

Interpretable Machine Learning for Species-Specific Prediction of *Helicobacter pylori* Urease Inhibitors

Anthony Peter

`anthony.peter@udsm.ac.tz`

University of Dar es Salaam

Research Article

Keywords: *Helicobacter pylori*, urease inhibitors, QSAR, machine learning, molecular fingerprints, SHAP

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11377834/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Interpretable Machine Learning for Species-Specific Prediction of *Helicobacter pylori* Urease Inhibitors

Anthony A. Peter^{1,2,*}

ORCID: 0009-0008-2030-6361

¹Department of Chemistry, Mkwawa University College of Education, P.O. Box 2513, Iringa, Tanzania

²University of Dar es Salaam, P.O. Box 35091, Dar es Salaam, Tanzania

*Correspondence: anthony.peter@udsm.ac.tz

Abstract

Helicobacter pylori depends on urease for survival in the acidic stomach, yet inhibitor-prediction studies frequently rely on measurements from different urease sources, limiting their relevance to the pathogen-specific target. To address this problem, an interpretable quantitative structure–activity relationship (QSAR) model was developed from 351 compounds with experimentally measured *H. pylori* urease inhibitory activity. PubChem molecular fingerprints were used to predict continuous pIC₅₀ values, with LightGBM selected through cross-validation. The model achieved a five-fold cross-validated Q^2_{CV} of 0.525 and an independent-test R^2 of 0.782, while response randomization substantially diminished test-set predictivity. SHAP analysis revealed contrasting contributions from chemically related carbonyl-containing patterns, alongside associations involving nitrogen, chlorine, and oxygenated molecular fragments. These findings suggest that the model distinguishes local structural environments rather than assigning uniform contributions to broad functional-group classes. Applicability-domain assessment further characterized the chemical space supporting its predictions. The trained model was made accessible through PanURA for individual and batch potency estimation. By integrating pathogen-specific bioactivity data with interpretable molecular predictions, this study establishes a practical framework for prioritizing candidate urease inhibitors and investigating structure–activity relationships through subsequent experimental studies.

Keywords: *Helicobacter pylori*; urease inhibitors; QSAR; machine learning; molecular fingerprints; SHAP

1 Introduction

Helicobacter pylori remains one of the most consequential bacterial pathogens in human health because it colonizes the stomach chronically, drives peptic ulcer disease, and is a major risk factor for gastric adenocarcinoma. Contemporary epidemiological analyses continue to show a substantial worldwide burden of infection, with prevalence differing markedly across regions and socioeconomic settings [19]. Clinically, the organism is important not only because infection is common, but also because long-term colonization can persist silently before progressing to severe gastroduodenal disease or malignancy [7]. International consensus guidance therefore emphasizes accurate diagnosis, eradication, and improved therapeutic decision-making as central priorities in the management of *H. pylori* infection [22].

Despite the availability of eradication regimens, treatment failure has become increasingly difficult to ignore. Rising resistance to clarithromycin, metronidazole, and levofloxacin has reduced the reliability of empirical antibiotic therapy in many settings, forcing guideline panels to recommend susceptibility-aware treatment strategies where possible [22, 33]. This therapeutic pressure creates a strong rationale for complementary

discovery strategies that move beyond conventional antibacterial killing and instead exploit pathogen-specific survival mechanisms. In this context, antivirulence targeting of *H. pylori* urease is attractive because urease is central to gastric colonization, acid acclimation, and persistence within the hostile gastric niche [31].

Urease enables *H. pylori* to hydrolyze urea into ammonia and carbon dioxide, thereby buffering the local environment and supporting survival under acidic stress. Structural and biochemical studies have shown that *H. pylori* urease is not merely a generic urease enzyme, but a highly organized nickel-dependent virulence system whose supramolecular assembly and active-site architecture are closely linked to acid resistance and inhibitor binding [8, 12]. Recent reviews further emphasize urease maturation, nickel transfer, auxiliary protein interactions, and covalent or metal-coordinating chemotypes as active directions for *H. pylori* urease inhibitor discovery [6, 11]. These features make urease a rational molecular target for selective inhibitor discovery, particularly because the enzyme is essential to the organism's colonization strategy but absent from human metabolism [15]. Very recent functional mapping further suggests that the *H. pylori* urease catalytic pocket is highly conserved and may represent a resistance-resilient target, especially when urease inhibitors are considered as adjuncts to antibiotic therapy [35]. A successful urease inhibitor could therefore weaken bacterial fitness in the stomach while potentially reducing direct selective pressure on canonical antibiotic targets.

However, the urease inhibitor literature is chemically rich but biologically fragmented. Many inhibitor studies use jack bean urease or other non-*H. pylori* ureases as experimental surrogates, even though early comparative enzymology showed that *H. pylori* urease and jack bean urease differ in kinetic behavior and inhibition profiles [5, 25]. The most recent experimental literature reinforces this heterogeneity: natural-product studies now compare *H. pylori* urease and jack bean urease or use crude *H. pylori* urease extracts, synthetic medicinal chemistry has reported *H. pylori*-focused ebselen/benzisoselenazolone and indole-based leads, whereas docking-led natural-product work shows that biological validation still varies across enzyme sources and assay systems [3, 4, 10, 16, 37, 42]. This matters because apparent urease inhibition in one enzyme source may not translate directly into species-specific activity against *H. pylori*. As a result, compound prioritization based only on pooled urease data risks favoring molecules that are broadly reactive, assay-dependent, or optimized for a surrogate enzyme rather than for the pathogen-relevant target. The unresolved species-specific translation problem is therefore a genuine gap in urease inhibitor discovery.

Machine learning has begun to address parts of this problem by extracting activity patterns from chemical descriptors, fingerprints, and molecular docking outputs. Recent studies have demonstrated that supervised learning can classify urease inhibitors, support virtual screening, and guide hydroxamic acid-based inhibitor design for *H. pylori* urease [2, 23, 39]. In parallel, a 2025 genome-wide interpretable machine learning study used feature-attribution analysis to explain *H. pylori* antibiotic-resistance prediction, demonstrating that explainable modeling is becoming increasingly relevant to *H. pylori* research beyond small-molecule classification [40]. Nevertheless, these advances also expose the next methodological need: models must be trained and interpreted in ways that respect enzyme source, assay context, chemical domain, and activity definition. Modern QSAR practice stresses rigorous curation, applicability-domain assessment, external validation, and transparent interpretation to avoid overfitting and misleading structure–activity conclusions [24]. Thus, the key gap is no longer whether machine learning can be applied to urease inhibitors, but whether interpretable models can deliver reliable, species-specific predictions for *H. pylori* urease.

This study develops an interpretable machine learning framework for species-specific prediction of *Helicobacter pylori* urease inhibitor potency using curated pathogen-specific bioactivity data. The framework integrates PubChem molecular fingerprints, supervised regression modeling, rigorous model validation, applicability-domain assessment, and SHAP-based feature attribution within a transparent QSAR workflow. By combining species-specific endpoint definition with interpretable prediction, the framework enables

biologically meaningful interpretation of model outputs and supports rational prioritization of candidate urease inhibitors, consistent with the growing role of explainable artificial intelligence in drug discovery [14]. The resulting framework provides a transparent benchmark for species-specific virtual screening and lead optimization of *H. pylori* urease inhibitors.

2 Materials and Methods

2.1 Species-Specific Endpoint Curation

Only experimentally determined IC₅₀ values measured against *H. pylori* urease were included for model construction. Bioactivity records were obtained primarily from the ChEMBL database using the *H. pylori* urease target identifier ChEMBL3885651, supplemented where necessary with peer-reviewed urease inhibitor studies and PubChem BioAssay records when compound structure, biological endpoint, enzyme source, and assay conditions were available [17, 41]. The initial ChEMBL retrieval yielded 617 compounds associated with *H. pylori* urease bioactivity records. Records reporting K_i, single-concentration percentage inhibition, qualitative activity calls, non-IC₅₀ endpoints, ambiguous structures, unspecified urease source, missing units, antibacterial growth inhibition without urease measurement, or mixtures without a dominant characterized constituent were excluded. Empty entries, duplicate compounds, records without IC₅₀ endpoints, and censored IC₅₀ values reported with inequality signs such as < or > were also removed. Surrogate data from jack bean urease or other non-*H. pylori* ureases were not merged with the *H. pylori*-specific regression endpoint.

Retrieved structures were standardized before descriptor calculation. Salts, counterions, duplicate fragments, and inorganic additives were removed when they did not represent the active molecule; valence states, aromaticity, charges, and tautomeric forms were normalized consistently; and duplicate entries were reconciled using canonical SMILES and source identifiers. Retained IC₅₀ values were converted to molar units, and replicate measurements for the same compound against *H. pylori* urease were summarized using the median IC₅₀ value. Activity values were transformed to pIC₅₀ according to Equation 1. These curation, standardization, and endpoint-transformation steps produced a final curated dataset of 351 compounds for regression model development.

$$\text{pIC}_{50} = -\log_{10}(\text{IC}_{50} [\text{M}]) \quad (1)$$

No binary activity labels were used for model training. A pIC₅₀ threshold of 5 was used only for descriptive exploratory summaries of dataset composition and was not used as a machine learning endpoint.

2.2 Molecular Representation and Modeling Cohort

2.2.1 Physicochemical Descriptor Calculation

Physicochemical descriptors were calculated to support exploratory data analysis and dataset characterization. The calculated summary descriptors included molecular weight (MW), hydrogen bond acceptor count (HBA), hydrogen bond donor count (HBD), and LogP. These descriptors were used exclusively for descriptive visualization of the curated compound set and were not included in the QSAR modeling matrix.

2.2.2 PubChem Fingerprint Generation

PubChem fingerprints were calculated for each compound using the PubChemFingerprint implementation in `skfp.fingerprints`. Each 881-bit vector encoded the presence or absence of predefined structural patterns and served as the QSAR descriptor space.

2.2.3 Dataset Partitioning and Feature Selection

The curated dataset of 351 compounds was partitioned into training (80%, $n = 280$) and independent test (20%, $n = 71$) sets using the Kennard–Stone algorithm based on the original 881-bit PubChem fingerprints. Descriptor selection was performed on the training set in two stages. First, low-variance fingerprint bits were removed using a variance threshold of 0.01. Subsequently, pairwise Pearson correlation analysis was applied, retaining one descriptor from each pair with an absolute correlation coefficient of $|r| \geq 0.85$. During five-fold cross-validation, both filtering steps were fitted separately within each training fold and applied to its corresponding validation fold. For final model development, the filters were fitted on the complete training set, and the resulting descriptor subset was applied unchanged to the independent test set. This procedure retained 142 descriptors for final model development.

