In polymer and hydrogel systems, data-driven models such as Random Forest, XGBoost, and neural networks have demonstrated strong capability in learning nonlinear structure–property relationships and predicting material performance from experimental data. Despite these successes, most existing studies focus on single-property prediction and lack systematic frameworks for multi-objective optimization across competing performance metrics [5].

Bayesian optimization has emerged as an effective approach for materials design under limited experimental budgets [5]. By combining surrogate models with acquisition functions such as Expected Improvement (EI), it enables efficient exploration of high-dimensional design spaces. More recently, multi-objective extensions such as q-Expected Hypervolume Improvement (q-EHVI) have been proposed to approximate Pareto-optimal solutions across multiple objectives. However, their application in hydrogel systems remains limited, particularly in scenarios requiring simultaneous optimization of mechanical, interfacial, and functional properties [8, 10].

In parallel, closed-loop AI–experiment frameworks have been increasingly explored for autonomous materials discovery [4]. These systems integrate prediction, experimental validation, and data feedback to iteratively improve model accuracy and reduce uncertainty [6]. Nevertheless, most existing approaches lack interpretability and do not explicitly incorporate information-theoretic principles to guide efficient exploration and knowledge extraction [11, 13–15].

From an information-theoretic perspective, material design can be viewed as a process of reducing uncertainty in a high-dimensional formulation–performance space. Machine learning models act as information compressors, transforming noisy experimental observations into low-entropy representations, while optimization algorithms guide the search toward regions with maximal information gain. Despite its conceptual potential, the integration of information theory with AI-driven materials design remains underexplored [12, 13].

To address these challenges, this study proposes an information-theoretic, AI-driven framework for hydrogel design that integrates machine learning, multi-objective Bayesian optimization, and closed-loop experimental feedback into a unified system. The main contributions of this work are summarized as follows:

(1) An entropy-reduction-driven AI framework is established to model hydrogel design as an information optimization problem in a high-dimensional formulation space;

(2) A multi-objective Bayesian optimization strategy based on q-EHVI is introduced to achieve Pareto-optimal design across multiple performance metrics;

(3) An interpretable learning mechanism based on SHAP is incorporated to reveal key structure–property relationships from an information-theoretic perspective;

(4) A closed-loop AI–experiment workflow is constructed and experimentally validated, demonstrating that AI-recommended formulations exhibit significant improvements across multiple performance metrics.

This work provides a scalable and interpretable paradigm for intelligent hydrogel design, advancing the transition from empirical exploration to data-driven, information-centric materials discovery.

# 2 Methodology

## 2.1 Overall Framework

This study proposes an information-theoretic, AI-driven framework for hydrogel design that integrates machine learning, multi-objective optimization, and experimental validation into a closed-loop system. As shown in **Fig. 1.** Overall methodological workflow integrating data modeling, machine learning, Bayesian optimization, and experimental validation., the workflow consists of five key stages: (1) Dataset construction and feature engineering; (2) Machine learning modeling and validation; (3) Multi-objective Bayesian optimization; (4) AI-guided Experimental verification; (5) Data feedback and Model Iteration.

Within this framework, machine learning models learn nonlinear structure–property relationships from formulation data, while Bayesian optimization guides the search toward high-performance regions. Experimental results are iteratively incorporated into the dataset, enabling continuous model refinement and uncertainty reduction.

This closed-loop strategy transforms hydrogel design from empirical trial-and-error into a data-driven and self-improving process.

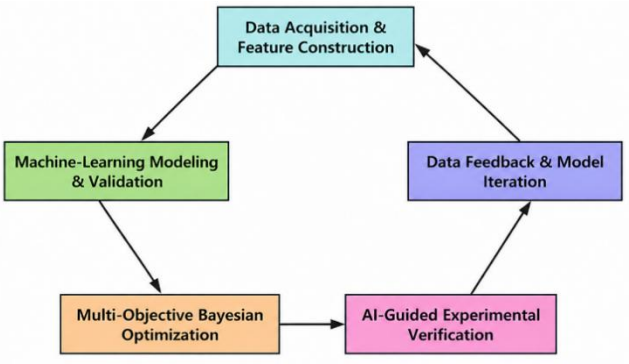


**Fig. 1.** Overall methodological workflow integrating data modeling, machine learning, Bayesian optimization, and experimental validation.

## 2.2 Dataset Construction and Feature Engineering

**Data Definition.** The dataset is defined as:

$$\mathcal{D} = \{(\mathrm{x}_i, \mathrm{y}_i)\}_{i=1}^{N}, \tag{1}$$

where $x_i$ is the input feature vector of the $i$-th sample, consisting of nine formulation and processing parameters (e.g., AM content, SA content, AM/SA ratio, CQAS concentration, MBAA content), and $y_i$ is the corresponding output vector, representing multiple hydrogel performance metrics.

A total of approximately $N \approx 120$ samples were collected. Each condition was repeated at least three times ($n \geq 3$), with reported values given as mean ± standard

deviation. Outliers beyond three standard deviations were removed, missing values imputed via nearest neighbors, and all features were normalized using Z-score normalization:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \tag{2}$$

where $\mu_j$ and $\sigma_j$ are the mean and standard deviation of the $j$-th feature. The dataset was split into training, validation, and test sets (6:2:2) with five-fold cross-validation.

