



RESEARCH ARTICLE

Enhancing Molecularly Imprinted Polymer Development with Machine Learning Techniques

Jisha R. C.¹, Niranjan Prasanth¹, Aswathi Perumkunnappala², Meenakshi Krishnan Praseetha Radhakrishnan², Manjusha Nair^{1,*} and Renuka Suravajhala^{3,*}

¹Amrita School of Computing, Amrita Vishwa Vidyapeetham-Amritapuri, India

²Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham-Amritapuri, India

³Department of Chemistry, Manipal University Jaipur, India

Abstract: Molecular imprinted polymers (MIPs) have various applications such as chemical sensing and drug delivery systems because they can selectively attach molecules that fit the template. The imprinting factor (IF) is used to assess the degree of contact between the binding sites on MIPs and target molecules, enabling quantification of the binding affinity for MIPs. The machine learning (ML) prediction was carried out in association with a study conducted with 5-fluorouracil, which is a chemotherapy drug used to fight cancer. ML techniques, including K-Nearest Neighbors, Random Forest, XGBoost, Decision Tree, Lasso, and Ridge Regressions, were used to predict IF values for MIP. The dataset includes 214 observations consisting of various template molecules, functional monomers, and crosslinkers. The performance of the model was evaluated and compared using root mean squared error (RMSE) values, mean absolute error (MAE), and R^2 . The Random Forest shows good performance with moderate RMSE (3.54), low MAE (2.19), and $R^2 = 0.55$, indicating acceptable error and moderate variance. A Streamlit-based web interface was developed for visualization of polymer interactions and usability of the models by researchers with similar interests. These results imply that with improved and more extensive data, we have the potential to achieve higher IF prediction accuracy with ML models. The limitations of conventional MIPs make experimental estimation of binding affinity challenging due to cost and time constraints; therefore, this study employs ML as a pre-screening tool for prediction.

Keywords: molecular imprinted polymers (MIPs), non-imprinted polymers (NIPs), imprinting factor (IF), binding affinity, machine learning algorithms

1. Introduction

Molecular imprinting is a technique in which synthetic receptor sites are developed within a polymeric matrix that can recognize and bind to specific target molecules. A brief history of molecular imprinting dates back nearly 40 years to the discovery of synthetic polymers imprinted at the molecular level, which created a sensation in this field. The conceptual foundation of this field can be traced back to Pauling's 1940 proposal of template-directed antibody formation, which introduced the idea that molecular recognition could be guided by structural complementarity [1]. This concept was translated into polymer chemistry by Wulff and Sarhan in 1972 through the first covalent molecularly imprinted polymer (MIP) synthesis [2].

Later expanded by Mosbach and co-workers in 1984 with the development of the non-covalent imprinting approach [3]. The non-covalent strategy significantly broadened the applicability of MIPs because it enabled simpler preparation and compatibility with a wider range of target molecules. Over the following decades, the field grew rapidly, with Alexander et al. documenting more than 1000 MIP-related publications within a single decade, reflecting both scientific interest and technological potential [4]. More recently, next-generation MIP systems have begun incorporating stimuli-responsive materials and artificial intelligence (AI)-guided synthesis strategies [5–8], indicating a shift from conventional empirical design toward intelligent and adaptive polymer engineering. The molecular imprinting process (MIPs) includes making polymer materials around a template molecule so that they take complementary shapes, which then enables them to selectively bind to targets. After polymerization, the template is removed, leaving behind cavities with shapes and functional groups complementary to the target molecule. MIPs are known for their high specificity and sensitivity toward their

*Corresponding authors: Manjusha Nair, Amrita School of Computing, Amrita Vishwa Vidyapeetham-Amritapuri, India. Email: manjusha.nair@am.amrita.edu and Renuka Suravajhala, Department of Chemistry, Manipal University Jaipur, India. Email: suravajhala.renuka@jaipur.manipal.edu

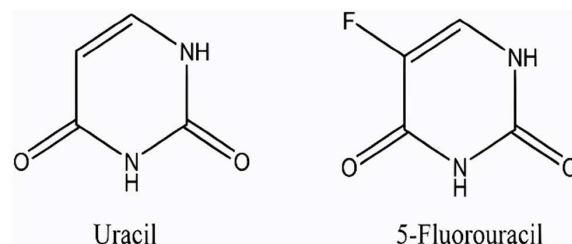
templates. This technology is used in numerous fields, such as chemical detection and separation processes, along with drug delivery systems [9, 10]. MIP has taken a huge leap forward due to radical developments and has become a major technology for material synthesis, finding wide applications in separations, assays, sensors, and optics [11]. Non-imprinted polymers (NIPs) are developed through the same procedure as MIPs but do not involve any template molecule during their production. NIPs serve as control materials in highlighting the specificity and binding affinity of MIPs, acting as a point of reference where comparisons can be made. This has helped to recognize specific binding sites within MIPs, those that involve nonspecific interactions [12, 13]. The approach has been made more effective by applying standardized synthesis protocols and exploring environmentally friendly methods to meet sustainability standards [14, 15].

Binding-site heterogeneity, template leakage, poor compatibility in physiological (aqueous) environments, and reduced binding capacity due to high crosslinking are some of the limitations of MIPs. The latest improvements in MIPs have focused on physicochemical properties to ensure homogeneity at the binding site, compatibility with water, and developing protein-tailored MIPs. This exhibits a commitment to the continuous enhancement and widening scope of an evolving technology [16]. Therefore, the proposed system focuses on physicochemical properties to develop the machine learning (ML) models.

In drug delivery, it is important to achieve precise targeting and controlled release of therapeutic drugs to enhance effectiveness and reduce potential side effects. MIPs have important advantages in this field because they selectively bind drug molecules or biomolecules related to disease targets [17]. The selectivity of MIPs allows them to act as smart carriers that can differentiate molecules with similar structures or target specific functional groups in complex biological environments [18]. MIPs can enclose drugs within their imprinted cavities [19]. As a result, drug molecules can perform in the same way that they were intended to work in the real world. Reaching their target area without any changes occurring to them. In addition to this, MIPs can respond to different external stimuli such as temperature, pH value, or certain biomarkers. These properties enable controlled release systems, which further improve the efficiency of drug delivery [20]. The strength and specificity of binding sites occurred during polymer synthesis influence the binding affinity between a template molecule and MIPs. Stronger binding can be seen in polymers with higher affinities, thus showing increased selectivity in molecular recognition, hence significantly useful across diverse fields, including chemical sensing, separation processes, and pharmaceutical science [21, 22]. An important measure, called the imprinting factor (IF), quantifies the strength of interaction between an imprinted polymer and a template molecule. This factor mirrors the specificity brought by molecular imprinting and the selective binding capacity of MIPs against their target compounds. It is counted as a significant indicator of the effectiveness of molecular recognition needed in different devices, such as sensors or drug delivery systems. According to various authors, higher IF values indicate increased affinities and specificities, thus improving the performance of these materials in general in such applications [23, 24]. The current study focused on 5-fluorouracil (5-FU), which is the synthetic analog of uracil used as a chemotherapy drug (Figure 1).

Uracil, the naturally occurring pyrimidine, consists of two nitrogen atoms and carbonyl groups, while in 5-FU, the fluorinated analog of uracil, the hydrogen atom in position 5 is replaced with a fluorine atom. The anticancer activity of this drug has been

Figure 1
Chemical structure of 5-fluorouracil (5-FU)



extensively studied in the recent past [25]. The responsiveness of cancer is measured using the sensitivity of the drug to cancer cells. There are multiple ways in which 5-FU can be consumed, such as applied directly to the skin or orally as tablets/capsules/completely soaked in water or through injections directly into a vein, following the prescription recommendations of the drug [26]. Molecular imprinting helps in the controlled release of this drug and can protect the drug from chemical breakdown. This process possesses many technical and practical challenges, like low specificity of binding sites, problems with residual templates, poor release kinetics, etc. The integration of the latest computational techniques in molecular imprinting not only makes the process cost-efficient but also lets people save more time. Recently, ML algorithms have been integrated more frequently with AI in the process of designing MIPs [27–29]. Complexities of the polymerization process due to many variables present, such as monomer–template interaction, polymerization conditions, or solvent properties, are addressed in many such studies using the latest ML techniques [30]. The current study also tries to predict the binding affinity of the drug based on existing data on MIP. K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Ridge Regression, Lasso Regression, and XGBoost are the selected six supervised ML methods [31–34]. Each model gives a different perspective on the problem. KNN focuses on local similarity patterns without assuming a predefined functional form; Decision Tree introduces interpretable rule-based partitioning; Random Forest extends the analysis by aggregating multiple trees to improve robustness and capture nonlinear interactions; Ridge and Lasso regression support regularized linear baselines; and XGBoost gives gradient boosting to model sequential residual structure and complex feature interactions [35, 36]. Table 1 shows the comparative analysis of recent MIP studies.

2. Methods

A detailed architectural framework for predicting the IF of MIPs is presented in Figure 2. The IF is quantified as:

$$IF = \frac{QMIP}{QNIP}$$

where $QMIP$ represents the binding capacity of the MIP and $QNIP$ represents the binding capacity of the corresponding NIP. The IF was predicted here using algorithmic techniques.