2.3 Model Benchmarking and Validation

2.3.1 Baseline Model Benchmarking

The 280-compound Kennard–Stone training set was randomly divided into development (80%, $n = 224$) and validation (20%, $n = 56$) subsets using a fixed random seed of 42. LazyPredict was used to compare 41 regression algorithms under their default parameter settings, without hyperparameter optimization. Low-variance fingerprint descriptors were removed using a variance threshold of 0.01, followed by pairwise Pearson correlation filtering at $|r| \geq 0.85$; both filtering steps were fitted on the development subset and applied unchanged to the validation subset. Successfully evaluated algorithms were ranked by validation R^2 , while RMSE and computational runtime were recorded as supplementary benchmarking measures. Up to five highest-ranked algorithms that successfully completed subsequent five-fold cross-validation were retained for comparative model evaluation.

2.3.2 Model Retraining and Performance Evaluation

The five selected candidates were LightGBM (LGBM), HistGradientBoosting (HGB), GradientBoosting (GB), RandomForest (RF), and OrthogonalMatchingPursuit (OMP). The selected regression algorithms were fitted using the complete 280-compound training set and evaluated by five-fold cross-validation with shuffled folds and a fixed random seed of 42. Each cross-validation iteration incorporated variance and correlation filtering fitted on the corresponding training fold and applied to its validation fold. Cross-validated predictions were used to calculate Q_{CV}^2 , $RMSE_{CV}$, and MAE_{CV} , while training-set predictions were evaluated using R_{train}^2 , $RMSE_{train}$, and MAE_{train} . The coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) were calculated using Equations 2, 4, and 5, respectively, whereas Q_{CV}^2 was calculated using Equation 3. Candidate models were ranked according to Q_{CV}^2 , and the highest-ranked model was selected for subsequent interpretation. Predictions for the independent 71-compound test set were evaluated using R_{test}^2 , $RMSE_{test}$, and MAE_{test} . External predictive performance was further examined using R_{ext}^2 (Equation 6), calculated relative to the training-set mean, and the regression-through-origin statistics k , k' , R_0^2 , and $R_0'^2$ (Equations 7 and 8). The external validation criteria summarized in Table 1 were applied to assess predictive performance and agreement between observed and predicted pIC_{50} values.

The coefficient of determination for an evaluated set was calculated as follows:

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

In Equation 2, n is the number of evaluated compounds, y_i is the observed pIC_{50} value for compound i , $\hat{y}_i$ is the corresponding predicted pIC_{50} value, and $\bar{y}$ is the mean observed pIC_{50} value for the evaluated set.

The five-fold cross-validated coefficient was calculated on the training set as follows:

$$Q_{\text{CV}}^2 = 1 - \frac{\sum_{i=1}^{n_{\text{train}}} (y_i - \hat{y}_{i,\text{CV}})^2}{\sum_{i=1}^{n_{\text{train}}} (y_i - \bar{y}_{\text{train}})^2} \quad (3)$$

In Equation 3, n_{train} is the number of training compounds, $\hat{y}_{i,\text{CV}}$ is the cross-validated prediction for training compound i , and $\bar{y}_{\text{train}}$ is the mean observed pIC_{50} value in the training set.

The root mean squared error was calculated as follows:

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

In Equation 4, n , y_i , and $\hat{y}_i$ are defined as in Equation 2. The mean absolute error was calculated as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

In Equation 5, n , y_i , and $\hat{y}_i$ are defined as in Equation 2.

The independent 71-compound test set was evaluated using R^2 , RMSE, and MAE as defined in Equations 2, 4, and 5. The external coefficient R_{ext}^2 was calculated with the training-set mean in the denominator [9, 26, 38]:

$$R_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y}_{\text{train}})^2} \quad (6)$$

In Equation 6, n_{test} is the number of test compounds, y_i and $\hat{y}_i$ are the observed and predicted test-set pIC_{50} values, and $\bar{y}_{\text{train}}$ is the mean observed pIC_{50} value in the training set.

OECD external validation statistics were computed for the selected models using the independent test-set predictions. For predicted test-set values regressed through the origin against observed test-set values, the quantities were calculated as follows [9, 26]:

$$k = \frac{\sum_{i=1}^{n_{\text{test}}} y_i \hat{y}_i}{\sum_{i=1}^{n_{\text{test}}} \hat{y}_i^2}$$

$$R_0^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - k y_i)^2}{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - \bar{\hat{y}})^2} \quad (7)$$

In Equation 7, k is the regression-through-origin slope for predicted versus observed test-set values, R_0^2 is the corresponding squared coefficient without intercept, and $\bar{\hat{y}}$ is the mean predicted pIC_{50} value in the test set.

For observed test-set values regressed through the origin against predicted test-set values, the quantities were calculated as follows:

$$k' = \frac{\sum_{i=1}^{n_{\text{test}}} y_i \hat{y}_i}{\sum_{i=1}^{n_{\text{test}}} \hat{y}_i^2}$$

$$R_0'^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y_i - k' \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y}_{\text{test}})^2} \quad (8)$$

In Equation 8, k' is the regression-through-origin slope for observed versus predicted test-set values, $R_0'^2$ is the corresponding squared coefficient without intercept, and $\bar{y}_{\text{test}}$ is the mean observed pIC₅₀ value in the test set. The implemented PASS_OECD decision was recorded from the external-validation statistics.

The threshold criteria used for model evaluation are summarized in Table 1. These criteria followed commonly used QSAR validation recommendations for goodness of fit, cross-validation, independent test-set predictivity, and regression-through-origin checks, while RMSE, MAE, runtime, and Y-randomization statistics were retained as comparative or robustness outputs [9, 26, 30, 38]. The PASS_OECD decision was assigned only when the independent test-set predictivity thresholds, slope checks, and origin-deviation criterion were collectively satisfied; failure of any required component resulted in a False decision.

Table 1

Model evaluation criteria used for benchmark screening, selected-model evaluation, OECD validation, and robustness assessment

Evaluation component	Criterion applied	Decision use
Training fit	$R_{\text{train}}^2 > 0.60$	Fit diagnostic for retrained models
Cross-validation	$Q_{\text{CV}}^2 > 0.50$	Internal validation criterion
Independent test-set predictivity	$R_{\text{test}}^2 > 0.60$ and $R_{\text{ext}}^2 > 0.60$	External validation criterion
Prediction error	Lower RMSE _{test} and MAE _{test} preferred	Comparative error ranking
OECD slope checks	$0.85 < k < 1.15$ and $0.85 < k' < 1.15$	PASS_OECD component
Origin-deviation check	$(R_{\text{test}}^2 - R_0^2)/R_{\text{test}}^2 < 0.10$ or $(R_{\text{test}}^2 - R_0'^2)/R_{\text{test}}^2 < 0.10$	PASS_OECD component
Y-randomization	$cR_p^2 > 0.5$; randomized R_{test}^2 substantially lower and near zero or negative	Chance-correlation check
Runtime	Shorter runtime preferred when predictive metrics were comparable	Computational benchmarking

2.3.3 Model Robustness Assessment

Y-randomization was conducted for the five selected regression models using 100 independent permutations of the training-set pIC₅₀ values, with a fixed random seed of 42 [24, 38]. For each permutation, response values were randomly reassigned among the training compounds while the molecular descriptors, Kennard–Stone partition, and model parameters remained unchanged. Each algorithm was refitted using the permuted responses, and predictions were generated for both training and independent test compounds. The resulting R_{train}^2 , R_{test}^2 , RMSE_{test}, and MAE_{test} values were recorded and summarized as means and standard deviations across the 100 iterations. Randomized-model results were compared with those obtained using the original response values, and the corrected randomization coefficient (cR_p^2) was calculated to assess the contribution of chance correlations using Equation 9. The criterion $cR_p^2 > 0.5$ was used as an additional robustness indicator. Y-randomization and external validation were used to characterize the performance of the evaluated models, while final model selection was based on five-fold cross-validation performance.

$$cR_p^2 = R \sqrt{R^2 - R_{\text{rand}}^2} \quad (9)$$

In Equation 9, R is the Pearson correlation between observed and predicted pIC_{50} values on the original-response training set, and R_{rand} is the mean Pearson correlation between the permuted training responses and their corresponding predictions across the 100 Y-randomized models. This corrected coefficient follows established QSAR formulations [27, 29].

2.4 Interpretable Feature Attribution

SHapley Additive exPlanations (SHAP) analysis was performed on the final regression model using the Python SHAP library and the selected PubChem fingerprint descriptors from the training set [20, 21, 28, 34]. SHAP values (ϕ_{ij}) quantified the contribution of each fingerprint descriptor j to the predicted pIC_{50} of compound i relative to the model’s baseline prediction. Positive and negative values indicated contributions toward higher and lower predicted pIC_{50} , respectively.

Global descriptor importance was quantified using the mean absolute SHAP value across the training compounds:

$$I_j = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} |\phi_{ij}| \quad (10)$$

The ten highest-ranked descriptors were visualized using SHAP bar and beeswarm plots to examine their relative importance, direction, and distribution of contributions. PubChem fingerprint definitions for these descriptors were mapped from the PubChem fingerprint dictionary to support chemical interpretation. SHAP values were used only for post hoc interpretation and were not used during model fitting, feature filtering, or endpoint definition.

2.5 Applicability-Domain Definition

A leverage-based applicability-domain assessment was performed for the final descriptor matrix used to train the selected LGBM. The conventional warning leverage threshold for sensitivity analysis was calculated as follows:

$$h^* = \frac{3(p+1)}{n} \quad (11)$$

where h^* is the conventional warning leverage threshold, $p = 142$ is the number of retained descriptors, and $n = 280$ is the number of training compounds. The primary leverage cutoff was the empirical 95th percentile of training-set leverage ($h_{p95} = 0.988$); the 99th-percentile cutoff ($h_{p99} = 1.000$) and conventional $h^* = 1.532$ were evaluated as sensitivity thresholds. Residuals were scaled by the training cross-validation RMSE and assessed against warning limits of ± 3 in the Williams-style diagnostic plot.