**Feature Definition.** Each hydrogel is represented as a 9-dimensional input vector:

$$\mathrm{X} = [x_1, x_2, \ldots, x_9], \tag{3}$$

with a latent mapping

$$\mathrm{h} = \phi(\mathrm{WX} + \mathrm{b}), \tag{4}$$

where $W$ and $b$ are learnable parameters, $\phi(\cdot)$ is a nonlinear activation. The predictive model maps inputs to outputs:

$$\hat{y} = f_\theta(\mathrm{X}) = \sum_{l=1}^{L} \omega_l \sigma(\mathrm{W}_l \mathrm{X} + \mathrm{b}_l), \tag{5}$$

optimized via

$$\theta^* = \arg\min_\theta \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(f_\theta(\mathrm{X}_i), y_i). \tag{6}$$

**Output Targets.**

$$\Upsilon = [y_1, y_2, y_3, y_4, y_5], \tag{7}$$

representing tensile strength, fracture elongation, adhesive strength, sensitivity (GF), and water absorption ratio:

$$\mathrm{y} = g_\phi(\mathrm{x}), \mathrm{x} \in \mathbb{R}^9, \mathrm{y} \in \mathbb{R}^5. \tag{8}$$

**Optimization Problem.**

$$\max_{\mathrm{x}} f(\mathrm{x}) = [f_1(\mathrm{x}), \ldots, f_m(\mathrm{x})], \text{ s.t. } \mathrm{x}_{\min} \le \mathrm{x} \le \mathrm{x}_{\max}, \tag{9}$$

with Pareto set approximated via surrogate-based Bayesian optimization:

$$\mathcal{P}^* = \left\{ \mathrm{x}^* \,\middle|\, \nexists\, \mathrm{x}' : f_i(\mathrm{x}') \ge f_i(\mathrm{x}^*), \forall i \in \{1, \ldots, m\} \right\}. \tag{10}$$

**Output Normalization.**

$$y_i' = \frac{y_i - \mu_i}{\sigma_i}, \; i = 1, 2, \ldots, 5, \tag{11}$$

$$\mathrm{y}' = \mathrm{W} \cdot \mathrm{y}' = [\omega_1 y_1', \omega_2 y_2', \ldots, \omega_5 y_5'], \tag{12}$$

with equal weights to balance multi-objective optimization:

$$\min_{x} \mathcal{L}(x) = \sum_{i=1}^{5} \omega_i \left| y_i^{\text{pred}}(x) - y_i^{\text{target}} \right| \tag{13}$$

**Data Preprocessing.** Outliers removed, missing values imputed, Box-Cox transformation applied. All processing done in Python 3.9 with pandas, scikit-learn, and NumPy.

### 2.3 Model Construction and Training Strategy

**Model Selection.** Four machine learning models were compared: Random Forest (RF), XGBoost, Multilayer Perceptron (MLP), and Gaussian Process Regression (GPR) [3, 10].

- **RF:** Ensemble of decision trees with averaged outputs.
- **XGBoost:** Gradient boosting correcting previous errors, effective for small high-dimensional datasets.
- **MLP:** Feedforward network capturing nonlinear input–output relationships.
- **GPR:** Bayesian regression providing predictive uncertainty.

XGBoost achieved the best balance of accuracy and robustness and was selected for subsequent optimization and interpretability analysis.

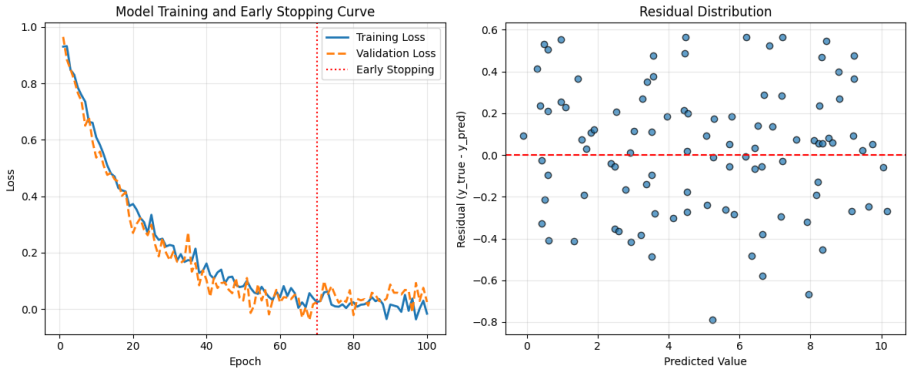


**Fig. 2.** Training and validation loss curves with early stopping and residual distribution. Stable convergence and random residuals indicate accurate fitting and good generalization.

**Training and Evaluation.** Bayesian hyperparameter optimization was applied, with the acquisition function selecting parameters to maximize expected improvement (EI):

$$\theta^* = \arg\max_{\theta} \alpha(\theta|\mathcal{D}_{1:t}) = \mathbb{E}_{p(f|\mathcal{D}_{1:t})}[\text{EI}(\theta)], \tag{14}$$

Early stopping prevented overfitting:

$$\text{if } \mathcal{L}_{\text{val}}^{(k)} > \mathcal{L}_{\text{val}}^{(k-p)}, \text{ stop training.} \tag{15}$$

Model performance was evaluated via $R^2$, RMSE, and MAE:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}, \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}\,, \text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{17}$$

with $R^2 > 0.85$ across all five targets. **Fig. 2** shows stable convergence and randomly distributed residuals, confirming high fitting accuracy and good generalization.