Using the molecular imprinting technique, the functional monomers and crosslinkers are involved in self-assembly and polymerization, which leads to the formation of NIP and MIP. During the template application, 5-FU is also used in this study. In the case of 5-FU being present in the polymerization of MIPs, the polymer matrix is realized with the specific cavities that are responsible for the selective bonding with the 5-FU molecule.

Table 1
Comparative analysis of MIP studies

No	Reference	Advantages	Limitations
1	Shevchenko et al. [37]	Exhibits high selectivity and excellent biocompatibility of nano MIPs, highlighting how surface engineering enhances drug loading capacity and enables controlled release	Limited quantitative kinetic data and in vivo validation pending
2	Yarahmadi et al. [38]	Introduces the first ML-based model for imprinting factor prediction, achieving high accuracy ($R^2 \approx 0.87$) with an interpretable link between synthesis conditions and polymer performance	Dataset diversity limited and no experimental validation
3	Rajpal et al. [39]	Employs Density Functional Theory (DFT) and molecular-dynamics simulations to optimize monomer–template binding and reduces experimental trial and error	High computational cost and lacks empirical benchmarking
4	Xu et al. [40]	Describes nanoMIPs for cancer targeting and recognition; balanced theoretical and experimental discussion	Limited large-scale synthesis reproducibility data
5	Abdel-Hamied et al. [41]	Highlights latest MIP-based biosensing technologies and diagnostic devices and discusses clinical potential	Focused on biosensing, not drug delivery, and lacks ML-related insights
6	Proposed work	The system enables the prediction of MIP imprinting performance, reducing experimental effort through a user-friendly machine learning framework	The system demonstrates predictive capability even with limited dataset. But there is significant potential to develop larger datasets and implement advanced structure-aware models such as GNNs to improve MIP design

The synergy of information from empirical results and data from the literature leads to a compilation that deals with the main aspects, like the respective templates, monomers, and crosslinkers. SMILES (Simplified Molecular Input Line Entry System) format was used to represent templates, monomers, and crosslinkers to calculate various molecular descriptors to represent the chemical characteristics of the molecules that are attested. The multisource dataset contains different kinds of chemical information. During the building of the training/testing dataset, we divided the data into two parts, where 80% of the data was used for training and the remaining 20% was used for testing.

Imprinting Factor: Affinity, the ability of the polymer to build target molecules, is the binding capacity of MIPs. MIP Binding Capacity (QMIP) is the level of the target molecule that the imprinted polymer can specifically bind in the presence of tailor-made cavities that are made according to the target's size, shape, and chemical properties. NIP Binding Capacity (QNIP) is the binding capability of an NIP, in which the aforementioned specific cavities are not present, and it serves as a control against nonspecific binding. The evaluation of QMIP and QNIP, in fact, helps to figure out the IF, which measures the specificity and selectivity of the MIP. A larger QMIP respects the QNIP to prove the uniqueness of the MIP and its ability to specifically bind the target molecule. This is a part of the dataset.

2.1. Data sources

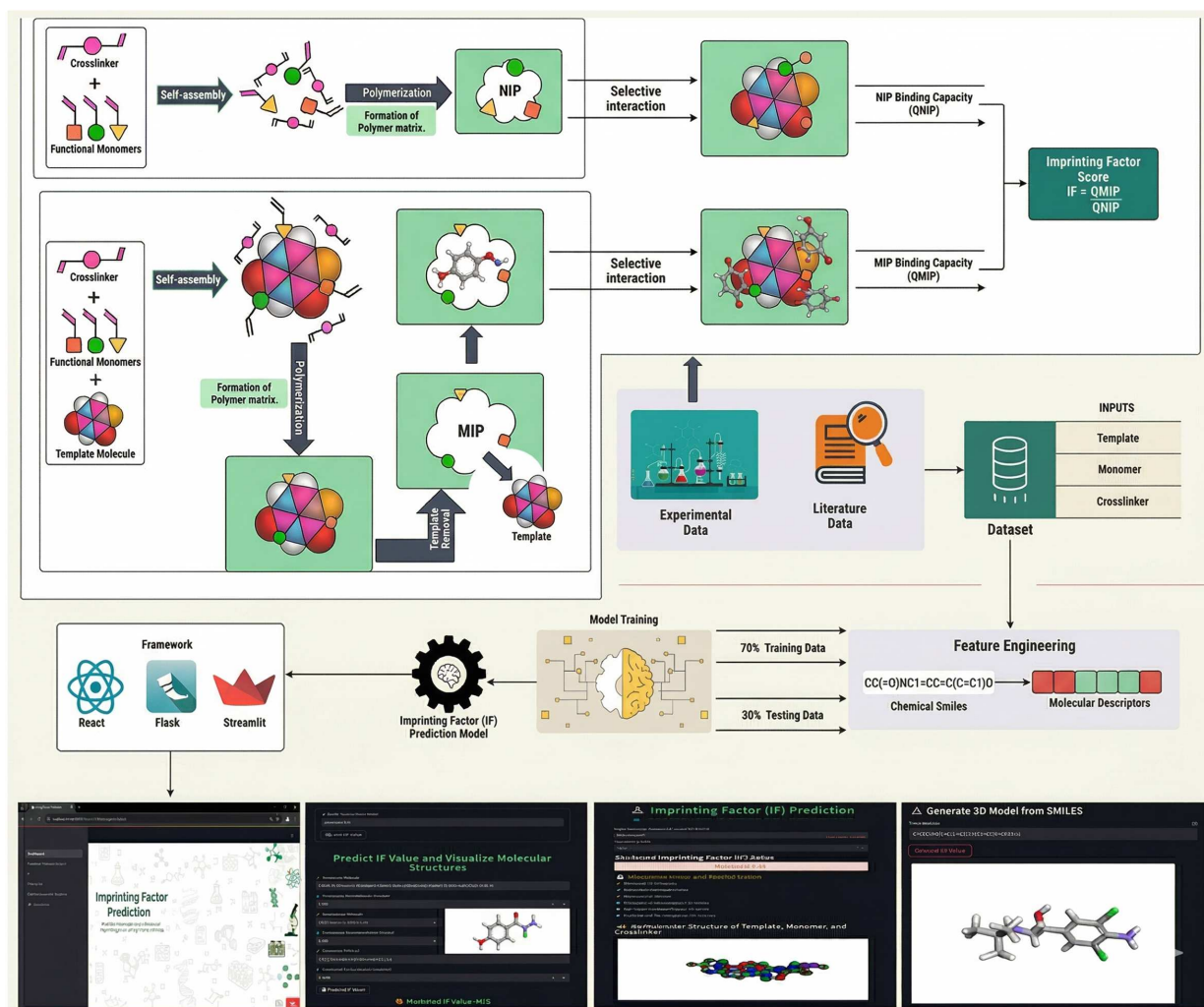
The major sources of data for this work are experimental measurements and literature-derived data. Experimental data were obtained through laboratory-based observations, providing direct evidence of molecular behavior under controlled conditions, while literature data were collected from peer-reviewed

publications to complement and extend the dataset. In this study, a dataset comprising 214 unique MIP formulations was assembled to predict the IF. Of these, 155 formulations were generated experimentally, while the remaining 59 were collected from literature sources [42–60]. Each entry represents a distinct polymer formulation characterized by its template, functional monomer, crosslinker, and associated physicochemical properties. The integration of empirical and literature data enabled comprehensive coverage of key design parameters, including templates, monomers, and crosslinkers, resulting in a diverse multisource dataset. The dataset was subsequently divided into 80% training and 20% testing sets for model development and evaluation.

2.2. Dataset composition and preprocessing

The dataset includes 214 MIP formulations, consisting of various n combinations of template molecules ($n = 134$), functional monomers ($n = 32$), and crosslinkers ($n = 12$), thereby addressing a broad chemical design space. The entries with missing values (IF) were excluded prior to modeling. All values for core structural identifiers like template, monomer, and crosslinker were complete. To reduce sensitivity to outliers, numerical features were further scaled using a *RobustScaler* method. The dataset is not uniformly balanced across polymer types because certain monomers and crosslinkers are more commonly represented due to their frequent usage in laboratory experimental studies. This is to avoid artificial sampling bias, and it reflects real-world research trends. The practical rarity of highly optimized, high-affinity MIP systems shows a right-skewed distribution for IF values. These properties used to maintain the realism and applicability of the predictive modeling framework. Each row in the dataset represents a unique MIP formulation, capturing key features such

Figure 2
Proposed workflow architecture



as the template molecule, functional monomer, crosslinker, and synthesis conditions. The features considered in this study are detailed below.

2.3. Molecular descriptors

Molecular descriptors are quantitative representations of molecular structure and properties used to model and predict chemical behavior in computational studies. The integration of electronic structure calculations with Python-based cheminformatics tools for efficient molecular descriptor generation and analysis has been explained in this study [61]. Byadi et al. [62] introduce DOPtools, a Python platform for automated molecular descriptor calculation and optimization of predictive models in cheminformatics workflows. Gray and Wu [63] provide a comparative evaluation of descriptor sets and modern AI representations for improving ML-based molecular property prediction. Based on these, the proposed work focuses on the following descriptors.