2.6 Web Application Development and Deployment

The final LightGBM QSAR model was integrated into PanURA, a web application developed using Streamlit [36]. The interface supports individual molecular input through SMILES notation or an interactive JSME editor, alongside batch screening using .csv, .txt, and .xlsx files. Molecular structures are processed using RDKit v2025.9.6 [18], and PubChem fingerprints are generated using scikit-fingerprints [1]. The stored feature-selection manifest preserves the descriptor subset and ordering established during model development, enabling predictions from the serialized LightGBM model without retraining. Predicted pIC_{50} values are converted to IC_{50} estimates in μM according to $\text{IC}_{50} (\mu\text{M}) = 10^{6-\text{pIC}_{50}}$. Results are presented interactively, and batch predictions can be exported as CSV files. The application architecture is illustrated in Figure 1; source code and software dependencies are available at <https://github.com/peterDataScientia/panURA>.

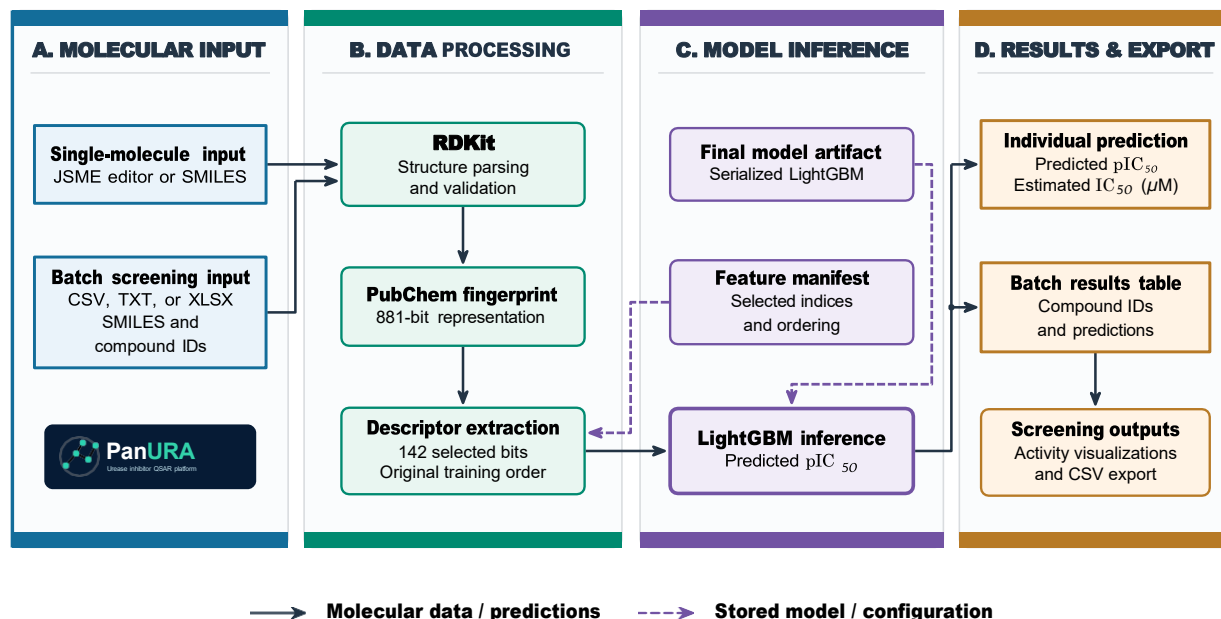


Fig. 1 Architecture and prediction workflow of PanURA, from compound input and molecular processing through fingerprint generation, model prediction, and result presentation

3 Results

3.1 Species-Specific Dataset and Descriptor Space

The initial ChEMBL retrieval for target ChEMBL3885651 yielded 617 compounds associated with *H. pylori* urease bioactivity records. The curation workflow described in Section 2.1 retained 351 compounds for regression model development, corresponding to 56.9% of the initially retrieved compound set. The final endpoint was represented as continuous pIC₅₀ values. For descriptive dataset characterization only, a pIC₅₀ threshold of 5 separated 133 descriptive Active entries (pIC₅₀ ≥ 5) from 218 descriptive Inactive entries (pIC₅₀ < 5); this threshold was not used as a model label. Fig. 2 summarizes the retained modeling cohort and descriptive activity distribution.

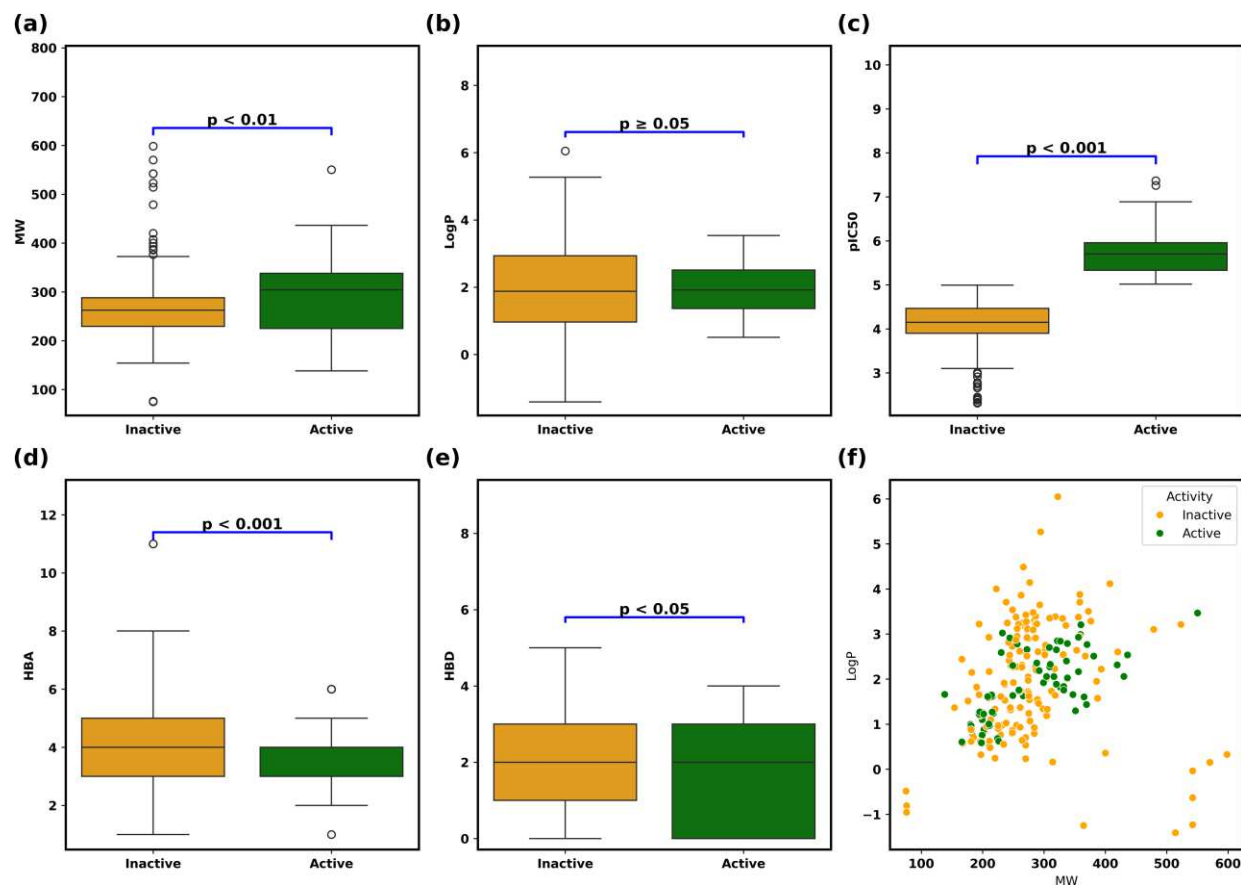


Fig. 2 Exploratory data analysis of the retained *H. pylori* urease inhibitor modeling cohort. The final dataset contained 351 compounds; for descriptive summary only, 133 entries were Active ($\text{pIC}_{50} \geq 5$) and 218 entries were Inactive ($\text{pIC}_{50} < 5$)

The curated compounds showed chemically diverse physicochemical profiles, reflecting the broad range of urease inhibitor chemotypes reported in the literature. The descriptor distributions in Figure 2 indicate that the dataset included both relatively small polar inhibitors and larger hydrophobic scaffolds, consistent with the known diversity of hydroxamate, heterocyclic, selenium-containing, and natural-product-derived urease inhibitor classes. PubChem fingerprint processing began with 881 bits. Variance filtering retained 415 features, removing 466 low-variance bits, and subsequent correlation filtering retained 142 descriptors. These filtering steps reduced descriptor redundancy while retaining a chemically informative representation for pIC_{50} regression.

3.2 Model Selection and Predictive Performance

The LazyPredict screen was used as a preliminary assessment of model families rather than as the final selection procedure. Under default settings, LGBM and HGB had nearly identical screening performance ($R^2 = 0.546$ and $\text{RMSE} = 0.665$ for each). Because this screen used a development-validation split, its values are reported separately from the final retraining results below.

The five shortlisted regressors were LGBM, HGB, GB, RF, and OMP. They were retrained and evaluated using the 142 retained PubChem fingerprint descriptors, five-fold cross-validation on the 280-compound training set, and independent evaluation on the 71-compound test set. Table 2 summarizes their

performance. LGBM had the highest cross-validated Q_{CV}^2 and test-set R_{test}^2 , although HGB performed virtually identically. The automatic selection procedure therefore chose LGBM, while the negligible difference between LGBM and HGB does not indicate a meaningful performance separation.

LGBM achieved $Q_{CV}^2 = 0.525$, $R_{test}^2 = 0.782$, $RMSE_{test} = 0.493$, and $MAE_{test} = 0.353$. Its external R_{ext}^2 was 0.783, and it passed the implemented OECD criteria. HGB produced nearly identical test metrics ($R_{test}^2 = 0.782$; $RMSE_{test} = 0.493$; $MAE_{test} = 0.353$). RF ranked next by test-set R_{test}^2 (0.715), followed by GB (0.659) and OMP (0.551). The corresponding observed-versus-predicted parity plots are shown in **Fig. 3**.