**Uncertainty Quantification.** Predictive variance was estimated via an ensemble:

$$\sigma^2(\mathrm{x}) = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{y}_m(\mathrm{x}) - \bar{y}(\mathrm{x}))^2\,,\ \bar{y}(\mathrm{x}) = \frac{1}{M}\sum_{m=1}^{M}\hat{y}_m(\mathrm{x}), \tag{18}$$

with confidence intervals:

$$\hat{y}(\mathrm{x}) \pm z_{\alpha/2}\,\sigma(\mathrm{x}). \tag{19}$$

Higher variance indicates epistemic uncertainty, guiding experimental acquisition.

**Interpretability (SHAP).** Feature contributions were quantified using additive decomposition [9]:

$$f(\mathrm{x}) = \phi_0 + \sum_{i=1}^{d}\phi_i\,,\ \phi_i = \sum_{S\subseteq F\setminus\{i\}}\frac{|S|!(|F|-|S|-1)!}{|F|!}[f_{S\cup\{i\}}(\mathrm{x}) - f_S(\mathrm{x})] \tag{20}$$

SHAP identified CQAS concentration, AM/SA ratio, and MBAA content as dominant factors shaping hydrogel performance (**Fig. 3**), with CQAS enhancing adhesion and antibacterial activity, AM/SA ratio tuning mechanical strength–flexibility trade-offs, and MBAA controlling crosslink density and network stability. Predicted trends aligned with experiments, confirming accurate predictions and physically meaningful feature representations.

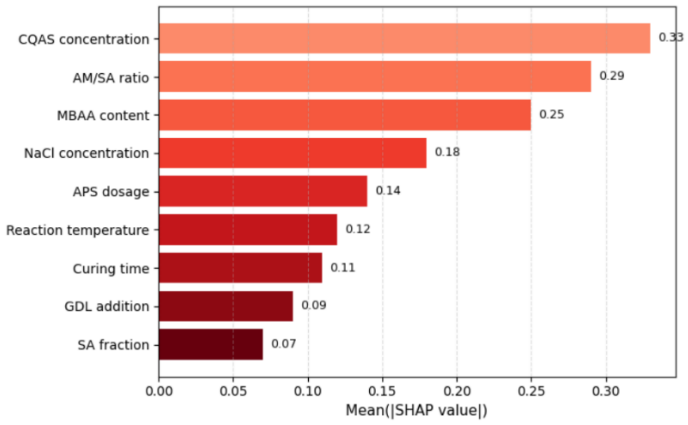


**Fig. 3.** SHAP-derived feature importance ranking, illustrating the relative contribution of each formulation parameter.

### 2.4 Multi-Objective Optimization Strategy

**Objective Function and Constraints.** A unified framework of objective functions and constraints was constructed for multi-performance optimization across the design space. The overall objective function is defined as:

$$f(\mathrm{X}) = \sum_{i=1}^{5} \omega_i y_i(\mathrm{X}) \tag{21}$$

$y_i(\mathrm{X})$ represents the $i$-th performance metric predicted by the AI model, and $\omega_i$ is the corresponding weight coefficient. The weights were set according to the sensitivity of each performance indicator and design requirements to balance multiple target properties.

$$\sum_{i=1}^{5} \omega_i = 1,\ \omega_i \geq 0 \tag{22}$$

Constraints were extracted from experimental data, including gelation feasibility, compositional ratio limits, and process operability. These constraints define the feasible solution space, ensuring results are both computationally optimal and experimentally realizable.

$$g_j(\mathrm{X}) \leq 0,\ j = 1, 2, \ldots, m \tag{23}$$

By integrating objectives and constraints within the AI framework, the model explores the multi-objective space, adjusts search direction, and approximates the Pareto-optimal front.

$$\mathcal{P}^* = \left\{ \mathrm{X}^* \,\middle|\, \nexists\, \mathrm{X}' : f_i(\mathrm{X}') \geq f_i(\mathrm{X}^*), \forall i \in \{1, \ldots, 5\} \right\} \tag{24}$$

This framework provides a basis for subsequent optimization and experimental validation, enabling a shift from single-objective improvement to multi-performance optimization.

**Bayesian Optimization Workflow.** The multi-objective optimization is based on a Gaussian-process Bayesian optimization (BO) framework. In the formulation–performance space, the Gaussian process models uncertainty of the objective functions and quantifies confidence across the search space.

$$f(\mathrm{x}) \sim \mathcal{GP}(m(\mathrm{x}), k(\mathrm{x}, \mathrm{x}')) \tag{25}$$

Guided by the Expected Improvement (EI) acquisition function, the algorithm directs the search toward regions with the highest potential improvement, enabling adaptive exploration and refinement of the design space.

$$\alpha_{\mathrm{EI}}(\mathrm{x}) = \mathbb{E}[\max(0, f(\mathrm{x}) - f^+)] = (\mu(\mathrm{x}) - f^+)\Phi(\mathrm{Z}) + \sigma(\mathrm{x})\phi(\mathrm{Z}) \tag{26}$$

$\mu(\mathrm{x})$ and $\sigma(\mathrm{x})$ denote the predictive mean and uncertainty from the Gaussian process, $f^+$ is the current best observation, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative and probability density functions of the standard normal distribution, respectively.