2.3.1. Molecular weight of crosslinker

The molecular weight of the crosslinker influences the polymer's pore structure, affecting template accessibility, binding stability, and rigidity. Low molecular weight leads to dense, stable networks, while high molecular weight improves

flexibility and mass transfer. So maintaining a balance between these two is important in achieving selective molecular recognition in imprinted polymers.

2.3.2. Molecular weight of template

The molecular weight of the template affects the size and precision of the imprinted cavities. Smaller templates enable uniform binding sites and efficient diffusion, while larger ones may cause irregular imprinting. Choosing an optimal template size is crucial for high imprinting efficiency.

2.3.3. Molecular weight of monomer

The molecular weight of the functional monomer influences its reactivity, mobility, and interaction with the template during polymerization. Low molecular weight monomers offer better uniform interaction with the template. High molecular weight monomers may reduce the number of effective binding interactions due to limited mobility. Additionally, they can affect the porosity and flexibility of the final polymer structure. Therefore, choosing the appropriate monomer molecular weight is important to ensure strong template-monomer interactions and efficient imprinting.

2.3.4. Concentration of the monomer

Monomer concentration affects the number and quality of binding sites in imprinted polymers. Low concentrations weaken binding, while high concentrations can cause nonspecific interactions. So maintaining a balance between these two is important to achieving strong, selective binding and maintaining polymer structure.

2.3.5. Concentration of the template

The template concentration in a polymerization mixture is one of the major factors that affect the formation and quality of imprinted sites in the polymer matrix. It is an essential factor that also affects the efficiency and selectivity of the growing MIPs. The optimal concentration of the functional monomer stands out as the most important one as it plays a crucial role in stable and selective binding sites in MIPs creation. Whereas higher template concentrations can improve capacity to bind, excessive amounts can ruin selectivity. Achieving an optimum balance in the template concentration is the key to enabling the exactness of the binding sites, and it guarantees the polymer can survive and the crosslinking can be done efficiently.

2.3.6. Concentration of the crosslinker

The concentration of the crosslinker is very crucial to stabilize the polymer network of MIPs besides its effect on their flexibility, structural stability, and porosity. The right concentration for the crosslinker is an important factor in sustained imprinted sites that are of suitable size, thus enhancing the binding and selectivity properties and the fabric of the material. It is necessary to find the right balance in the concentration of the crosslinker and functional monomers to manufacture MIPs that are highly efficient and have a remarkable variety of applications.

2.3.7. XlogP

The XlogP value, which indicates the partition coefficient between octanol and water, is the key factor in the assessment of the hydrophobicity of a molecule and its influence on MIPs. A higher XlogP value increases hydrophobic interactions, resulting in higher imprint stability and improved binding affinity.

2.3.8. Hydrogen bond acceptor count

Hydrogen bond acceptors (HBAs) are instrumental in providing a template molecule with the required selectivity for a specific target in MIPs formation. Thus, it not only plays the role of selective binding but also binds more tightly. The higher the HBA count is, the better it is for the hydrogen bonding to the template, which is responsible for the stability of the template-polymer complex and the maintenance of the imprint shape. By considering the HBA count in predictive models, we can accurately predict MIP performance and design the correct recognition of the target molecule.

2.3.9. Hydrogen bond donor count

A hydrogen bond donor (HBD) is an atom or group within a molecule that donates a hydrogen atom for hydrogen bonding. To enhance the binding affinity of the target molecule during polymerization, MIPs utilize HBD.

2.3.10. Solubility

Solubility influences the uniform interaction of the template, monomer, and crosslinker during polymerization. Poor solubility can cause weak binding and irregular structures, while good solubility ensures consistent imprinting. Selecting the right solvent is important for optimal polymer porosity and molecular recognition. For this study, the data were organized to train the ML models. Every possible combination of the parameters was explored using the GridSearchCV method. Such an approach was most beneficial since we aimed at the complete usage of the provided information. To preprocess the data, the features were standardized using RobustScaler to make the data less sensitive to outliers. The parameter grids for each model are shown in Table 2.

For the convenience of users, a web application was also developed using Streamlit. Access to the tool and its user-friendly features for submitting data and viewing predictions was possible through this web interface. It will enable researchers to input the required chemical data and quickly receive their respective IF predictions. In addition to making IF predictions, the tool can also

Table 2
Model hyperparameter grid

Model	Hyperparameter	Values
Random Forest Regressor	n_estimators	[10, 50, 100]
	max_depth	[5, 10, 15, None]
	min_samples_split	[2, 5, 10]
	min_samples_leaf	[1, 2, 4]
Decision Tree Regressor	max_depth	[None, 10, 20, 30]
	min_samples_split	[2, 5, 10]
	min_samples_leaf	[1, 2, 4]
XGBoost Regressor	n_estimators	[50, 100, 200]
	max_depth	[2, 3, 4]
	learning_rate	[0.01, 0.05, 0.1]
K-Nearest Neighbors	n_neighbors	[3, 5, 7]
Ridge Regression	Alpha	[0.01, 0.1, 1.0, 10.0]
Lasso Regression	Alpha	[0.01, 0.1, 1.0, 10.0]

be used to visualize the structure of the component and extract the molecular descriptors.

represented as boxplots and heatmaps in Figures 3 and 4, respectively.

3. Results

Ten different features of the data were considered, such as molecular weight and concentration of template, monomer, and crosslinker, solubility, XlogP, HBA count, and HBD count. IF values were predicted as the output for our analysis. The distribution of the features and exploratory data analysis are

3.1. Exploratory data analysis

Let the dataset be represented by a feature matrix

$$X \in \mathbb{R}^{n \times 10}$$

where:

$n = 214$ is the number of MIP formulations,

Figure 3
Feature correlation heatmap

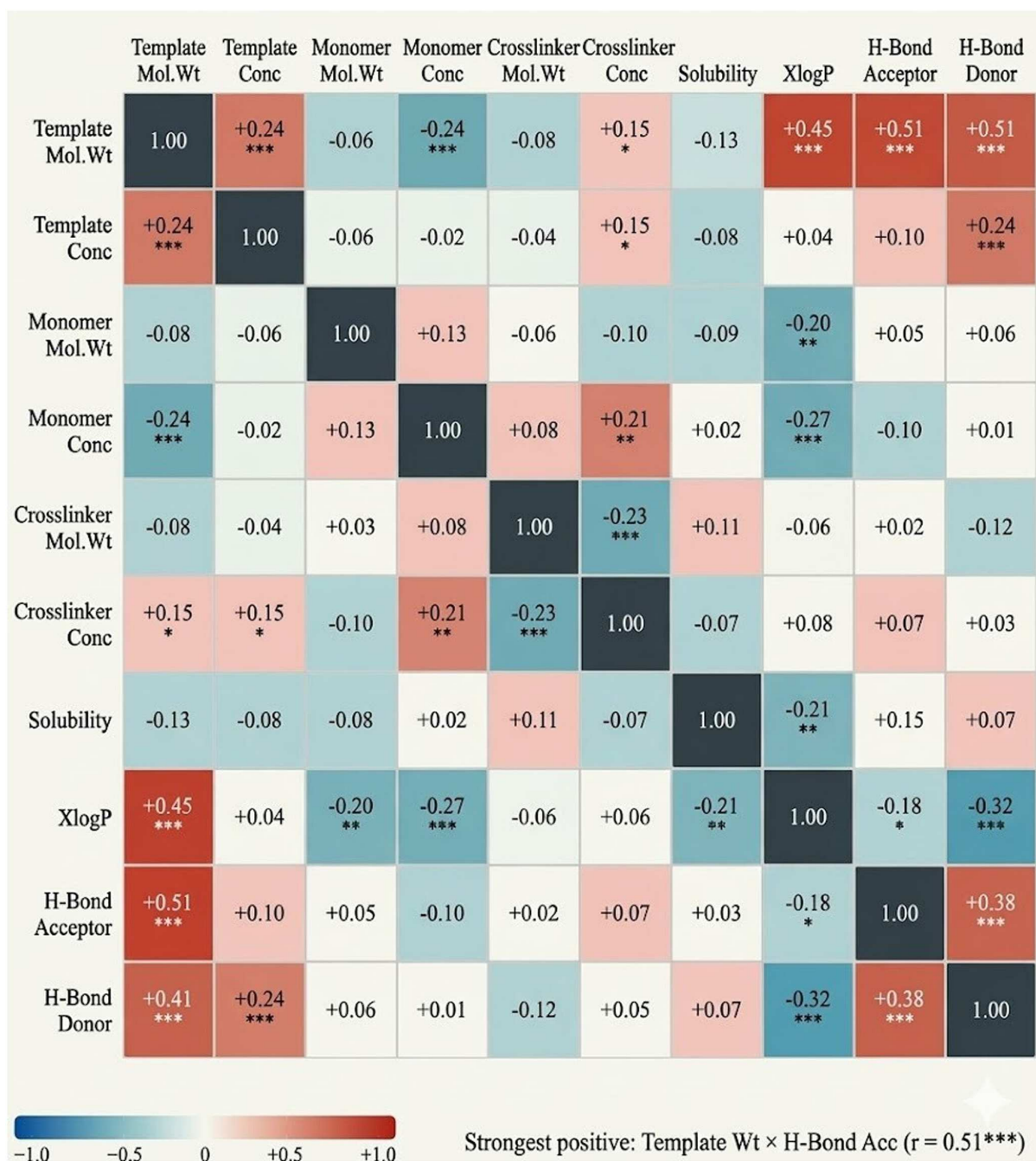
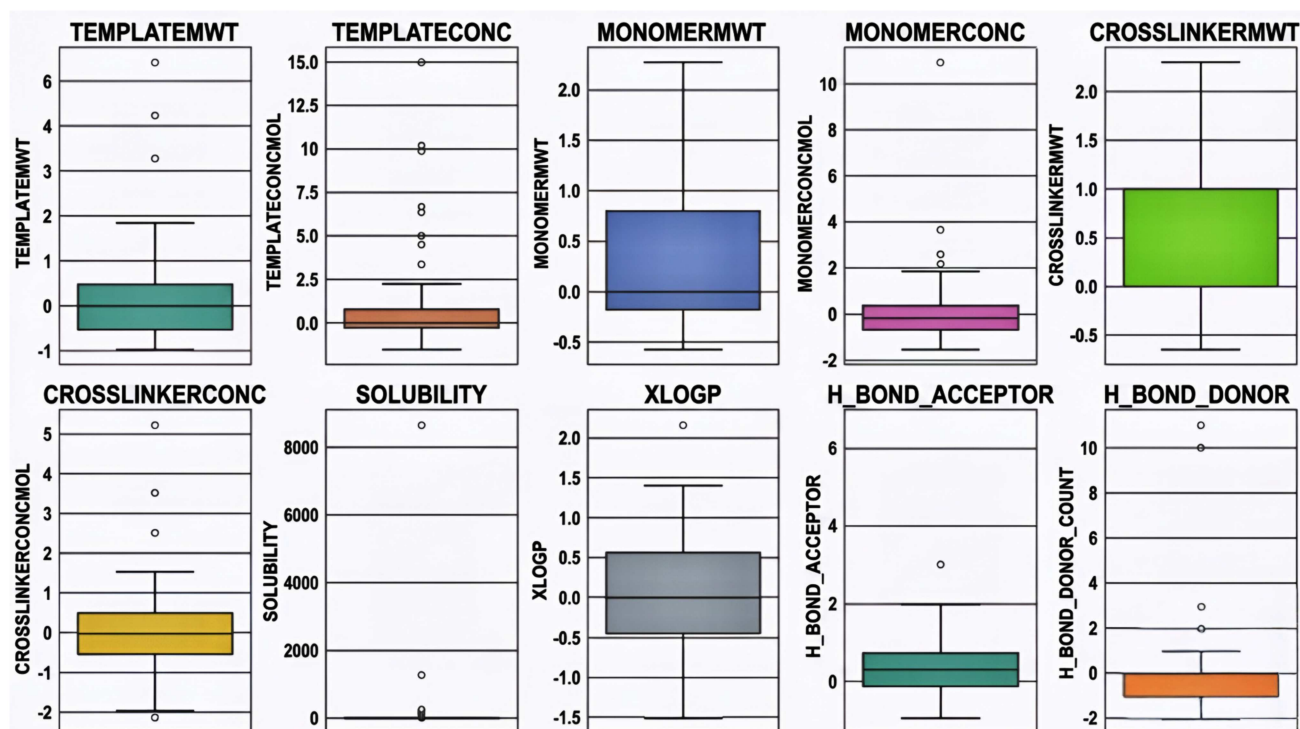