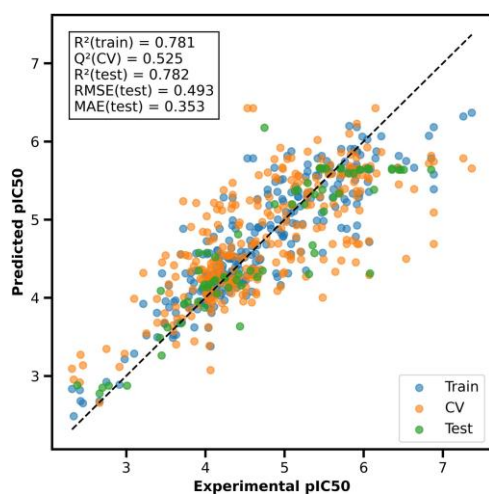
Table 2

Training, five-fold cross-validation, and independent test-set performance of the five retrained regressors using 142 retained PubChem fingerprint descriptors

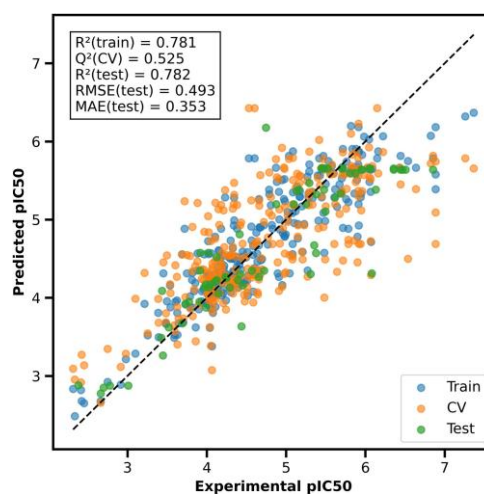
Model	R_{train}^2	Q_{CV}^2	R_{test}^2	$RMSE_{train}$	$RMSE_{CV}$	$RMSE_{test}$	MAE_{train}	MAE_{CV}	MAE_{test}
LGBM	0.781	0.525	0.782	0.431	0.635	0.493	0.317	0.479	0.353
HGB	0.781	0.525	0.782	0.431	0.635	0.493	0.317	0.479	0.353
GB	0.844	0.505	0.659	0.364	0.649	0.617	0.279	0.489	0.439
RF	0.930	0.501	0.715	0.244	0.651	0.563	0.175	0.484	0.412
OMP	0.607	0.485	0.551	0.578	0.662	0.708	0.445	0.497	0.490

3.3 External Validation and Y-Randomization

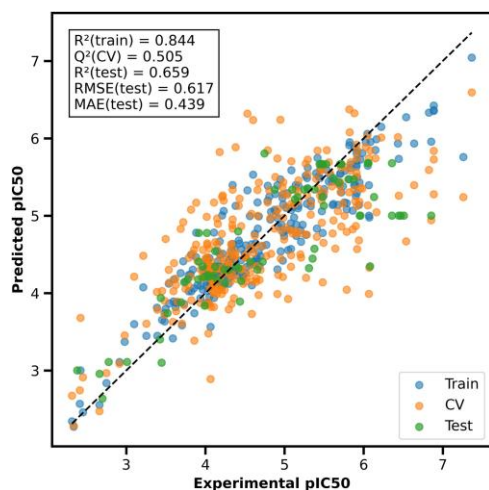
OECD external validation provided regression-consistency checks beyond ordinary test-set R_{test}^2 for the five retrained models. As summarized in Table 3, LGBM and HGB both passed the implemented criteria and had nearly identical external R_{ext}^2 values (0.783). GB and RF also passed, whereas OMP did not. Thus, four of the five retrained models met the implemented OECD validation criteria.



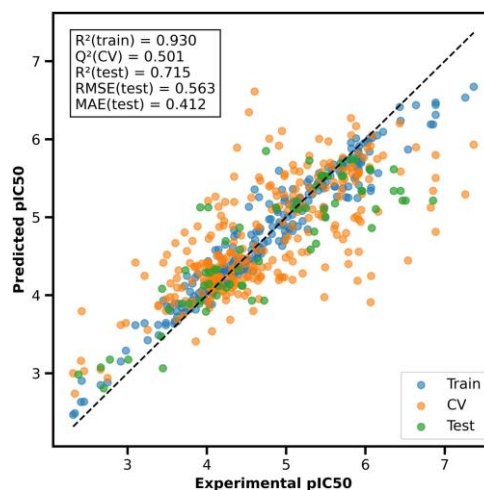
(a) LGBM



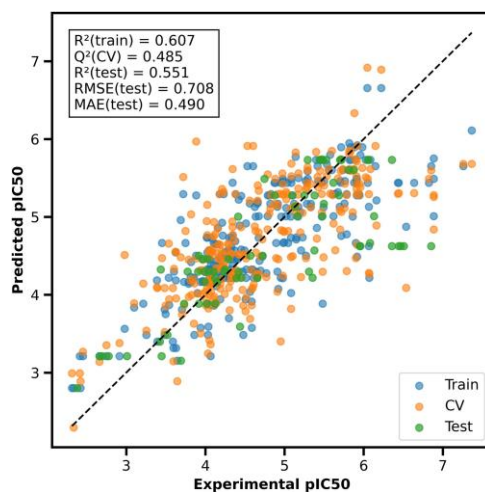
(b) HGB



(c) GB



(d) RF



(e) OMP

Fig. 3 Observed-versus-predicted pIC_{50} parity plots for the five retrained regression models using 142 retained PubChem fingerprint descriptors. Model abbreviations are defined in the Methods

Table 3*OECD external validation statistics for the five retrained regression models*

Model	Q_{CV}^2	R_{test}^2	R_{ext}^2	k	k'	R_0^2	$R_0'^2$	OECD Pass
LGBM	0.525	0.782	0.783	0.965	1.026	0.717	0.796	True
HGB	0.525	0.782	0.783	0.965	1.026	0.717	0.796	True
GB	0.505	0.659	0.660	0.953	1.033	0.459	0.681	True
RF	0.501	0.715	0.716	0.960	1.029	0.556	0.732	True
OMP	0.485	0.551	0.553	0.940	1.043	0.309	0.587	False

Y-randomization served as a negative-control analysis in which pIC_{50} response values were randomized before refitting the selected models. Table 4 summarizes the mean and standard deviation of randomized-model performance and reports the corrected coefficient and strict-threshold result for each model. All five models produced negative mean randomized test R_{test}^2 values. LightGBM ($cR_p^2 = 0.502$), HistGradientBoosting ($cR_p^2 = 0.5000526$), and OrthogonalMatchingPursuit ($cR_p^2 = 0.531$) passed the strict $cR_p^2 > 0.5$ criterion; GradientBoosting (0.432) and RandomForest (0.187) did not. HistGradientBoosting’s value rounds to 0.500 at three decimal places but is above the threshold at full precision. OMP’s higher coefficient did not offset its weaker predictive performance. These results indicate that Y-randomization is a robustness diagnostic and should be interpreted alongside cross-validation and independent test-set performance [29, 30, 38].

Table 4*Y-randomization summary for selected regressors after response-value randomization*

Model	R_{train}^2 (mean $\pm$ SD)	R_{test}^2 (mean $\pm$ SD)	RMSE _{test} (mean $\pm$ SD)	MAE _{test} (mean $\pm$ SD)	cR_p^2	Pass?
LGBM	0.419 $\pm$ 0.032	-0.178 $\pm$ 0.253	1.140 $\pm$ 0.121	0.953 $\pm$ 0.102	0.502	PASS
HGB	0.419 $\pm$ 0.032	-0.154 $\pm$ 0.255	1.128 $\pm$ 0.124	0.946 $\pm$ 0.108	0.5000526	PASS
GB	0.547 $\pm$ 0.033	-0.177 $\pm$ 0.219	1.141 $\pm$ 0.105	0.955 $\pm$ 0.088	0.432	FAIL
RF	0.816 $\pm$ 0.015	-0.264 $\pm$ 0.270	1.181 $\pm$ 0.122	0.973 $\pm$ 0.115	0.187	FAIL
OMP	0.144 $\pm$ 0.027	-0.099 $\pm$ 0.203	1.102 $\pm$ 0.102	0.929 $\pm$ 0.089	0.531	PASS

3.4 Chemical Determinants of Predicted Potency

SHAP analysis of the automatically selected LGBM regressor identified the PubChem fingerprint descriptors that contributed most strongly to predicted pIC_{50} values. PubChem_439 had the largest mean absolute SHAP value, followed by PubChem_646 and PubChem_535. Figure 4 presents schematic representations of the top ten fingerprint patterns and their mean absolute SHAP values; these are fingerprint definitions, not complete inhibitor structures or causal pharmacophores. Figure 5a ranks the descriptors by mean absolute SHAP value, while Figure 5b shows the direction and distribution of their contributions across training compounds. Positive SHAP values indicate contributions toward higher predicted pIC_{50} relative to the model baseline, and negative values indicate contributions toward lower predictions. These attributions are model-based associations, not evidence that individual fingerprint bits cause urease inhibition.

Fingerprint enrichment provided a descriptive comparison of the top SHAP-ranked bits across the activity groups (Figure 6). PubChem_439 was present in 91.0% of active and 54.4% of inactive compounds, while PubChem_14 was present in 97.0% and 72.2%, respectively. PubChem_535, PubChem_37, and

PubChem_143 were also more frequent among active compounds. In contrast, PubChem_516 occurred at similar frequencies in the two groups, and PubChem_700, PubChem_566, PubChem_646, and PubChem_374 were more frequent among inactive compounds. Activity grouping was used only for descriptive interpretation and was not the regression endpoint; frequency differences do not establish causal structure–activity relationships. Figure 7 shows selected high- and low-potency compounds from the curated dataset. The five highest-potency examples included CHEMBL4167553 ($\text{pIC}_{50} = 7.37$), while the lower-potency examples included CHEMBL2377157 ($\text{pIC}_{50} = 2.31$). These structures illustrate chemical diversity within the dataset and should not be interpreted as a matched-pair comparison.

SHAP-value distributions for all ten top-ranked features, grouped by fingerprint-bit absence or presence, are shown across the two parts of Figure 8. The y-axes show each compound's SHAP contribution for the indicated bit, not its measured pIC_{50} ; these grouped distributions describe model attributions and do not establish independent or causal effects.