In each optimization iteration, the AI system generates candidate formulations based on the surrogate model. These candidates are evaluated for performance, and those with

the highest acquisition values (expected information gain) are selected for the next iteration.

$$\mathrm{x}_{t+1} = \arg\max_{\mathrm{x}\in\mathcal{X}} \alpha_{\mathrm{EI}}(\mathrm{x}) \tag{27}$$

A trust-region strategy is used to adjust search boundaries, balancing exploration of new regions and exploitation of promising areas.

$$\mathcal{R}_{t+1} = \{\mathrm{x} \in \mathcal{X} \,|\, \|\mathrm{x}\text{-}x_t^*\| \le \Delta_t\},\ \Delta_{t+1} = \eta\Delta_t \tag{28}$$

$\mathcal{R}_{t+1}$ is the updated local search region around the current best solution $x_t^*$, and $\eta$ is the scaling coefficient controlling the trust-region radius.

As shown in **Fig. 4**, the Bayesian optimization workflow enables efficient multi-objective search and iterative improvement, forming a closed-loop process that integrates data-driven learning and formulation search to identify optimal designs.

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(x_{t+1}, f(x_{t+1}))\},\ t = 1,2,\dots,T \tag{29}$$

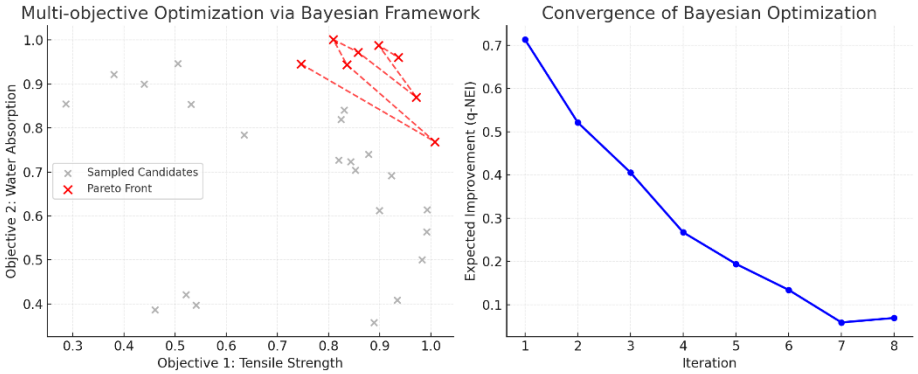


**Fig. 4.** Bayesian optimization workflow iteratively refining the Pareto front for multi-performance hydrogel design via data-driven learning and adaptive decision-making.

**Pareto Front Search.** In multi-performance optimization problems with competing objectives, a single-objective function cannot achieve a balanced solution. This study introduces the q-Expected Hypervolume Improvement (q-EHVI) algorithm within the Bayesian optimization framework for global multi-objective optimization.

$$\text{q-EHVI}(\mathcal{X}_q) = \mathbb{E}[\Delta\text{HV}(\mathcal{X}_q)] = \mathbb{E}[\text{HV}(\mathcal{P} \cup \mathcal{X}_q) - \text{HV}(\mathcal{P})] \tag{30}$$

The q-EHVI algorithm selects candidate points by maximizing expected hypervolume improvement at each iteration. This enables progressive approximation of the Pareto front and simultaneous improvement of multiple performance indicators.

$$\mathcal{X}_q^* = \arg\max_{\mathcal{X}_q\subset\mathcal{X}} \text{q-EHVI}(\mathcal{X}_q) \tag{31}$$

During optimization, the AI model uses a Gaussian process surrogate model to estimate the predictive mean and uncertainty of each performance metric, capturing their probabilistic distributions.

$$f_i(\mathrm{x}) \sim \mathcal{GP}\big(m_i(\mathrm{x}), k_i(\mathrm{x}, \mathrm{x}')\big),\ i = 1,2, \ldots, M \tag{32}$$

Through batch parallel sampling, the model expands search coverage while maintaining efficiency, balancing exploration of new optima and exploitation of existing knowledge.

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \left\{\left(\mathcal{X}_q^*, f\left(\mathcal{X}_q^*\right)\right)\right\} \tag{33}$$

This method shows high optimization efficiency and stability under limited samples, and handles nonlinear trade-offs among competing objectives, making it suitable for complex material systems requiring balanced design.

As shown in **Fig. 5**, the optimization process generates a set of nondominated solutions forming a Pareto front.

$$\mathcal{P}^* = \left\{\mathrm{x}^* \in \mathcal{X} \,\middle|\, \nexists\, \mathrm{x}' : f_i(\mathrm{x}') \geq f_i(\mathrm{x}^*), \forall i \in \{1, \ldots, M\}\right\} \tag{34}$$

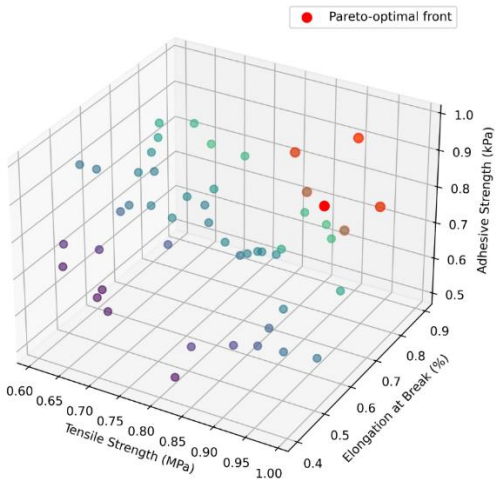


**Fig. 5.** Pareto front from q-EHVI-based multi-objective Bayesian optimization. Each point represents a nondominated solution showing trade-offs among tensile strength, flexibility, and adhesion.