Figure 4
Distribution of molecular descriptors



the 10 columns correspond to the molecular descriptors: $X_1, X_2, \dots, X_{10}$

For any two descriptors X_i and X_j , the Pearson correlation coefficient is defined as

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

where:

x_{ki} is the value of descriptor i for sample k ,

$\bar{x}_i$ is the mean of descriptor i .

Figure 3 illustrates the Pearson correlation matrix for 10 molecular descriptors, with asterisks denoting statistical significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Template molecular weight emerged as the most interconnected descriptor across the matrix. It showed its strongest associations with HBA count ($r = +0.51$, $p < 0.001$) and HBD count ($r = +0.51$, $p < 0.001$), showing that templates with higher molecular weight carry greater hydrogen bonding capacity. A positive correlation also exists between template molecular weight and XlogP ($r = +0.45$, $p < 0.001$), indicating that larger templates in this dataset tend to be more lipophilic. Template molecular weight further correlated positively with template concentration ($r = +0.24$, $p < 0.001$) and crosslinker concentration ($r = +0.15$, $p < 0.05$), while exhibiting a significant negative association with monomer concentration ($r = -0.24$, $p < 0.001$).

Among hydrogen bonding descriptors, acceptor and donor counts were moderately correlated with each other ($r = +0.38$, $p < 0.001$), which is chemically expected since both reflect the polarity of the template. XlogP showed negative associations with monomer concentration ($r = -0.27$, $p < 0.001$), monomer molecular weight ($r = -0.20$, $p < 0.01$), solubility ($r = -0.21$, $p < 0.01$),

and HBD count ($r = -0.32$, $p < 0.001$), collectively reflecting the inverse relationship between hydrophobicity and both polarity and aqueous solubility.

Crosslinker molecular weight and crosslinker concentration were negatively correlated ($r = -0.23$, $p < 0.001$), consistent with typical formulation practice where denser crosslinkers are applied at lower concentrations to maintain network flexibility. Template concentration and crosslinker concentration show a weak relation ($r = +0.15$, $p < 0.05$), while monomer concentration and crosslinker concentration show a slightly positive relationship ($r = +0.21$, $p < 0.01$).

The remaining feature pairs produced correlation coefficients below ± 0.15 with no statistical significance. The nonoverlapping chemical information present in those descriptors leads to unbiased IF prediction. The low multicollinearity supports the use of all 10 features simultaneously in the ML models.

RobustScaler has been applied to perform post-scaling distributions of 10 molecular descriptors (Figure 4). The changes in the descriptor values based on median centering and interquartile range scaling normalize the feature space by minimizing the influence of outliers. This is important in the current dataset, where descriptors such as solubility, template concentration, and HBD count exhibit right skewness and multiple upper outliers. The model stability and interpretability are achieved by applying RobustScaler, which in turn removes the impact of these outliers. RobustScaler enables all features to contribute proportionally to the ML model. This preprocessing step ensures that descriptors maintain their predictive contribution without permitting their outlier-driven variance to distort the learned feature space during model training. The most affected by distributional extremes are solubility, monomer concentration, and template concentration. So it is essential to adopt outlier handling strategies during the application of ML

models. The performance of each model was assessed by root mean squared error (RMSE), mean absolute error (MAE), and R^2 .

Table 3 displays the hyperparameter configuration and its corresponding RMSE, MAE, and R^2 values for each model. The Random Forest model shows the best performance among all evaluated models, with the lowest RMSE (3.54), MAE (2.19), and highest R^2 (0.55), indicating strong predictive capability when compared with other models.

The Random Forest outperforms all other models, as it has the lowest RMSE (3.54), suggesting better predictive accuracy (Figure 5). In contrast, the Lasso Regressor delivered an RMSE of 5.19. Decision Tree with an RMSE of 4.85 performs slightly better than Ridge (5.18) and KNN (4.93). XGBoost gives an RMSE value of 5.16.

Further data analysis was performed using the Random Forest model. Figure 6 demonstrates the training and validation loss of the model. Its learning curve demonstrates low training error, which is typical for ensemble models like Random Forest that can easily fit training data. The initially high and variable validation error decreases significantly as training size increases, suggesting that the model benefits greatly from more data. The flattening of the validation curve at larger training sizes indicates stable generalization and minimal overfitting. The scatter plot (Figure 7) visualizes the performance of a Random Forest Regression model in predicting the IF values. The X-axis represents the true, observed values of the target variable (IF) from the test dataset, while the Y-axis represents the values of (IF) predicted by the trained Random Forest model for the same test data points. Each point represents a single data point from the test set. The horizontal position indicates the actual IF value, and the vertical position shows the model's prediction. The diagonal line represents a perfect prediction. Points falling exactly on this line mean the model predicted the IF value correctly. Deviations from this line indicate prediction errors. The RMSE value (Table 3) shows the model's performance, and the scatter plot provides visual information about Random Forest predictions. The red dashed line represents the ideal prediction line ($y = x$), indicating that most predictions are close to the actual values, with a few deviations at higher IF ranges.