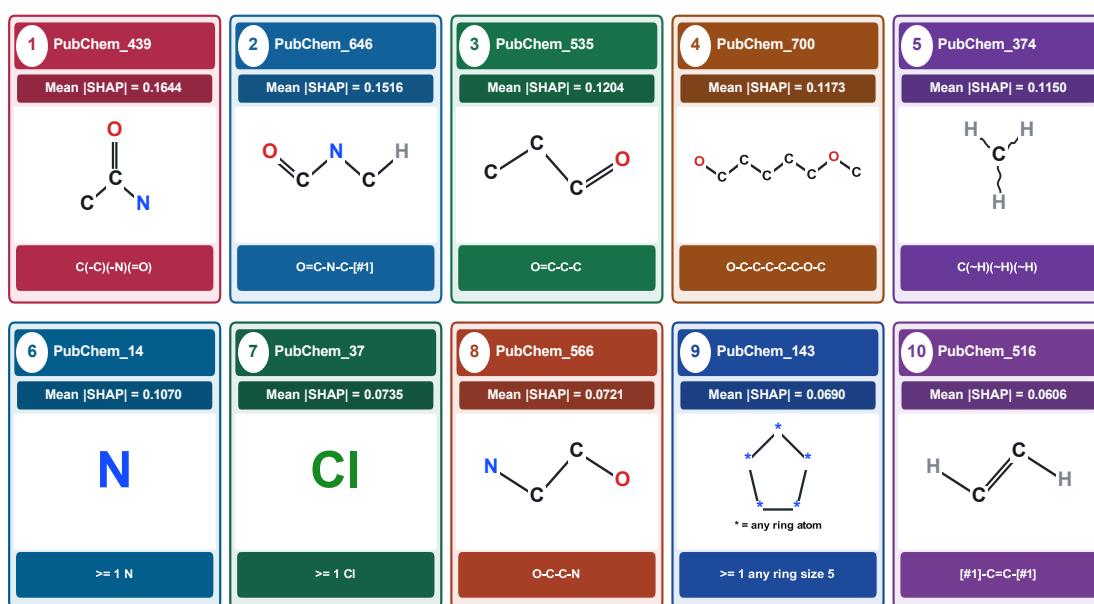


Fig. 4 Schematic representations of the ten PubChem fingerprint patterns with the highest mean absolute SHAP values in the selected LGBM model. The drawings represent fingerprint definitions rather than complete compounds or experimentally confirmed binding motifs

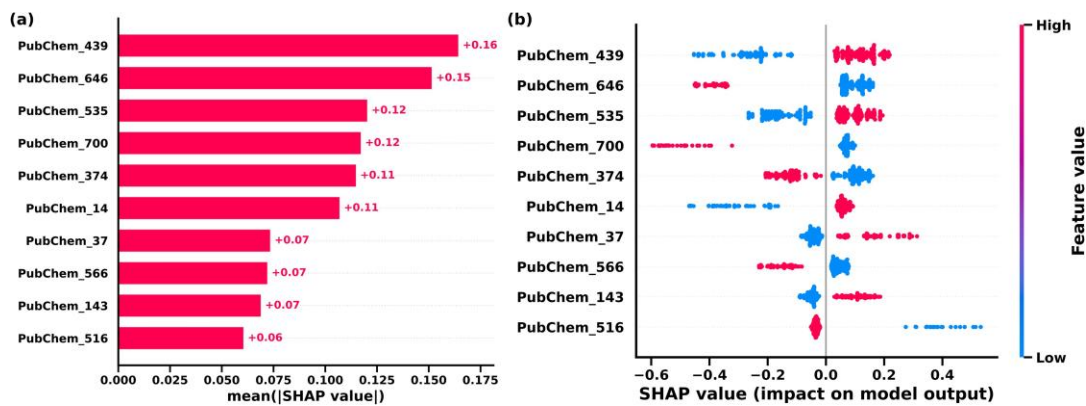


Fig. 5 SHAP interpretation of the selected LGBM model using the top ten PubChem fingerprint descriptors:

(a) mean absolute SHAP importance and (b) distribution and direction of SHAP contributions

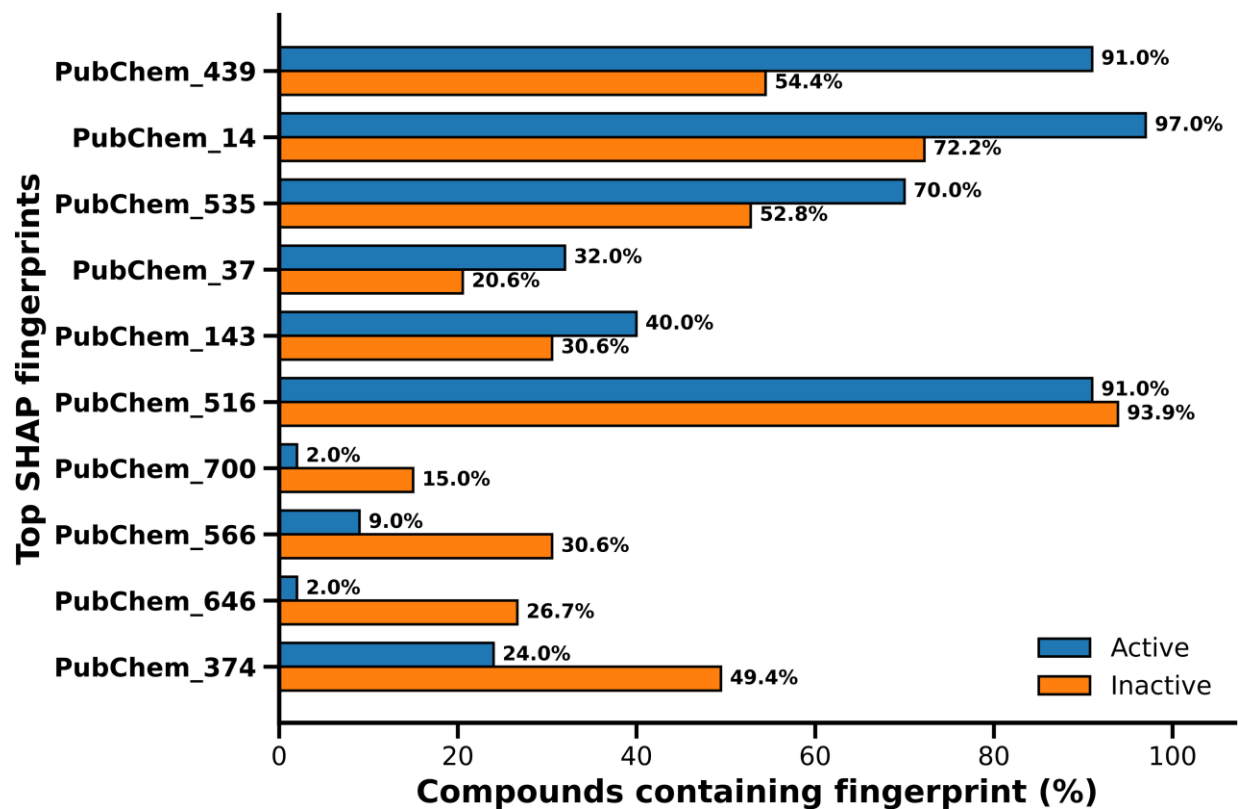


Fig. 6 Frequency of the top SHAP-ranked PubChem fingerprint descriptors in the descriptive activity groups

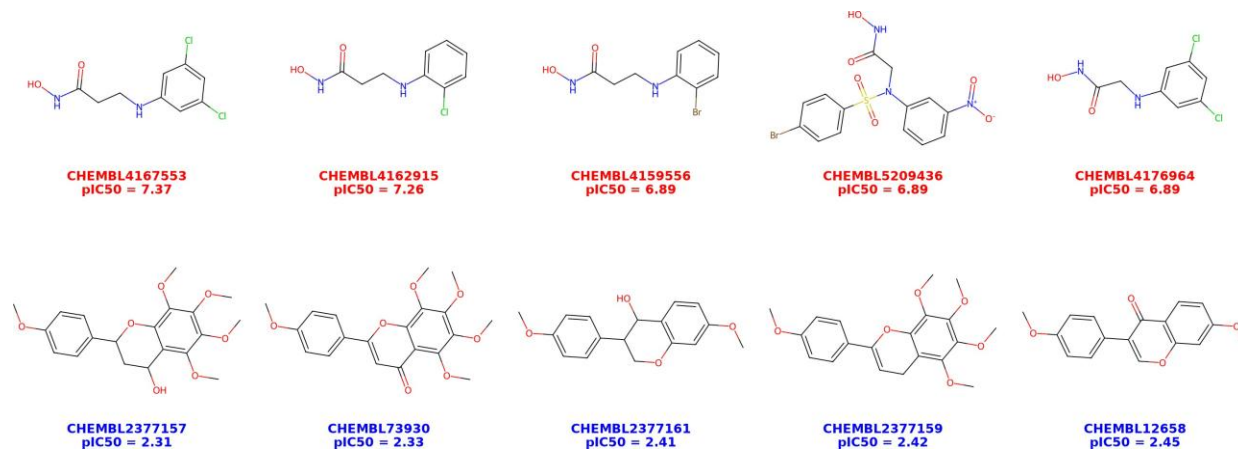


Fig. 7 Representative high- and low-potency compounds from the curated *H. pylori* urease dataset