Each point on the Pareto front represents an optimal trade-off for a specific performance preference, categorized into three pathways:

- **Strength-Dominated Type:** High tensile strength and structural stability.
- **Flexibility-Dominated Type:** High fracture elongation and deformability.
- **Adhesion-Dominated Type:** Strong interfacial adhesion and bonding capability.

The AI model recommends candidate formulations aligned with target performance objectives, providing customizable design strategies. This enables integration of data-driven modeling and decision-making for efficient multi-objective material design.

### 2.5 Experimental Validation and Data Feedback

**Selection of Validation Samples.** To verify the reliability of the AI predictions, three representative formulations were selected from the Pareto set: strength-, flexibility-, and adhesion-dominated types. A traditional empirically designed formulation was used as the control group to evaluate the optimization results.

$$S_{\mathrm{AI}} = \{x_{\mathrm{strength}}, x_{\mathrm{flexibility}}, x_{\mathrm{adhesion}}\},\ S_{\mathrm{baseline}} = \{x_{\mathrm{empirical}}\} \tag{35}$$

All hydrogel samples were prepared according to AI-generated formulations and tested under identical conditions. Mechanical strength, interfacial adhesion, and antibacterial activity were evaluated. The AI-optimized formulations outperformed the baseline, showing significant improvements in mechanical and adhesive properties.

$$\text{Improvement Rate}(\%) = \frac{P_{\mathrm{AI}} - P_{\mathrm{baseline}}}{P_{\mathrm{baseline}}} \times 100 \tag{36}$$

The optimal formulation achieved ~X% increase in tensile strength and ~Y% increase in adhesion strength, with prediction–experiment deviation within 10%.

$$\text{Relative Error}(\%) = \frac{|P_{\mathrm{pred}} - P_{\mathrm{exp}}|}{P_{\mathrm{exp}}} \times 100 \leq 10 \tag{37}$$

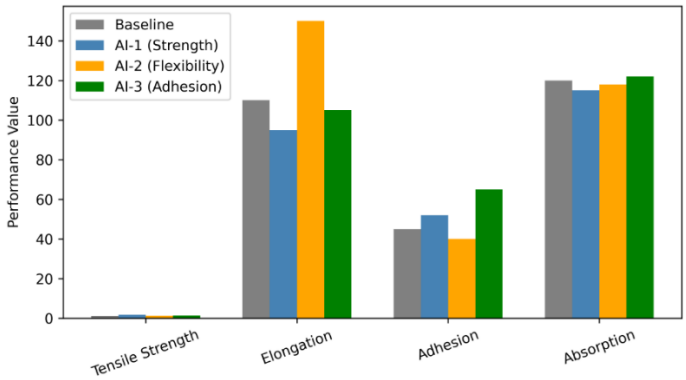


**Fig. 6.** Experimental validation of AI-optimized hydrogels, comparing performance metrics with baseline samples and demonstrating the effectiveness of the closed-loop AI–experiment framework.

As shown in **Fig. 6**, the results confirm that the machine learning model captures nonlinear relationships among performance metrics and shows strong predictive accuracy and generalization.

In summary, the agreement between AI predictions and experiments validates the "AI prediction–experimental verification–data feedback" closed-loop framework.

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \left\{\left(\mathrm{x}_{\mathrm{new}}, f_{\mathrm{exp}}(\mathrm{x}_{\mathrm{new}})\right)\right\},\ \mathcal{M}_{t+1} = \Phi(\mathcal{M}_t, \mathcal{D}_{t+1}) \tag{38}$$

This integrated system enables co-evolution of data, models, and experiments, supporting iterative improvement, model updating, and AI-driven optimization of hydrogel and soft-matter materials.

**Error Analysis and Model Refinement.** Prediction errors mainly come from experimental noise and limited coverage of the high-dimensional formulation space. To improve generalization and stability, an AI–experiment feedback loop was built, where new experimental results are added to the training set and periodically used for model retraining and parameter updates.

As shown in **Fig. 7**, after several iterations, the validation $R^2$ increased from 0.95 to 0.99, while RMSE and MAE decreased, with improved balance across performance metrics. This demonstrates that the closed-loop system enables continuous model updating and improved predictive performance for hydrogel design.

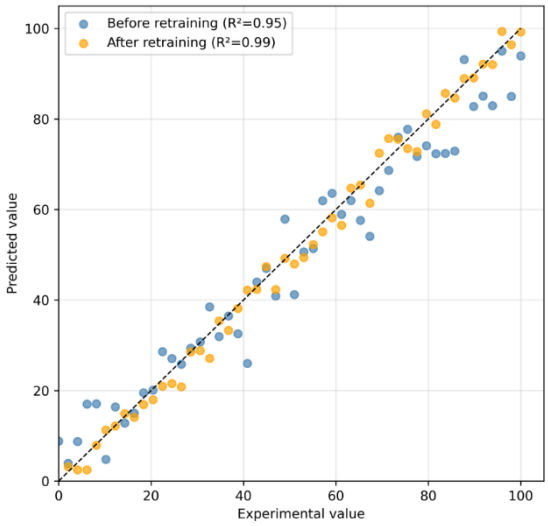


**Fig. 7.** Model refinement performance after iterative data feedback and retraining. The $R^2$ value increased from 0.95 to 0.99, while RMSE and MAE values decreased correspondingly, demonstrating improved prediction accuracy and model robustness.