3.2. Random Forest analysis

The best prediction accuracy is obtained in the Random Forest model, outperforming all other models, achieving the lowest RMSE (3.54). By combining multiple decision trees, Random Forest can capture complex, nonlinear relationships in the data. Its ensemble approach also reduces overfitting. In comparison,

Figure 5
Model RMSE comparison

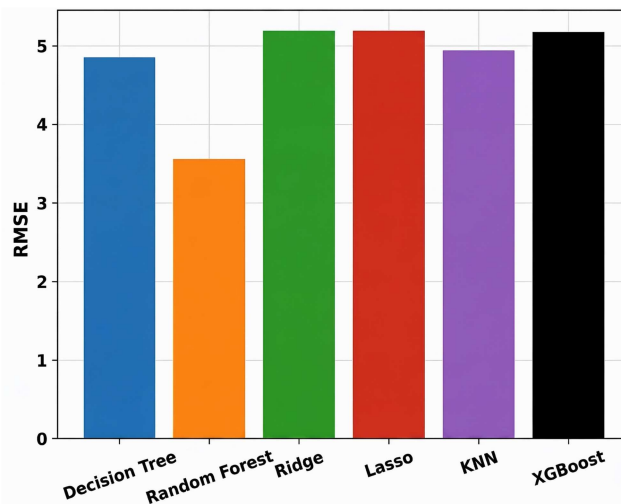
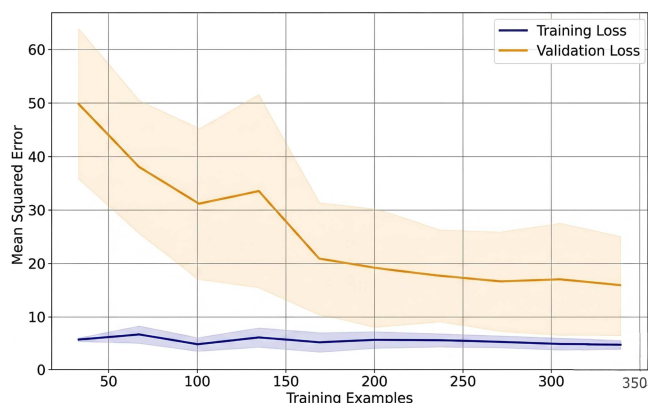


Figure 6
Random Forest learning curve



other models like Lasso, Decision Tree, and KNN showed higher errors, denoting weaker predictive performance.

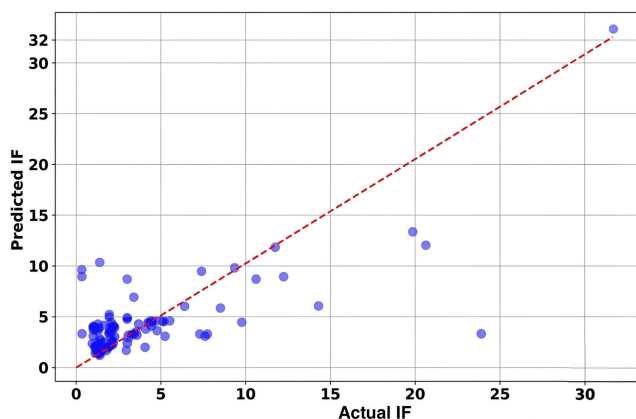
3.3. Effect of crosslinker concentration

A web-based tool was also developed as part of the study; a snapshot of the same is shown in Figure 8. The interface enables input of molecular data, visualization of 3D structures, and retrieval of predicted IF values with corresponding compound details. The user can input the SMILES/compound name with concentration details of the template, crosslinker, and monomer

Table 3
Model performance comparison

Model	Optimal hyperparameters	RMSE	MAE	R^2
Decision Tree Regressor	max_depth:10, min_samples_leaf:4, min_samples_split:10	4.85	2.66	0.16
Random Forest Regressor	n_estimators:100, max_depth:10	3.54	2.19	0.55
XGBoost Regressor	n_estimators:50, max_depth:2, learning_rate:0.01	5.16	3.2	0.094
Ridge Regression	alpha:1.0	5.18	3.52	0.04
Lasso Regression	alpha:0.1	5.19	3.51	0.04
K-Nearest Neighbors	n_neighbors:7	4.93	2.85	0.13

Figure 7
Actual vs predicted IF values



into the system. The system will predict the IF values based on the selected features. It also provides a 3D view of the compounds. Figure 9 shows a structural comparison between two MIP networks synthesized with identical numbers of functional monomers ($n = 20$), but differing in crosslinker count (10) (Figure 9(a)) vs 3 (Figure 9(b)). The mechanistic interpretation is grounded in polymer physics: higher crosslinker density increases network junction density, reducing mesh size and pore diameter, thereby restricting template diffusion during the imprinting step and leading to shallower, less-accessible binding cavities consistent with findings reported by Akgönüllü et al. [64]. Excessive crosslinking can cause template entrapment rather than template-directed cavity formation, resulting in reduced QMIP.

Lower crosslinker concentration results in a more flexible and open polymer network, promoting better monomer–template alignment, reduced steric hindrance, and the formation of

well-defined, accessible binding sites, thereby enhancing template recognition and increasing the IF. This behavior is further supported by mesh-size estimates from rubber elasticity theory [65]. Both binding capacity and selectivity are consistent with these observations. Comparative analysis in Table 4 shows that reduced crosslinker ratios improve both binding capacity and selectivity, indicating the important role of crosslinker concentration in the field of polymer morphology and functional performance.

Figure 9 shows a structural comparison between two MIP networks synthesized with identical numbers of functional monomers ($n = 20$), but differing in the number of crosslinkers, 10 in Figure 9(a) and 3 in Figure 9(b).

In Figure 9(a), a higher crosslinker concentration exhibits a denser and more rigid network, where crosslinkers dominate the spatial architecture. This high degree of crosslinking likely leads to restricted mobility of monomer–template complexes during polymerization, potentially resulting in deformed or buried binding cavities. Even though the mechanical stability may be higher, the reduced accessibility of these cavities is expected to yield lower binding affinity and reduced imprinting efficiency. In contrast, the structure prepared with only three crosslinkers demonstrates a more open and flexible polymer matrix. The improved monomer–template distribution results from slowing or preventing chemical reactions from crosslinkers, enabling the formation of well-defined and accessible binding cavities. This type of structure is anticipated to exhibit higher binding site fidelity, stronger template recognition, and, as a result, a higher IF. These morphological differences strongly support the binding results as shown in Table 4. The MIP with a lower crosslinker ratio demonstrated enhanced performance in both binding capacity and selectivity. The comparative analysis between high and low crosslinker MIPs demonstrates that crosslinker concentration has a substantial impact on polymer morphology and functionality. The variant of the crosslinker (10 crosslinkers) formed a dense and rigid structure, in contrast to the more open and flexible network

Figure 8
IF prediction web interface

Figure 9
Effect of crosslinkers on MIP structure: (a) comparison with 10 crosslinkers and (b) comparison with three crosslinkers

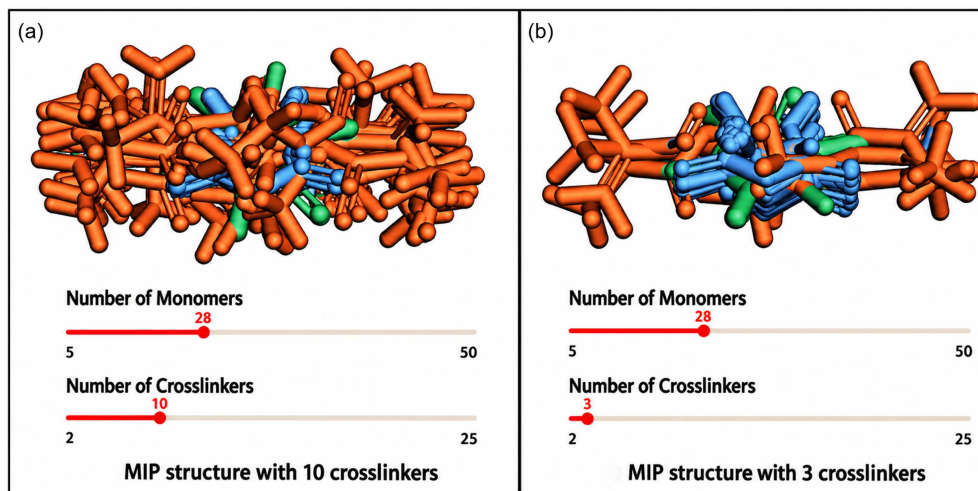
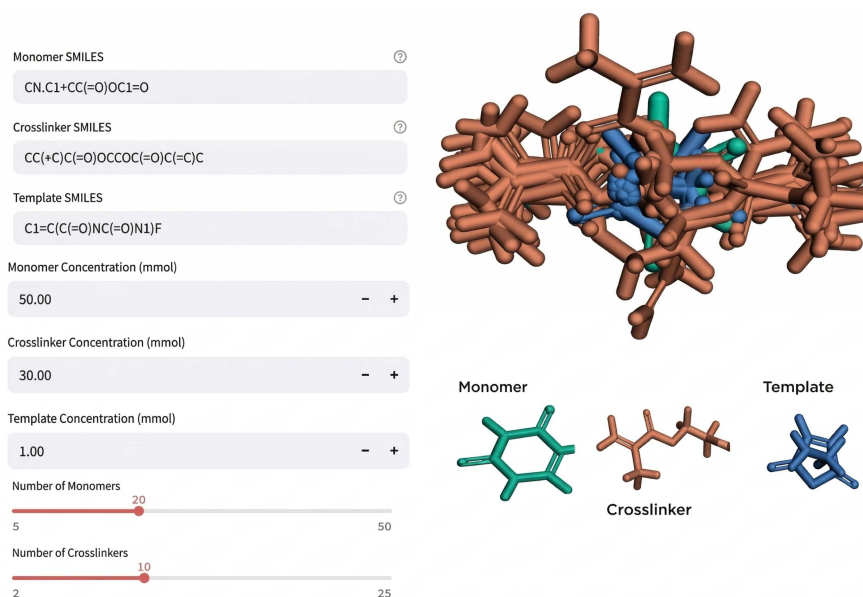