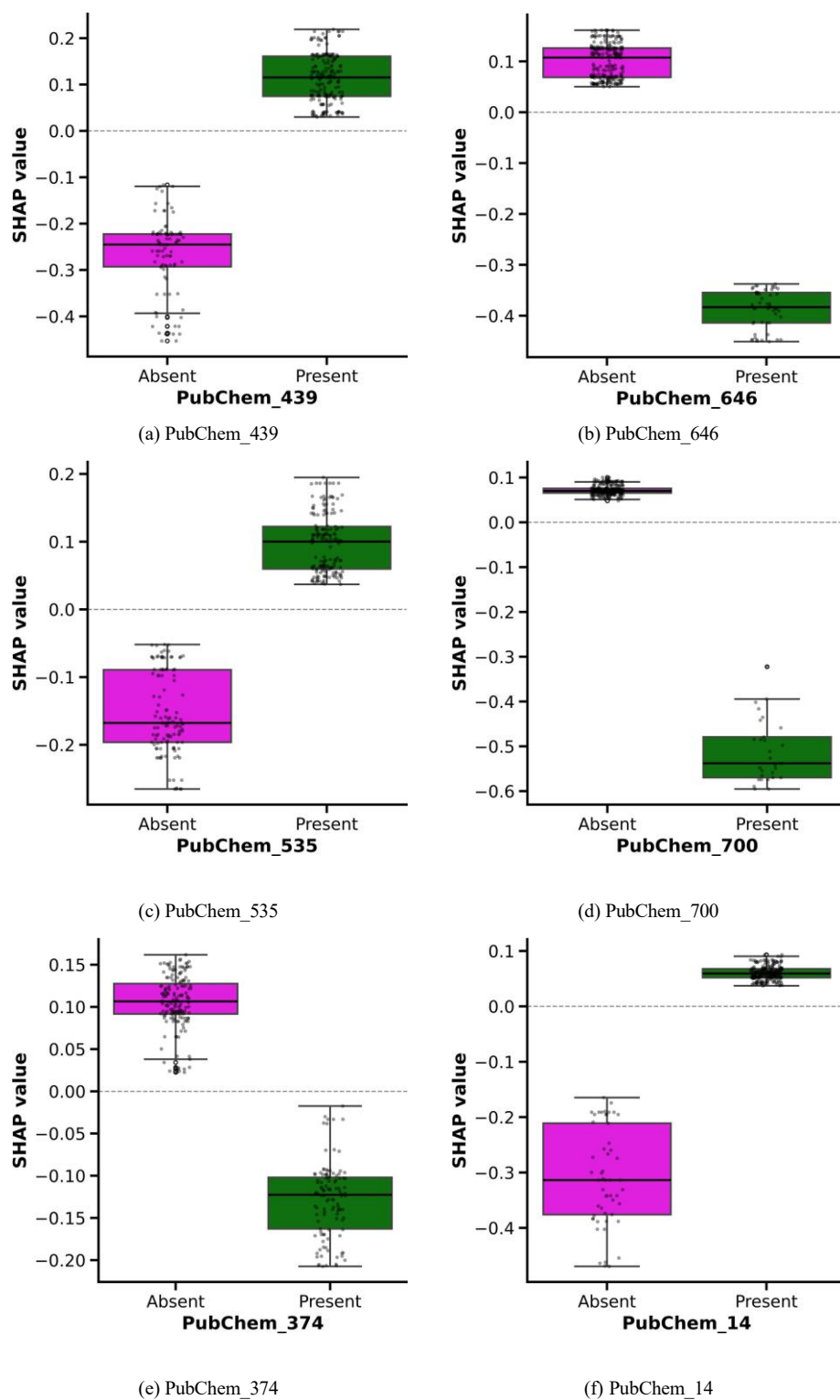


Fig. 8 Per-compound SHAP-value distributions grouped by absence or presence of six top-ranked PubChem fingerprint bits

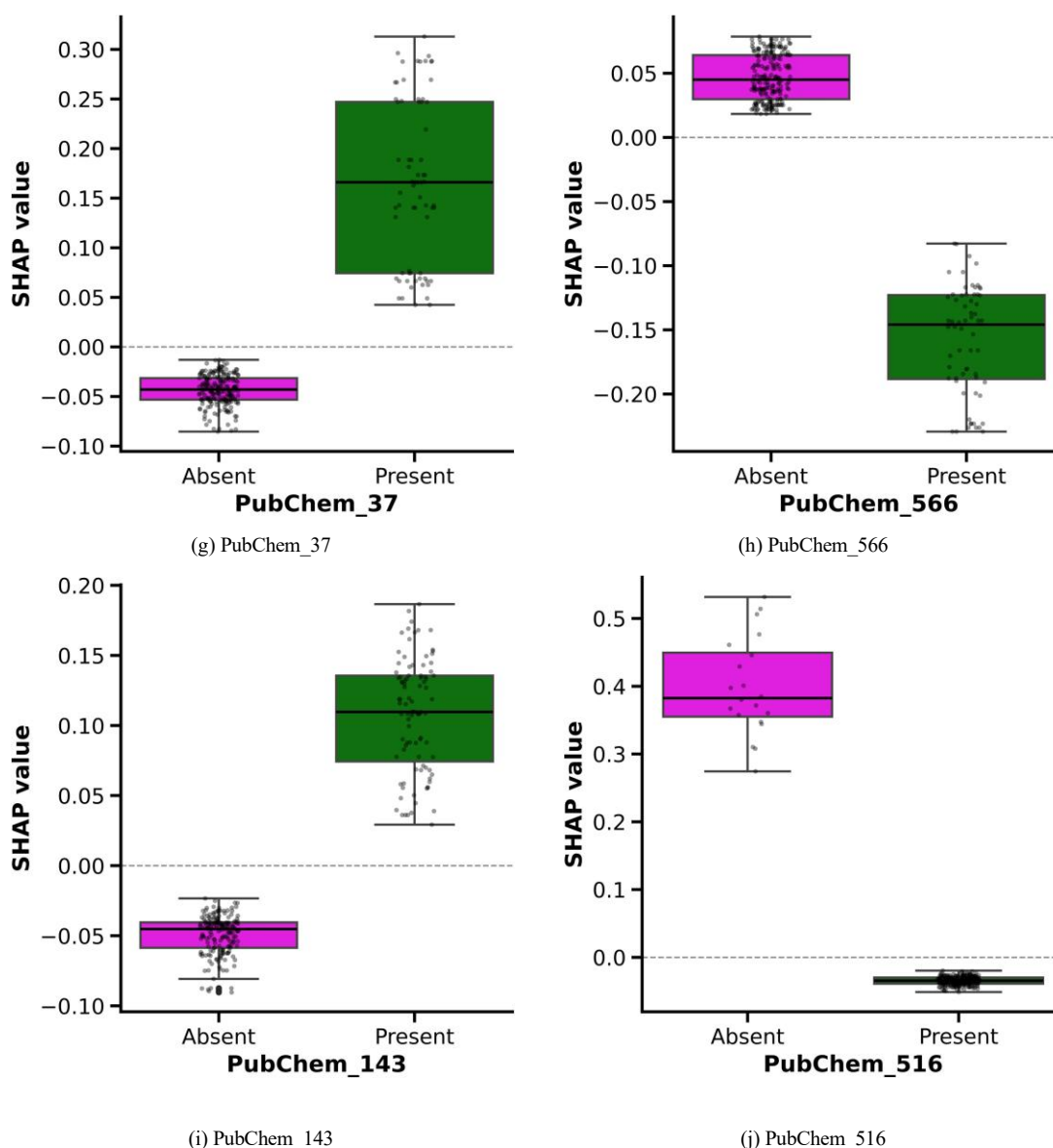


Fig. 8 SHAP-value distributions for the remaining four top-ranked PubChem fingerprint bits (continued)

3.5 Applicability Domain of the Final Model

Applicability-domain assessment was performed for the selected LGBM model using the 142 retained descriptors, with 280 training compounds and 71 held-out test compounds (Figure 9). At the primary empirical 95th-percentile leverage cutoff ($h_{p95} = 0.988$), 9 compounds in each split exceeded the leverage threshold; thus, 271/280 training compounds (96.8%) and 62/71 test compounds (87.3%) were within the leverage range. Two training compounds exceeded the positive residual warning limit of three training cross-validation RMSE units. Applying the combined leverage and residual criteria, 270/280 training compounds (96.4%) and 62/71 test compounds (87.3%) remained within the applicability domain. Sensitivity analysis showed that the empirical 99th-percentile cutoff ($h_{p99} = 1.000$) classified 280/280 training and 62/71 test compounds within the leverage range, whereas the conventional $3(p + 1)/n$ cutoff ($h^* = 1.532$) classified 280/280 training and 70/71 test compounds within that range. The empirical P95 cutoff was retained as the primary, more stringent leverage criterion.

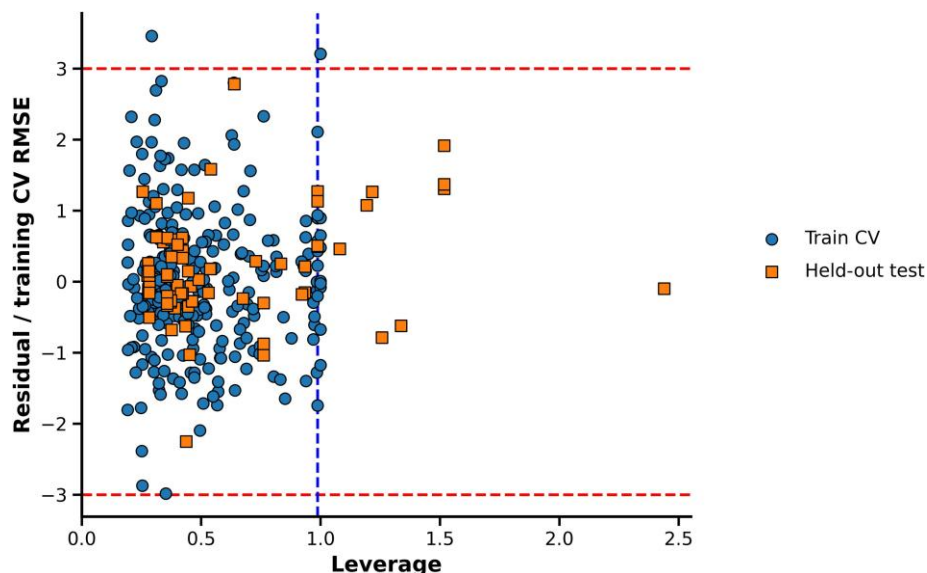


Fig. 9 Williams plot of leverage versus cross-validation-scaled residuals for the selected LGBM model. Dashed blue and red lines mark the leverage and residual warning thresholds, respectively

3.6 PanURA Web Application

The selected LGBM model was deployed as PanURA, a Streamlit web application for individual compound prediction and batch screening. The deployed application is available at <https://panura.streamlit.app>. It returns predicted pIC_{50} and corresponding IC_{50} estimates in μM , with batch results available for interactive review and .csv export. The application workflow and input-processing steps are described in Section 2.6.

4 Discussion

The present study establishes a quantitative relationship between molecular structure and *Helicobacter pylori* urease inhibitory potency using species-specific bioactivity data and PubChem molecular fingerprints. Among the evaluated regression algorithms, LightGBM (LGBM) achieved the highest five-fold cross-validated performance ($Q_{\text{CV}}^2 = 0.525$) and maintained favorable predictive accuracy on the independent test set ($R_{\text{test}}^2 = 0.782$, $\text{RMSE}_{\text{test}} = 0.493$, and $\text{MAE}_{\text{test}} = 0.353$). HistGradientBoosting (HGB) produced nearly identical results, indicating that both histogram-based gradient-boosting implementations captured the predictive relationships consistently. The combination of quantitative potency modeling and molecular fingerprint interpretation offers a means of investigating structural characteristics associated with inhibition of the pathogen-specific enzyme.