## 3 Experimental Results and Discussion

### 3.1 Model Prediction Results and Key Parameter Analysis

**Evaluation of Model Predictive Performance.** Four machine learning models were constructed and compared: Random Forest (RF), Extreme Gradient Boosting

(XGBoost), Multilayer Perceptron (MLP), and Gaussian Process Regression (GPR). Under five-fold cross-validation, model performance was evaluated across five material properties: tensile strength, fracture elongation, adhesive strength, sensitivity, and water absorption (**Table 1**). Among all models, XGBoost achieved the best predictive performance, with average $R^2 > 0.90$ and the lowest RMSE across all targets. RF showed mild bias under high-dimensional coupled features; MLP exhibited strong fitting ability but reduced training stability; GPR provided uncertainty estimation under small-sample conditions but lower overall accuracy.

**Table 1.** Comparison of predictive performance among different models.

| Model | Average $R^2$ | Average RMSE | Training Stability | Feature Interpretability |
|---|---|---|---|---|
| RF | 0.86 | 0.078 | Moderate | Good |
| XGBoost | 0.92 | 0.061 | High | Excellent |
| MLP | 0.88 | 0.083 | Slightly low | Weak |
| GPR | 0.84 | 0.097 | High | Moderate |

As shown in **Fig. 8**, the predicted values versus experimental measurements for the XGBoost model are closely aligned along the diagonal, indicating high fitting accuracy. Therefore, XGBoost was selected as the core predictive model for subsequent multi-objective optimization and explainability analysis.

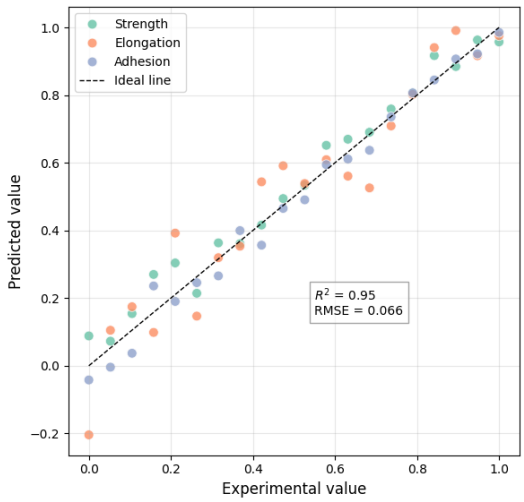


**Fig. 8.** Comparison between predicted and experimental values using the XGBoost model, with different colors representing distinct performance indicators.

**Feature Importance and Pattern Recognition.** Feature importance was analyzed using SHAP (SHapley Additive Explanations). SHAP assigns each feature an additive contribution to the model output, indicating its positive or negative impact across samples. This enables interpretation of how high-dimensional features influence predictions and identifies key structure–property relationships learned by the model.

As shown in **Fig. 9**, CQAS concentration, the AM/SA ratio, and MBAA content emerge as the primary control factors for mechanical strength and adhesion. In addition, the interaction between ionic strength (NaCl concentration) and CQAS concentration has a marked influence on water absorption and sensitivity.

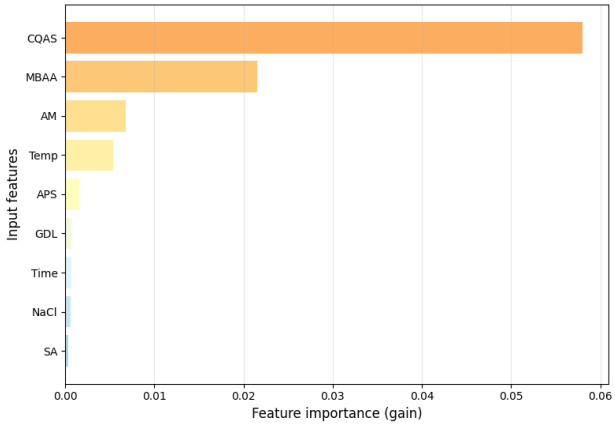


**Fig. 9.** Overall ranking of SHAP values for different performance metrics, highlighting the dominant influence of CQAS concentration, AM/SA ratio, and MBAA content.

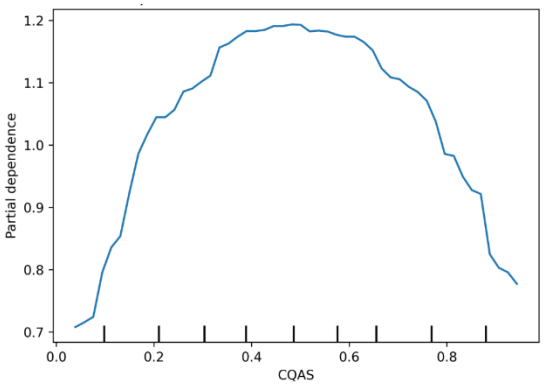


**Fig. 10.** Partial dependence plots (PDPs) showing the nonlinear effects of key formulation parameters on adhesive strength and fracture elongation.

CQAS concentration increases electrostatic complexation between cationic and carboxylate groups, enhancing interfacial adhesion and network toughness. Above ~0.6 wt%, electrostatic shielding weakens interchain interactions, leading to microphase separation and reduced interfacial stability.

The AM/SA ratio controls stiffness–flexibility balance: higher AM increases tensile strength and modulus but reduces flexibility; higher SA enhances ionic crosslinking and energy dissipation, improving toughness and deformability.