Table 4
Effect of crosslinker concentration

Feature evaluated	High crosslinker	Low crosslinker	Scientific interpretation/effect
Number of monomers	20	20	Consistent polymer backbone
Number of crosslinkers	10	3	Top has higher crosslinking
Polymer density	Dense, entangled	Looser, linear	Crosslinking tightens structure
Structure appearance	Bulky	Open, elongated	Related to steric hindrance
Potential porosity	Low	High	Looser structure allows voids
Flexibility	Lower	Higher	Fewer crosslinkers = more mobility
Crosslinker visibility	Evenly distributed	Sparse	Matches slider inputs
Binding capacity	Moderate to Low	High	More accessible binding sites
Cavity information	Shallow, distorted	Well-formed	Better cavity formation

Figure 10
5-FU-based MIP structure



observed in the low-crosslinker MIP (three crosslinkers). Increased crosslinking reduced porosity and flexibility, limiting molecular accessibility and distorting cavity shapes, while reduced crosslinking enhanced analyte diffusion and cavity integrity. As a result, the low crosslinker MIP demonstrated higher binding capacity and better selectivity due to preserved cavity structure and reduced steric hindrance.

Overall, reducing the crosslinker concentration improves the functional performance of MIPs, making them more effective for molecular recognition applications. The figure (Figure 10) represents an MIP structure based on 5-FU with specified concentrations of monomer (50 mmol), crosslinker (30 mmol), and template (1 mmol). A high monomer concentration aids the formation of abundant functional sites around the template, improving interaction specificity. The crosslinker concentration gives a rigid structural network that maintains the integrity of the binding cavities without restricting template access. The ratio of 20 monomers to 16 crosslinkers shows a balance allowing sufficient flexibility while preserving cavity shape fidelity. Overall, this configuration is likely to yield an MIP with high imprinting efficiency and a strong binding affinity, leading to a high IF.

4. Conclusion

This study demonstrates that ML can reliably predict the IF of MIPs synthesized with the specific template molecule. Among six models evaluated, Random Forest Regressor achieved the strongest performance (RMSE = 3.54, MAE = 2.19, $R^2 = 0.55$), outperforming Decision Tree (RMSE = 4.85, MAE = 2.55, $R^2 = 0.6$), XGBoost (RMSE = 5.16, MAE = 3.2, $R^2 = 0.094$), Ridge (RMSE = 5.18, MAE = 3.52, $R^2 = 0.04$), Lasso regression (RMSE $\approx$ 5.19, MAE = 3.51, $R^2 = 0.04$), and KNN (RMSE = 4.93, MAE = 2.85, $R^2 = 0.13$). A user-friendly web interface was also deployed for IF estimation, making the process of evaluating polymer performance more accessible and practical.

The identified limitations provide a basis for future work including the dataset of 214 usable entries represents a narrow slice of the full MIP formulation space, increasing variance in test-set estimates and limiting generalizability to underrepresented chemistries, the right-skewed IF distribution means the model is less calibrated for high-affinity formulations, which are precisely the ones of greatest practical interest, and 3D conformational and electronic interaction features that govern actual template-monomer binding are missing. The dataset is not uniformly balanced across polymer types; certain monomers and crosslinkers (predominantly acrylic-acid-based monomers and Ethylene Glycol Dimethacrylate (EGDMA) crosslinkers) are over-represented due to the experimental work. This reflects real-world research trends rather than artificial sampling bias and is identified as a generalizability limitation. So the future work should focus on the following directions: dataset expansion via Natural Language Processing (NLP)-assisted systematic mining of the MIP literature, targeting a broader chemical diversity of templates, monomers, and crosslinkers, and graph neural network (GNN) models trained directly on molecular graphs of the template, monomer, and crosslinker, enabling end-to-end learning of structure-property relationships [66]. Transfer learning from related polymer property databases (PolyInfo, MatWeb) to improve performance under data scarcity and accelerate model convergence on new MIP chemistries [67–69], and prospective experimental validation, in which the top-10 Random Forest-predicted formulations are synthesized and characterized via Quartz Crystal Microbalance with Dissipation

rebinding assays, providing the ground truth needed to close the prediction validation loop and validate the computational predictions in vitro.

Acknowledgment

This work derives direction and ideas from the chancellor of Amrita University, Sri Mata Amritanandamayi Devi.

Funding Support

This work was supported by funding from DBT BIRAC IKP Knowledge Park.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Author Contribution Statement

Jisha R. C.: Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Niranjan Prasanth:** Software, Formal analysis, Writing – original draft. **Aswathi Perumkunnappala:** Investigation, Data curation. **Meenakshi Krishnan Praseetha Radhakrishnan:** Investigation, Data curation. **Manjusha Nair:** Conceptualization, Validation, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration. **Renuka Suravajhala:** Conceptualization, Validation, Investigation, Resources, Writing – review & editing, Supervision, Funding acquisition.

References

- [1] Pauling, L. (1940). A theory of the structure and process of formation of antibodies. *Journal of the American Chemical Society*, 62(10), 2643–2657. <https://doi.org/10.1021/ja01867a018>
- [2] Garg, S., Singla, P., Kaur, S., Canfarotta, F., Velliou, E., Dawson, J. A., ..., & Peeters, M. (2025). Future perspectives on the automation and biocompatibility of molecularly imprinted polymers for healthcare applications. *Macromolecules*, 58(3), 1157–1168. <https://doi.org/10.1021/acs.macromol.4c01621>
- [3] Arshady, R., & Mosbach, K. (1981). Synthesis of substrate-selective polymers by host-guest polymerization. *Die Makromolekulare Chemie*, 182(2), 687–692. <https://doi.org/10.1002/macp.1981.021820240>
- [4] Alexander, C., Andersson, H. S., Andersson, L. I., Ansell, R. J., Kirsch, N., Nicholls, I. A., ..., & Whitcombe, M. J. (2006). Molecular imprinting science and technology: A survey of the literature for the years up to and including 2003.