Previous machine-learning investigations have addressed urease inhibition through different biological endpoints, chemical datasets, and computational strategies. Aniceto et al. [2] developed a random-forest classifier using 2640 compounds tested against jack bean urease, integrating molecular docking and experimental evaluation of predicted hits. Morales et al. [23] similarly approached urease inhibitor identification as a classification problem, emphasizing physicochemical properties as screening criteria. More directly related to the present work, Wang et al. [39] developed an ensemble-learning model for *H. pylori* urease inhibitors, combining potency prediction with activity-cliff analysis, molecular docking, and molecular dynamics simulations to investigate hydroxamic acid derivatives. The present study contributes a complementary perspective through continuous pIC_{50} modeling of 351 curated compounds encompassing chemically diverse *H. pylori* urease inhibitors, followed by fingerprint-level interpretation and applicability-domain assessment. This approach permits potency-associated structural patterns to be examined across different molecular frameworks rather

than being restricted to a single inhibitor class. The distinction between urease sources remains biologically relevant because kinetic properties and inhibitor responses can differ between bacterial and plant-derived enzymes [5, 25]. The methodological differences among these studies are summarized in Table 5.

Table 5

Comparison of the present model with representative urease QSAR and machine-learning studies

Study	Dataset/target	Endpoint	Main method	Reported performance
ul Haq and Wadood [13]	27 hydroxamic acids; urease source not restricted to the present dataset	Urease potency	Docking and 3D-QSAR	LOO $q^2 = 0.848/0.755$; no comparable independent test R^2
Aniceto et al. (2023)	Urease-screening dataset with docking	Activity prioritization	ML and docking	Screening study; no continuous pIC_{50} test R^2
Morales et al. (2024)	Urease-inhibitor dataset	Binary classification	Supervised classifiers	Classification endpoint; no regression R^2
Wang et al. (2024)	Focused hydroxamic-acid dataset targeting <i>H. pylori</i> urease	Continuous potency	Ensemble ML, docking, and MD	$R^2 = 0.761$
Present study	351 curated <i>H. pylori</i> urease records; 280/71 train-test split	Continuous pIC_{50}	LGBM with 142 PubChem fingerprints	Test $R^2 = 0.782$; RMSE = 0.493; MAE = 0.353

The selection of LGBM was supported by complementary validation procedures that examined different aspects of model performance. Four of the five retrained regressors satisfied the implemented external-validation criteria, while response randomization substantially disrupted the predictive relationships, producing negative mean independent-test R^2 values across all five algorithms. The results also revealed differences between predictive accuracy and the corrected randomization coefficient. OrthogonalMatching-Pursuit (OMP), despite exhibiting weaker cross-validated and external performance, achieved the highest corrected randomization coefficient. Conversely, LGBM and HGB remained closely comparable across the principal predictive metrics. These findings support the predefined cross-validation-based selection of LGBM and demonstrate the importance of considering predictive accuracy and sensitivity to chance associations jointly. The observed differences between cross-validation and independent-test performance further highlight the influence of evaluation-set composition on reported QSAR performance, a consideration emphasized in established model-validation literature [29, 30, 38].

The physicochemical characteristics of the curated compounds revealed an additional dimension of the structure-activity relationships. Higher-potency compounds generally had greater molecular weights but fewer hydrogen-bond acceptors and donors, whereas lipophilicity showed no statistically significant difference between the descriptive potency groups (Figure 2). Aniceto et al. [2] similarly observed overlapping physicochemical distributions among active and inactive urease inhibitors, illustrating the limited discrimination afforded by broad molecular properties alone. Structural investigations of *H. pylori* urease provide a complementary molecular perspective. Cunha et al. [8] demonstrated that a hydroxamic acid-containing inhibitor coordinates the catalytic nickel ions while its aromatic and heterocyclic regions interact with residues surrounding the active-site flap. Such interactions illustrate how the chemical identity and spatial arrangement of functional groups can influence inhibitor recognition. The physicochemical differences observed in the present dataset may therefore reflect variation in scaffold composition and functional-group organization, particularly among inhibitor chemotypes containing distinct carbonyl, nitrogen, and oxygenated structural arrangements.

A more specific account of these structural relationships emerged from SHAP analysis of the selected LGBM model. Carbonyl-containing fragments featured prominently among the highest-ranked molecular

fingerprints, but their contributions differed according to local atomic connectivity (Figures 4 and 5). The carbonyl carbon bonded to carbon and nitrogen atoms (PubChem_439) and the carbonyl-linked carbon-chain arrangement (PubChem_535) were associated with positive contributions to predicted pIC₅₀. In contrast, the carbonyl–nitrogen–carbon–hydrogen pattern (PubChem_646) contributed predominantly toward lower predictions. The contrasting behavior of these related fragments is chemically informative because it differentiates carbonyl-containing environments that would otherwise be grouped under broad functional-group descriptions. Similar distinctions appeared among other influential fingerprints: nitrogen-containing and chlorine-associated patterns generally favored higher predicted potency, whereas extended oxygenated carbon chains and saturated methyl-associated arrangements showed predominantly negative contributions. The resulting attribution profiles emphasize the relevance of neighboring atoms and molecular connectivity to the structure–activity relationships captured by the model.

The relationship between fingerprint attribution and experimental potency became clearer when the SHAP findings were examined alongside structural enrichment and compound-level distributions (Figures 6 and 8). The carbonyl–carbon–nitrogen pattern was present in 91.0% of higher-potency compounds compared with 54.4% of lower-potency compounds, consistent with its predominantly positive contribution to model predictions. The carbonyl–nitrogen–carbon–hydrogen arrangement displayed the opposite tendency, occurring in 26.7% of lower-potency compounds but only 2.0% of higher-potency compounds. These associations suggest that differences in nitrogen substitution and neighboring carbonyl connectivity contribute to the discrimination of potency-associated chemotypes. The SHAP distributions nevertheless revealed variation in attribution magnitude among compounds sharing identical fingerprint states, demonstrating that the contribution assigned to a fragment also reflects the molecular context in which it occurs. This distinction is particularly relevant to nonlinear QSAR models, where combinations of structural descriptors can influence predicted activity beyond the information conveyed by any individual feature [14, 28].

An especially informative example was the hydrogen-substituted carbon–carbon double-bond pattern (PubChem_516). Despite ranking among the influential descriptors, it occurred at similar frequencies in the higher- and lower-potency groups (91.0% and 93.9%, respectively). Its compound-level SHAP distribution showed that the strongest positive contributions were associated with the absence of this fingerprint, whereas its presence produced comparatively small negative contributions. The importance of this descriptor therefore arose primarily from how the model treated molecules lacking a common structural pattern, rather than from enrichment of the corresponding unsaturated fragment among potent inhibitors. This behavior illustrates a distinction between structural prevalence and predictive contribution that is not apparent from activity-group frequencies alone. It also reinforces the need to interpret highly ranked fingerprints according to the direction and distribution of their SHAP values, particularly when examining chemically diverse datasets.

The representative structures in Figure 7 place these statistical associations within recognizable chemical frameworks. Highly potent examples, including CHEMBL4167553 (pIC₅₀ = 7.37) and CHEMBL4162915 (pIC₅₀ = 7.26), feature nitrogen-containing carbonyl functionalities within substituted aromatic molecular architectures. Their structural characteristics are consistent with the prominence of carbonyl–nitrogen patterns in the SHAP analysis and with the relevance of hydroxamic acid-containing inhibitors established in previous *H. pylori* urease investigations [8, 39]. By comparison, lower-potency representatives, including CHEMBL2377157 (pIC₅₀ = 2.31), display distinct oxygen-rich cyclic architectures. The contrast between these molecular frameworks is broadly consistent with the opposing contributions of selected nitrogen-containing and extended oxygenated fingerprints. However, the occurrence of some influential fragments across both potency groups reveals that broad structural features do not uniquely characterize active inhibitor classes. The representative compounds consequently support an interpretation centered on the organization

and coexistence of molecular features, providing a rationale for subsequent examination of related analogues in which individual structural modifications can be compared more directly.

The relevance of chemical-space coverage becomes particularly important when these molecular relationships are applied to the evaluation of new compounds. At the empirical 95th-percentile leverage threshold, 62 of the 71 independent test compounds (87.3%) fell within the selected model's structural applicability domain (Figure 9). The combined leverage and residual assessment retained the same number of test compounds, supporting the coverage of the evaluated chemical space while identifying structures with comparatively unusual fingerprint profiles. The choice of domain boundary also influenced structural coverage: the conventional leverage threshold included 70 test compounds, compared with 62 under the more stringent empirical criterion. The selected P95 threshold therefore provides a defined reference for interpreting prospective predictions and distinguishing candidate molecules according to their representation within the training descriptor space. Incorporating this assessment into molecular prioritization is particularly relevant when chemically modified candidates are evaluated using a model trained on a finite collection of inhibitor structures [32, 38].

The deployment of the selected model through PanURA connects these predictive and structural findings with interactive molecular exploration. The application allows existing molecular structures to be modified or new structures to be constructed while their predicted inhibitory potency is updated using the same trained LGBM model. Successive designs can be retained and compared, enabling researchers to examine how molecular modifications alter predicted pIC_{50} and estimated IC_{50} within a consistent biological endpoint. Integrating applicability-domain assessment into this workflow adds an important dimension to compound evaluation. A structural modification may improve predicted potency while simultaneously moving the resulting molecule outside the chemical space represented during model development. Assessing both outcomes enables candidate selection to consider predicted activity alongside the model's structural coverage rather than relying exclusively on numerical potency estimates. PanURA thus supports iterative, model-guided exploration of *H. pylori* urease inhibitor candidates and provides a practical connection between interpretable QSAR analysis and prospective molecular design.

The scope of these findings is influenced by the chemical composition of the curated dataset, potential differences among experimental assay conditions, and the validation strategy employed. Although Kennard–Stone partitioning provided systematic coverage of descriptor space, it did not establish scaffold-independent predictive performance. The species-specific endpoint also does not demonstrate inhibitory selectivity against ureases from other biological sources. Furthermore, PubChem fingerprints encode two-dimensional structural characteristics without explicitly representing molecular conformations, enzyme–inhibitor binding geometries, or cellular activity. The associations identified through SHAP describe relationships learned by the predictive model and do not independently establish biochemical mechanisms or the effects of individual substituent modifications. Likewise, favorable potency predictions generated during interactive molecular editing remain computational hypotheses. Future studies incorporating scaffold-aware validation, matched molecular-pair investigations, prospective biochemical assays, and complementary structure-based analyses could extend the framework's predictive scope and establish the experimental relevance of the molecular relationships identified here.

Acknowledgements

The author gratefully acknowledges Mkwawa University College of Education, University of Dar es Salaam for providing a supportive academic environment during the conduct of this research.

Statements and Declarations

Funding

No external funding was received for this study.

Competing Interests

The author declares no competing interests.

Author Contributions

A.A.P. conceived and designed the study, curated the data, developed the methodology and software, conducted the formal analyses, interpreted the results, and wrote and revised the manuscript. The author read and approved the final manuscript.