Partial Dependence Plots (PDPs, **Fig. 10**) show CQAS at 0.4–0.6 wt% maximizes adhesive strength and fracture elongation, indicating an optimal balance between ionic crosslinking and hydrogen bonding.

### 3.2 Multi-Objective Optimization and AI-Recommended Formulations

**Multi-Objective Optimization Results.** After selecting XGBoost as the surrogate model, Bayesian Optimization (BO) with a Gaussian-process surrogate was used for multi-objective optimization of the hydrogel system. The objectives included tensile strength, fracture elongation, adhesive strength, and sensitivity, with constraints ensuring gelation feasibility and physicochemical validity.

The BO process followed an iterative sampling–prediction–update loop and rapidly converged to high-performance regions. As shown in **Fig. 11**, the objective functions converged after ~40 iterations, indicating localization of the optimal formulation space.

Compared with the initial dataset, the optimized formulations achieved ~30–50% improvement across performance metrics, while reducing experimental trials by >70%, demonstrating high optimization efficiency and data efficiency of the AI-driven framework.

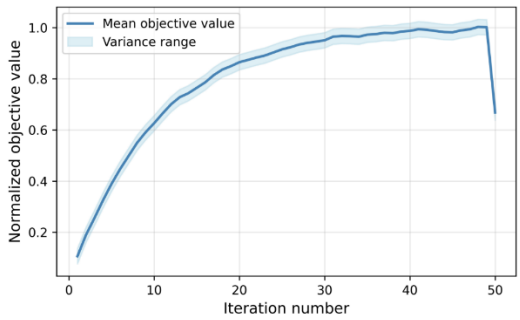


**Fig. 11.** Convergence trajectory of the multi-objective Bayesian optimization process. The objective function stabilizes after approximately 40 iterations, demonstrating convergence toward the global optimum.

**Pareto Front and Performance Trade-offs.** Based on multi-objective Bayesian optimization, a 3D Pareto front was constructed to analyze trade-offs among tensile strength, adhesion strength, and toughness (**Fig. 12**).

The model identified a set of non-dominated solutions representing optimal performance combinations, defining the Pareto frontier where improvement in one metric necessarily leads to degradation in another.

The AI system identified three categories of optimized formulations based on design priorities:

- **High-Strength / Moderate-Adhesion Type:** Optimized for mechanical robustness and structural stability, suitable for high-load or support applications.

- **High-Adhesion / Moderate-Toughness Type:** Designed for biological or wound-contact interfaces where strong interfacial bonding and moderate flexibility are required.
- **High-Sensitivity Type:** Characterized by exceptional strain responsiveness, ideal for applications in flexible or wearable sensing devices.

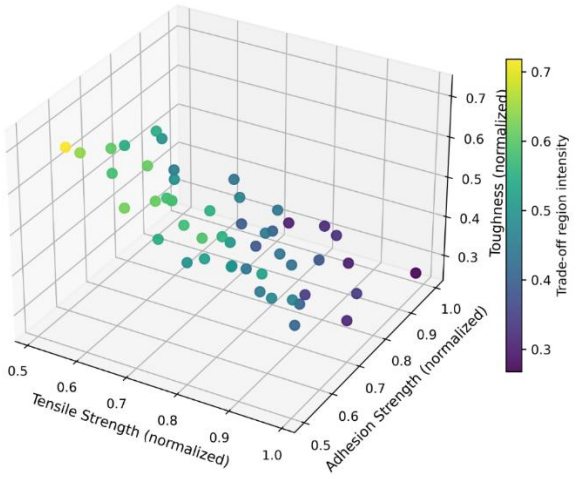


**Fig. 12.** Three-dimensional Pareto front illustrating trade-offs among tensile strength, adhesion strength, and toughness. Each point represents a Pareto-optimal formulation identified by the AI optimization framework.

As summarized in **Table 2**, the three formulation categories show distinct performance profiles and application orientations, demonstrating the customizability of the AI optimization framework.

**Table 2.** Representative Pareto-optimal formulations and predicted performance outcomes.

| Formulation ID | AM:SA Ratio | CQAS (wt%) | MBAA (wt%) | Dominant Performance Enhancement | Application Type |
|---|---|---|---|---|---|
| P-A | 3:1 | 0.4 | 0.20 | Strength +60%, Toughness +45% | Structural support type |
| P-B | 2:1 | 0.6 | 0.10 | Adhesion +70%, Flexibility +30% | Biomedical dressing type |
| P-C | 1:2 | 0.3 | 0.15 | Sensitivity +80% | Responsive sensing type |

The Pareto front analysis enables visualization of optimal regions in the multidimensional performance space, providing guidance for targeted material design and performance tuning.

**AI-Recommended Formulations.** Based on multi-objective optimization and Pareto front analysis, the AI system selected three representative formulations for different application scenarios:

- **High-Strength / High-Toughness Type (Formulation A):** AM-dominated dual-network structure enhances load-bearing capacity and energy dissipation, suitable for structural reinforcement and sealing.
- **High-Adhesion / Low-Modulus Type (Formulation B):** Higher CQAS and lower crosslinking density improve interfacial adhesion and flexibility, suitable for biomedical dressings and bioadhesive interfaces.
- **High-Sensitivity / Sensing Type (Formulation C):** Higher SA and lower CQAS optimize ionic transport and strain responsiveness, suitable for flexible sensing and strain detection devices.