- Journal of Molecular Recognition*, 19(2), 106–180. <https://doi.org/10.1002/jmr.760>
- [5] Li, Z., Deng, J., Ma, P., Bai, H., Jin, Y., Zhang, Y., . . . , & Dong, A. (2024). Stimuli-responsive molecularly imprinted polymers: Mechanism and applications. *Journal of Separation Science*, 47(19), e202400441. <https://doi.org/10.1002/jssc.202400441>
 - [6] Sullivan, M. V., Lasserre, P., Blackburn, C., Turner, N. W., & Selligren, B. (2025). Stimuli-responsive molecularly imprinted materials: Fundamentals and applications. *Responsive Materials*, 3(1), e20240032. <https://doi.org/10.1002/rpm.20240032>
 - [7] Mohsenzadeh, E., Ratautaite, V., Brazys, E., Ramanavicius, S., Zukauskas, S., Plausinaitis, D., & Ramanavicius, A. (2024). Application of computational methods in the design of molecularly imprinted polymers (review). *TrAC Trends in Analytical Chemistry*, 171, 117480. <https://doi.org/10.1016/j.trac.2023.117480>
 - [8] Kapase, S. A., Jadhav, S. A., Peralta, M. E., & Carlos, L. (2025). Molecularly imprinted polymers (MIPs): Synthesis, applications and recent advances in water remediation. *ACS Applied Polymer Materials*, 7(16), 10358–10394. <https://doi.org/10.1021/acsapm.5c02039>
 - [9] Pasquardini, L., & Bossi, A. M. (2021). Molecularly imprinted polymers by epitope imprinting: A journey from molecular interactions to the available bioinformatics resources to scout for epitope templates. *Analytical and Bioanalytical Chemistry*, 413(24), 6101–6115. <https://doi.org/10.1007/s00216-021-03409-1>
 - [10] Tse Sum Bui, B., Mier, A., & Haupt, K. (2023). Molecularly imprinted polymers as synthetic antibodies for protein recognition: The next generation. *Small*, 19(13), 2206453. <https://doi.org/10.1002/smll.202206453>
 - [11] Chen, S., Sun, J., Li, Y., Qiao, A., Hou, Y., Jin, L., & Wang, H. (2025). Advances in molecularly imprinted polymers for clinical biomarker detection (2021–2025). *Microchimica Acta*, 192(9), 572. <https://doi.org/10.1007/s00604-025-07443-z>
 - [12] Urriza-Arsuaga, I., Guadaño-Sánchez, M., & Urraca, J. L. (2023). Current trends in molecular imprinting: Strategies, applications and determination of target molecules in Spain. *International Journal of Molecular Sciences*, 24(3), 1915. <https://doi.org/10.3390/ijms24031915>
 - [13] Leibl, N., Haupt, K., Gonzato, C., & Duma, L. (2021). Molecularly imprinted polymers for chemical sensing: A tutorial review. *Chemosensors*, 9(6), 123. <https://doi.org/10.3390/chemosensors9060123>
 - [14] Gavrilă, A.-M., Diacon, A., Iordache, T.-V., Rotariu, T., Ionita, M., & Toader, G. (2024). Hazardous materials from threats to safety: Molecularly imprinted polymers as versatile safeguarding platforms. *Polymers*, 16(19), 2699. <https://doi.org/10.3390/polym16192699>
 - [15] Yu, M., Li, H., Xie, J., Xu, Y., & Lu, X. (2022). A descriptive and comparative analysis on the adsorption of PPCPs by molecularly imprinted polymers. *Talanta*, 236, 122875. <https://doi.org/10.1016/j.talanta.2021.122875>
 - [16] Büyüktiryaki, S. (2025). Molecularly imprinted polymer nanoparticles for pharmaceutical applications: Sample preparation, sensor-based detection, and controlled drug release. *Polymers*, 17(17), 2283. <https://doi.org/10.3390/polym17172283>
 - [17] Morsi, S. M. M., Morsi, M. M., & Abd El-Aziz, M. E. (2026). Advances in molecularly imprinted synthetic polymers: From structural principles to functional technologies. *European Polymer Journal*, 245, 114542. <https://doi.org/10.1016/j.eurpolymj.2026.114542>
 - [18] Fu, L., Zhang, X., Zhao, X., Wang, L., & Fu, Y. (2025). Advances in biomass-based molecularly imprinted polymers: From sustainable design to diverse applications. *Green Chemistry*, 27(30), 9055–9093. <https://doi.org/10.1039/D5GC01018G>
 - [19] Liu, R., Zhao, M., Zhang, X., Zhang, C., Ren, B., & Ma, J. (2026). Advances and challenges in molecularly imprinted electrochemical sensors for application in environmental, biomedicine, and food safety. *Critical Reviews in Analytical Chemistry*, 56(4), 962–980. <https://doi.org/10.1080/10408347.2025.2460751>
 - [20] Seo, Y., Kim, Y., Kim, J., Jung, H. G., Jang, J. W., Kim, H., . . . , & Yoon, D. S. (2025). Recent advances in molecularly imprinted polymers toward biomedical healthcare devices. *Biosensors and Bioelectronics*, 287, 117637. <https://doi.org/10.1016/j.bios.2025.117637>
 - [21] Haupt, K., Medina Rangel, P. X., & Tse Sum Bui, B. (2020). Molecularly imprinted polymers: Antibody mimics for bioimaging and therapy. *Chemical Reviews*, 120(17), 9554–9582. <https://doi.org/10.1021/acs.chemrev.0c00428>
 - [22] Testa, V., Anfossi, L., Cavallera, S., di Nardo, F., Serra, T., & Baggiani, C. (2025). Evidence of positive cooperativity in the binding behavior of molecularly imprinted polymers prepared by solid phase polymerization synthesis. *ACS Applied Polymer Materials*, 7(5), 3286–3295. <https://doi.org/10.1021/acsapm.4c03726>
 - [23] Dashtaki, R. M., Dashtaki, S. M., Heydari-Bafrooei, E., & Piran, M. J. (2025). Enhancing the predictive performance of molecularly imprinted polymer-based electrochemical sensors using a stacking regressor ensemble of machine learning models. *ACS Sensors*, 10(4), 3123–3133. <https://doi.org/10.1021/acssensors.5c00364>
 - [24] Suravajhala, R., Burri, H. R., & Malik, B. (2022). Development of highly specific and selectively recognizing caffeine imprinted polymer nanomaterials with EGDMA crosslinker. *Current Nanomaterials*, 7(1), 65–72. <https://doi.org/10.2174/2405461506666210603121936>
 - [25] Zhang, T., Berghaus, M., Li, Y., Song, Q., Stollenwerk, M. M., Persson, J., . . . , & Lv, Y. (2025). PSMA-targeting imprinted nanogels for prostate tumor localization and imaging. *Advanced Healthcare Materials*, 14(3), 2401929. <https://doi.org/10.1002/adhm.202401929>
 - [26] Sethy, C., & Kundu, C. N. (2021). 5-fluorouracil (5-FU) resistance and the new strategy to enhance the sensitivity against cancer: Implication of DNA repair inhibition. *Biomedicine & Pharmacotherapy*, 137, 111285. <https://doi.org/10.1016/j.biopha.2021.111285>
 - [27] Sarkar, S., Kiren, S., & Gmeiner, W. H. (2024). Review of pro-drug and nanodelivery strategies to improve the treatment of colorectal cancer with fluoropyrimidine drugs. *Pharmaceutics*, 16(6), 734. <https://doi.org/10.3390/pharmaceutics16060734>
 - [28] Malashin, I. P., Tynchenko, V. S., Nelyub, V. A., Borodulin, A. S., & Gantimurov, A. P. (2024). Estimation and prediction of the polymers' physical characteristics using the machine learning models. *Polymers*, 16(1), 115. <https://doi.org/10.3390/polym16010115>
 - [29] Melge, A. R., Parate, S., Pavithran, K., Koyakutty, M., & Mohan, C. G. (2022). Discovery of anticancer hybrid molecules by supervised machine learning models and in

- vitro validation in drug resistant chronic myeloid leukemia cells. *Journal of Chemical Information and Modeling*, 62(4), 1126–1146. <https://doi.org/10.1021/acs.jcim.1c01554>
- [30] Yarahmadi, B., Hashemianzadeh, S. M., & Milani Hosseini, S. M.-R. (2023). A new approach to prediction riboflavin absorbance using imprinted polymer and ensemble machine learning algorithms. *Heliyon*, 9(7), e17953. <https://doi.org/10.1016/j.heliyon.2023.e17953>
- [31] Mofazali, P., Lalinia, M., Gross, J. D., & Samadi, A. (2026). Molecularly imprinted polymers: Applications, computational approaches, and the transformative role of artificial intelligence. *Coordination Chemistry Reviews*, 559, 217808. <https://doi.org/10.1016/j.ccr.2026.217808>
- [32] Jayaram, R. S., Prasanth, P. V., Saravanamuthukumar, P., Abdullah, A. B., & Ramalingam, K. (2026). Impact performance enhancement of nano-clay-reinforced sandwich panels: A machine learning approach. *Polymer Composites*, 47(2), 1218–1234. <https://doi.org/10.1002/pc.70212>
- [33] Speiser, J. L. (2021). A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. *Journal of Biomedical Informatics*, 117, 103763. <https://doi.org/10.1016/j.jbi.2021.103763>
- [34] Anbarasan, P., Rohith, P. S., Pranav, R., & Surya, A. V. K. S. (2025). Mental health state identification using machine learning algorithms. *SSRN*. <https://doi.org/10.2139/ssrn.5195594>
- [35] di Masi, S., de Benedetto, G. E., & Malitesta, C. (2024). Optimisation of electrochemical sensors based on molecularly imprinted polymers: From OFAT to machine learning. *Analytical and Bioanalytical Chemistry*, 416(9), 2261–2275. <https://doi.org/10.1007/s00216-023-05085-9>
- [36] Arjun, A. M., Deshpande, S., Liu, G., Miura, D., Pawlak, K., Takahiko, T., . . . , & Sharma, S. (2025). Low-cost polyphe-nol–polypyrrole molecularly imprinted sensor for point-of-care Alzheimer's detection. *ACS Sensors*, 10(11), 8435–8446. <https://doi.org/10.1021/acssensors.5c01816>
- [37] Shevchenko, K. G., Garkushina, I. S., Canfarotta, F., Piletsky, S. A., & Barlev, N. A. (2022). Nano-molecularly imprinted polymers (nanoMIPs) as a novel approach to targeted drug delivery in nanomedicine. *RSC Advances*, 12(7), 3957–3968. <https://doi.org/10.1039/D1RA08385F>
- [38] Yarahmadi, B., Hashemianzadeh, S. M., & Milani Hosseini, S. M.-R. (2023). Machine-learning-based predictions of imprinting quality using ensemble and non-linear regression algorithms. *Scientific Reports*, 13(1), 12111. <https://doi.org/10.1038/s41598-023-39374-1>
- [39] Rajpal, S., Mishra, P., & Mizaikoff, B. (2023). Rational in silico design of molecularly imprinted polymers: Current challenges and future potential. *International Journal of Molecular Sciences*, 24(7), 6785. <https://doi.org/10.3390/ijms24076785>
- [40] Xu, S., Wang, L., & Liu, Z. (2021). Molecularly imprinted polymer nanoparticles: An emerging versatile platform for cancer therapy. *Angewandte Chemie International Edition*, 60(8), 3858–3869. <https://doi.org/10.1002/anie.202005309>
- [41] Abdel-Hamied, M., Rajpal, S., Mishra, P., Kranz, C., & Mizaikoff, B. (2026). Molecularly imprinted polymers for biomarker detection: Advances in design, modeling, and sensing strategies. *TrAC Trends in Analytical Chemistry*, 194 (Part A), 118487. <https://doi.org/10.1016/j.trac.2025.118487>
- [42] Brycht, M., Poltorak, L., Baluchová, S., Sipa, K., Borgul, P., Rudnicki, K., & Skrzypek, S. (2024). Electrochemistry as a powerful tool for investigations of antineoplastic agents: A comprehensive review. *Critical Reviews in Analytical Chemistry*, 54(5), 1017–1108. <https://doi.org/10.1080/10408347.2022.2106117>
- [43] Chen, L., Xu, J., Wang, T., Huang, Y., Yuan, D., & Gong, Z. (2021). Toward a versatile flow technique: Development and application of reverse flow dual-injection analysis (rFDIA) for determining dissolved iron redox species and soluble reactive phosphorus in seawater. *Talanta*, 232, 122404. <https://doi.org/10.1016/j.talanta.2021.122404>
- [44] Cetinkaya, A., Kaya, S. I., Alahmad, W., Bellur Atici, E., & Ozkan, S. A. (2023). Designing an electrochemical sensor based on ZnO nanoparticle-supported molecularly imprinted polymer for ultra-sensitive and selective detection of sorafenib. *Analytica Chimica Acta*, 1280, 341866. <https://doi.org/10.1016/j.aca.2023.341866>
- [45] Martins, R. O., Batista Junior, A. C., Melo Brito, C. C. S., Rocha, Y. A., & Chaves, A. R. (2024). Molecularly imprinted polymers in solid-phase microextraction: Enhancing the analysis of pharmaceutical compounds across diverse sample matrices. *Journal of Pharmaceutical and Biomedical Analysis Open*, 4, 100037. <https://doi.org/10.1016/j.jpba.2024.100037>
- [46] Ahmed, S. A., Roushani, M., & Karazan, Z. M. (2025). Development of a molecularly imprinted electrochemical sensor for erlotinib detection using a modified screen-printed carbon electrode. *Microchimica Acta*, 192(3), 138. <https://doi.org/10.1007/s00604-025-07012-4>
- [47] Hong, D., Wang, C., Gao, L., & Nie, C. (2024). Fundamentals, synthetic strategies and applications of non-covalently imprinted polymers. *Molecules*, 29(15), 3555. <https://doi.org/10.3390/molecules29153555>
- [48] Farshi Azhar, F., Ahmadi, M., & Khoshmaram, L. (2024). Chitosan-graphene quantum dot-based molecular imprinted polymer for oxaliplatin release. *Journal of Biomaterials Science, Polymer Edition*, 35(14), 2115–2136. <https://doi.org/10.1080/09205063.2024.2366645>
- [49] Preda, D., Radu, G. L., Iorgulescu, E.-E., Cheregi, M.-C., & David, I. G. (2024). Curcumin-based molecularly imprinted polymer electropolymerized on single-use graphite electrode for dipyrindamole analysis. *Molecules*, 29(19), 4630. <https://doi.org/10.3390/molecules29194630>
- [50] Guo, Z., Zheng, H., Ma, J., Xu, G., & Jia, Q. (2024). Design of pH-responsive molecularly imprinted polymer as a carrier for controlled and sustainable capecitabine release. *Analytica Chimica Acta*, 1317, 342881. <https://doi.org/10.1016/j.aca.2024.342881>
- [51] Fernandes, E., Lopes, C. M., & Lúcio, M. (2024). Lipid biomimetic models as simple yet complex tools to predict skin permeation and drug-membrane biophysical interactions. *Pharmaceutics*, 16(6), 807. <https://doi.org/10.3390/pharmaceutics16060807>
- [52] Mota Díaz, I. I., Douda, J., García López, P., Cabrera Becerra, S. E., Gómez Álvarez, M. A., Jiménez Rodríguez, R., . . . , & López Sánchez, P. (2024). Co-encapsulation of paclitaxel and doxorubicin in liposomes layer by layer. *Colloids and Interfaces*, 8(4), 42. <https://doi.org/10.3390/colloids8040042>
- [53] Migliorini, F., Cini, E., Dreassi, E., Finetti, F., Ievoli, G., Macrì, G., . . . , & Taddei, M. (2022). A pH-responsive crosslinker platform for antibody-drug conjugate (ADC) targeting delivery. *Chemical Communications*, 58(75), 10532–10535. <https://doi.org/10.1039/D2CC03052G>