Ethics Approval

Not applicable. This study analyzed chemical structures and previously reported in vitro bioactivity data and did not involve new studies of human participants or animals.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Data Availability

The data supporting the findings of this study are provided in the manuscript.

Code Availability

The PanURA source code and software dependencies are available at <https://github.com/peterDataScientia/panURA>.

References

- [1] Adamczyk J, Ludynia P (2024) Scikit-fingerprints: Easy and efficient computation of molecular fingerprints in python. *SoftwareX* 28:101944. <https://doi.org/10.1016/j.softx.2024.101944>
- [2] Aniceto N, Albuquerque TS, Bonifacio VDB, et al (2023) Using machine learning and molecular docking to leverage urease inhibition data for virtual screening. *International Journal of Molecular Sciences* 24(9):8180. <https://doi.org/10.3390/ijms24098180>

- [3] Cao W, Jin F, Liu B, et al (2025) Synthesis and evaluation of novel ebselen derivatives as urease inhibitors for combating *Helicobacter pylori* infections. European Journal of Medicinal Chemistry 287:117359. <https://doi.org/10.1016/j.ejmech.2025.117359>
- [4] Cardos IA, Danila C, Yahya G, et al (2026) Targeting *Helicobacter pylori* enzymes using *Viscum album* l. extract: In silico molecular docking and in vitro study. Frontiers in Cellular and Infection Microbiology 15:1690969. <https://doi.org/10.3389/fcimb.2025.1690969>
- [5] Cesareo SD, Langton SR (1992) Kinetic properties of *Helicobacter pylori* urease compared with jack bean urease. FEMS Microbiology Letters 99(1):15–21. <https://doi.org/10.1111/j.1574-6968.1992.tb05535.x>
- [6] Chen X, Gou L, Wang Y, et al (2025) New directions in *Helicobacter pylori* urease inhibitors: Focusing on nickel ions transfer and auxiliary protein interactions during urease maturation. Infection and Drug Resistance 18:3037–3053. <https://doi.org/10.2147/IDR.S519194>
- [7] Crowe SE (2019) *Helicobacter pylori* infection. New England Journal of Medicine 380(12):1158–1165. <https://doi.org/10.1056/NEJMc1710945>
- [8] Cunha ES, Chen X, Sanz-Gaitero M, et al (2021) Cryo-em structure of *Helicobacter pylori* urease with an inhibitor in the active site at 2.0 Å resolution. Nature Communications 12:230. <https://doi.org/10.1038/s41467-020-20485-6>
- [9] Golbraikh A, Tropsha A (2002) Beware of q^2 ! journal of molecular graphics and modelling, 20(4), 269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
- [10] Grabarek M, Tabor W, Krzyzek P, et al (2025) Synthesis of phosphonate derivatives of benzoselenazolones and their remarkable antiureolytic activity in *Helicobacter pylori* cells. Journal of Medicinal Chemistry 68(19):20324–20339. <https://doi.org/10.1021/acs.jmedchem.5c01385>
- [11] Guzel-Akdemir O, Akdemir A (2025) Urease inhibitors for the treatment of *H. pylori*. Expert Opinion on Therapeutic Patents 35(1):17–30. <https://doi.org/10.1080/13543776.2024.2423004>
- [12] Ha NC, Oh ST, Sung JY, et al (2001) Supramolecular assembly and acid resistance of *Helicobacter pylori* urease. Nature Structural Biology 8(6):505–509. <https://doi.org/10.1038/88563>
- [13] ul Haq Z, Wadood A (2009) Prediction of binding affinities for hydroxamic acid derivatives as urease inhibitors by molecular docking and 3d-qsar studies. Letters in Drug Design & Discovery 6(2):93–100. <https://doi.org/10.2174/157018009787582598>
- [14] Jimenez-Luna J, Grisoni F, Schneider G (2020) Drug discovery with explainable artificial intelligence. Nature Machine Intelligence 2(10):573–584. <https://doi.org/10.1038/s42256-020-00236-4>
- [15] Kafarski P, Talma M (2018) Recent advances in design of new urease inhibitors: A review. Journal of Advanced Research 13:101–112. <https://doi.org/10.1016/j.jare.2018.01.007>

- [16] Khoder G, Mustafa EM, Zaib S, et al (2026) Novel indole-based sulfonate and sulfamate derivatives as potent and selective urease inhibitors targeting *Helicobacter pylori*: A promising therapeutic strategy. *Gut Pathogens* 18:27. <https://doi.org/10.1186/s13099-026-00814-8>
- [17] Kim S, Chen J, Cheng T, et al (2023) Pubchem 2023 update. *Nucleic Acids Research* 51(D1):D1373–D1380. <https://doi.org/10.1093/nar/gkac956>
- [18] Landrum G, Tosco P, Kelley B, et al (2026) rdkit/rdkit: 2025_09_6 (q3 2025) release. Computer software, <https://doi.org/10.5281/zenodo.18797641>, URL <https://doi.org/10.5281/zenodo.18797641>
- [19] Li Y, Choi H, Leung K, et al (2023) Global prevalence of *Helicobacter pylori* infection between 1980 and 2022: A systematic review and meta-analysis. *The Lancet Gastroenterology & Hepatology* 8(6):553–564. [https://doi.org/10.1016/S2468-1253\(23\)00070-5](https://doi.org/10.1016/S2468-1253(23)00070-5)
- [20] Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems* 30, pp 4765–4774
- [21] Lundberg SM, Erion G, Chen H, et al (2020) From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence* 2(1):56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- [22] Malfertheiner P, Megraud F, Rokkas T, et al (2022) Management of *Helicobacter pylori* infection: The maastricht vi/florence consensus report. *Gut* 71(9):1724–1762. <https://doi.org/10.1136/gutjnl-2022-327745>
- [23] Morales N, Valdes-Munoz E, Gonzalez J, et al (2024) Machine learning-driven classification of urease inhibitors leveraging physicochemical properties as effective filter criteria. *International Journal of Molecular Sciences* 25(8):4303. <https://doi.org/10.3390/ijms25084303>
- [24] Muratov EN, Bajorath J, Sheridan RP, et al (2020) Qsar without borders. *Chemical Society Reviews* 49(11):3525–3564. <https://doi.org/10.1039/D0CS00098A>
- [25] Odake S, Morikawa T, Tsuchiya M, et al (1994) Inhibition of *Helicobacter pylori* urease activity by hydroxamic acid derivatives. *Biological and Pharmaceutical Bulletin* 17(10):1329–1332. <https://doi.org/10.1248/bpb.17.1329>
- [26] OECD (2024) (q)sar assessment framework: Guidance for the regulatory assessment of (quantitative) structure activity relationship models and predictions. Second edition, OECD, URL https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/11/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions-second-edition_cc89955e/bbdac345-en.pdf

- [27] Ojha PK, Roy K (2011) Comparative qsars for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection. *Chemometrics and Intelligent Laboratory Systems* 109(2):146–161. <https://doi.org/10.1016/j.chemolab.2011.08.007>
- [28] Rodriguez-Perez R, Bajorath J (2020) Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design* 34:1013–1026. <https://doi.org/10.1007/s10822-020-00314-0>
- [29] Roy PP, Paul S, Mitra I, et al (2009) On two novel parameters for validation of predictive qsar models. *Molecules* 14(5):1660–1701. <https://doi.org/10.3390/molecules14051660>
- [30] Rucker C, Rucker G, Meringer M (2007) Y-randomization and its variants in qspr/qsar. *Journal of Chemical Information and Modeling* 47(6):2345–2357. <https://doi.org/10.1021/ci700157b>
- [31] Sachs G, Weeks DL, Melchers K, et al (2003) The gastric biology of *Helicobacter pylori*. *Annual Review of Physiology* 65:349–369. <https://doi.org/10.1146/annurev.physiol.65.092101.142156>
- [32] Sahigara F, Mansouri K, Ballabio D, et al (2013) Comparison of different approaches to define the applicability domain of qsar models. *Molecules* 18(5):4791–4810. <https://doi.org/10.3390/molecules18054791>
- [33] Savoldi A, Carrara E, Graham DY, et al (2018) Prevalence of antibiotic resistance in *Helicobacter pylori*: A systematic review and meta-analysis in world health organization regions. *Gastroenterology* 155(5):1372–1382.e17. <https://doi.org/10.1053/j.gastro.2018.07.007>
- [34] Shapley LS (1953) A value for n-person games. In: *Contributions to the Theory of Games II*. Princeton University Press, p 307–317
- [35] Song Q, Wu H, Ma Z, et al (2026) The functional and catalytic landscape of urease reveals a conserved target against *Helicobacter pylori*. *Gut Microbes* 18(1):2653575. <https://doi.org/10.1080/19490976.2026.2653575>
- [36] Streamlit (2026) Streamlit documentation. <https://docs.streamlit.io/>, accessed: 13 October 2026
- [37] Tang Y, Yang F, Wen X, et al (2025) Component characterization of *Smilax glabra* roxb. and its inhibitory activity against *Helicobacter pylori* through targeted suppression of its secreted urease. *Frontiers in Cellular and Infection Microbiology* 15:1617330. <https://doi.org/10.3389/fcimb.2025.1617330>
- [38] Tropsha A (2010) Best practices for qsar model development, validation, and exploitation. *Molecular Informatics* 29(6–7):476–488. <https://doi.org/10.1002/minf.201000061>

- [39] Wang N, Wu X, Liang J, et al (2024) Molecular design of hydroxamic acid-based derivatives as urease inhibitors of *Helicobacter pylori* using ensemble learning model, molecular docking and molecular dynamics simulations. *Molecular Diversity* 28:2229–2244. <https://doi.org/10.1007/s11030-024-10914-9>
- [40] Wang Y, Zheng S, Guo R, et al (2025) Assessment for antibiotic resistance in *Helicobacter pylori*: A practical and interpretable machine learning model based on genome-wide genetic variation. *Virulence* 16(1):2481503. <https://doi.org/10.1080/21505594.2025.2481503>
- [41] Zdrazil B, Felix E, Hunter F, et al (2024) The chembl database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research* 52(D1):D1180–D1192. <https://doi.org/10.1093/nar/gkad1004>
- [42] Zhao L, Chen H, Zheng D, et al (2026) Oleuropein derived from olive leaves inhibits the activity of *Helicobacter pylori* urease: Network pharmacology, inhibition kinetics, and mechanistic insights. *RSC Advances* 16(26):23637–23650. <https://doi.org/10.1039/D6RA00410E>