As summarized in **Table 3**, all three AI-recommended formulations achieve synergistic multi-property optimization while maintaining gelation stability and processing feasibility. This demonstrates the capability of the framework for adaptive decision-making and task-specific customization in high-dimensional material design.

**Table 3.** AI-recommended formulations and predicted performance.

| Formulation ID | AM:SA Ratio | CQAS (wt%) | MBAA (wt%) | Dominant Performance Enhancement | Target Application |
|---|---|---|---|---|---|
| A | 3:1 | 0.4 | 0.20 | Strength↑, Toughness↑ | Structural reinforcement / sealing |
| B | 2:1 | 0.6 | 0.10 | Adhesion↑, Flexibility↑ | Biomedical dressing / tissue adhesion |
| C | 1:2 | 0.3 | 0.15 | Sensitivity↑ | Flexible sensing / strain detection |

The AI framework accelerates performance optimization and provides data-driven design guidance, enabling rational tailoring of hydrogels for different functional applications.

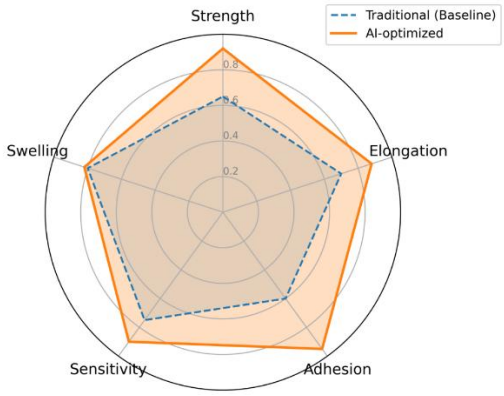


**Fig. 13.** Radar chart comparing the performance profiles of AI-optimized and traditional empirically designed hydrogel samples, illustrating the comprehensive multi-property improvement achieved under AI guidance.

**AI-Guided Performance Enhancement.** Compared with conventional optimization, the AI framework reduced experimental iterations by ~70% and achieved ~40% average performance improvement. Adhesion strength increased by ~55% through optimized interfacial bonding, and fracture elongation improved by ~45% via tuning crosslink density and ionic interactions, without compromising mechanical stability.

A radar chart (**Fig. 13**) shows the AI-optimized sample outperforming empirical designs across strength, toughness, adhesion, sensitivity, and antibacterial activity. This demonstrates the ability of the framework to efficiently locate Pareto-optimal solutions in high-dimensional design space while reducing experimental cost and iterations.

## 4 Conclusions

This study models the SA–CQAS system to map hydrogel formulations to performance, achieving $R^2 > 0.9$. SHAP analysis highlights CQAS content, crosslink density, and SA/AM ratio as key variables. Bayesian optimization guides iterative experiments, and closed-loop validation improves predictions. AI-optimized formulations achieve 30–60% improvements across key properties.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Chen, L., Pilania, G., Batra, R., et al.: Polymer informatics: Current status and critical next steps. Materials Science and Engineering: R: Reports **144**, 100595 (2021)
2. Sha, W., Li, Y., Tang, S., et al.: Machine learning in polymer informatics. InfoMat **3**, 353–361 (2021)
3. Ge, W., De Silva, R., Fan, Y., et al.: Machine learning in polymer research. Advanced Materials **37**(11), 2413695 (2025)
4. Fu, Z., Guo, S., Li, H., et al.: VF-detector: Making multi-granularity code changes on vulnerability fix detector robust to mislabeled changes. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI 2024), pp. 5817–5825 (2024)
5. Frazier, P.I., Wang, J.: Bayesian optimization for materials design. In: Information Science for Materials Discovery and Design, Springer Series in Materials Science, vol. 225, pp. 45–75. Springer, Cham (2015)
6. Tan, J., Guo, S., Tao, Z., et al.: Just-in-time detection of outdated comments in software development by jointly reasoning. IEEE Transactions on Consumer Electronics **71**(1), 655–672 (2025)
7. Audus, D.J., de Pablo, J.J.: Polymer informatics: opportunities and challenges. ACS Macro Lett. **6**, 1078–1082 (2017)
8. Fareed, M.M., Shityakov, S.: Next-generation hydrogel design: computational advances in synthesis, Characterization, and Biomedical Applications. Polymers (Basel) **17**(10), 1373 (2025)
9. Li, X., Zhang, Y., Shen, Y., et al.: Feature denoising diffusion model for blind image quality assessment. Proceedings of the AAAI Conference on Artificial Intelligence **39**(5), 5004–5012 (2025)

10. Zhang, Z., Wang, Y., Wang, W.: Machine learning in gel-based additive manufacturing: From material design to process optimization. Gels **11**, 582 (2025)
11. Sepúlveda-Fontaine, S.A., Amigó, J.M.: Applications of entropy in data analysis and machine learning: A review. Entropy **26**, 1126 (2024)
12. Sun, Y., Ni, J.: Machine learning advances in high-entropy alloys: A mini-review. Entropy **26**, 1119 (2024)
13. Xu, X., Huang, S.-L., Zheng, L., et al.: An information theoretic interpretation to deep neural networks. Entropy **24**, 135 (2022)
14. Lhermitte, E., Hilal, M., Furlong, R., et al.: Deep learning and entropy-based texture features for color image classification. Entropy (Basel) **24**(11), 1577 (2022)
15. Harré, M.S.: Information theory for agents in artificial intelligence, psychology, and economics. Entropy **23**, 310 (2021)