- [54] Gmeiner, W. H., & Okechukwu, C. C. (2023). Review of 5-FU resistance mechanisms in colorectal cancer: Clinical significance of attenuated on-target effects. *Cancer Drug Resistance*, 6(2), 257–272. <https://doi.org/10.20517/cdr.2022.136>
- [55] de Moraes, F. C. A., de laia, E. A., Sano, V. K. T., dos Santos, A. G. E., Pereira, C. R. M., & Burbano, R. M. R. (2025). A systematic review and meta-analysis of users versus non-users: Unveiling the influence of proton pump inhibitors on capecitabine efficacy in colorectal cancer. *Expert Review of Clinical Pharmacology*, 18(1–2), 67–75. <https://doi.org/10.1080/17512433.2024.2443183>
- [56] Önal Acet, B., İnanan, T., Salieva, K., Borkoev, B., Odabaşı, M., & Acet, Ö. (2024). Molecular imprinted polymers: Important advances in biochemistry, biomedical and biotechnology. *Polymer Bulletin*, 81(12), 10439–10459. <https://doi.org/10.1007/s00289-024-05238-5>
- [57] Liu, S., Han, S., Song, Y., Sun, R., Zhao, L., & Yan, C. (2023). Disulfide-bridged dendritic organosilicas-based biodegradable molecularly imprinted polymers for multiple targeting and pH/redox-responsive drug release toward chemical/photodynamic synergistic tumor therapy. *Advanced Healthcare Materials*, 12(20), 2300184. <https://doi.org/10.1002/adhm.202300184>
- [58] Ahadi, H. M., Fardhan, F. M., Rahayu, D., Pratiwi, R., & Hasanah, A. N. (2024). Molecularly imprinted microspheres in active compound separation from natural product. *Molecules*, 29(17), 4043. <https://doi.org/10.3390/molecules29174043>
- [59] Nasiriani, T., Javanbakht, S., Shaabani, A., & Kazeminava, F. (2024). Gelatin-based carbon quantum dot-molecularly imprinted polymer: Safe photoluminescent core-shell nanocarrier for the pH-responsive anticancer drug delivery. *International Journal of Biological Macromolecules*, 278, 134669. <https://doi.org/10.1016/j.ijbiomac.2024.134669>
- [60] Dobrzyńska, J., Dąbrowska, M., Olchowski, R., Zięba, E., & Dobrowolski, R. (2021). Development of a method for removal of platinum from hospital wastewater by novel ion-imprinted mesoporous organosilica. *Journal of Environmental Chemical Engineering*, 9(4), 105302. <https://doi.org/10.1016/j.jece.2021.105302>
- [61] Han, D., Han, H., Chen, E., Cesarski, W. J., Lim, H., Shin, H., . . . , & Jeong, K. (2025). Integrating electronic structure theory calculations with Python and cheminformatics tools. *Journal of Chemical Education*, 102(9), 4115–4122. <https://doi.org/10.1021/acs.jchemed.4c01171>
- [62] Byadi, S., Gantzer, P., Gimadiev, T., & Sidorov, P. (2025). DOPtools: A Python platform for descriptor calculation and model optimization. *Digital Discovery*, 4(5), 1188–1198. <https://doi.org/10.1039/D4DD000399C>
- [63] Gray, M., & Wu, L. (2025). Comparative study of molecular descriptors and AI-based embeddings for toxicity prediction. *Chemical Research in Toxicology*, 38(12), 2061–2070. <https://doi.org/10.1021/acs.chemrestox.5c00289>
- [64] Akgönüllü, S., Kılıç, S., Esen, C., & Denizli, A. (2023). Molecularly imprinted polymer-based sensors for protein detection. *Polymers*, 15(3), 629. <https://doi.org/10.3390/polym15030629>
- [65] Zhang, J., Zhang, M., Dong, B., & Ma, H. (2022). Quantitative evaluation of steel corrosion induced deterioration in rubber concrete by integrating ultrasonic testing, machine learning and mesoscale simulation. *Cement and Concrete Composites*, 128, 104426. <https://doi.org/10.1016/j.cemconcomp.2022.104426>
- [66] Aldeghi, M., & Coley, C. W. (2022). A graph representation of molecular ensembles for polymer property prediction. *Chemical Science*, 13(35), 10486–10498. <https://doi.org/10.1039/D2SC02839E>
- [67] Gupta, V., Choudhary, K., Tavazza, F., Campbell, C., Liao, W., Choudhary, A., & Agrawal, A. (2021). Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. *Nature Communications*, 12(1), 6595. <https://doi.org/10.1038/s41467-021-26921-5>
- [68] Jiang, S., & Webb, M. (2025). Physics-guided neural networks for transferable property prediction in architecturally diverse copolymers. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv-2024-ld2k6-v2>
- [69] Huang, Q., Li, Y., Zhu, L., Zhao, Q., & Yu, W. (2025). Unified multimodal multidomain polymer representation for property prediction. *npj Computational Materials*, 11(1), 153. <https://doi.org/10.1038/s41524-025-01652-z>

How to Cite: R. C., J., Prasanth, N., Perumkunnappala, A., Radhakrishnan, M. K. P., Nair, M., & Suravajhala, R. (2026). Enhancing Molecularly Imprinted Polymer Development with Machine Learning Techniques. *Journal of Computational and Cognitive Engineering*. <https://doi.org/10.47852/bonviewJCCE62027832>