Trop J Nat Prod Res, August 2025; 9(8): 3784 - 3793 **ISSN 2616-0684 (Print)**
ISSN 2616-0692 (Electronic)

Tropical Journal of Natural Product Research

Available online at https://www.tjnpr.org

Original Research Article

Enhancing Quantitative Structure–Activity Relationship Predictive Power and Explainability: Meta-Modeling and Shapley Additive Explanations Feature Importance Analysis for Drug Discovery

Ardo Sanjaya [1,2]*, Hana Ratnawati[2,3], Nathanael A. Mianto[1], Keyshia V. Camillo[4], Aryo Tedjo[5,6], Wisnu A. Kusuma [7,8]

[1]*Department of Anatomy, Faculty of Medicine, Maranatha Christian University, Bandung, 40164, Indonesia*
[2]*Biomedical Research Laboratory, Faculty of Medicine, Maranatha Christian University, Bandung, 40164, Indonesia*
[3]*Department of Histology, Faculty of Medicine, Maranatha Christian University, Bandung, 40164, Indonesia*
[4]*Undergraduate Program in Medicine, Faculty of Medicine, Maranatha Christian University, Bandung, 40164, Indonesia*
[5]*Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia*
[6]*Bioinformatics Core Facilities - IMERI, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia*
[7]*Bioinformatics Study Program, Faculty of Mathematics and Natural Science, IPB University, Bogor, 16680, Indonesia*
[8]*Tropical Biopharmaca Research Center, Institute of Research and Community Empowerment, IPB University, Bogor, 16128, Indonesia*

ARTICLE INFO

Article history:
Received 07 June 2025
Revised 07 July 2025
Accepted 09 July 2025
Published online 01 September 2025

ABSTRACT

Quantitative structure-activity relationship (QSAR) modeling plays a crucial role in drug discovery by predicting biological endpoints based on molecular structure. Existing studies lack consensus on the optimal fingerprint, and many function as black boxes with limited explainability. This study addresses these gaps by integrating multiple fingerprints through meta-modeling and applying SHAP analysis to enhance prediction accuracy and interpretability. The performance of several fingerprints was evaluated across ten proteins to predict pIC50. Concordance analysis using the Concordance Correlation Coefficient (CCC) was used to evaluate prediction agreement and reproducibility. Shapley Additive exPlanations (SHAP) analysis was used to analyze the feature importance of the molecular substructure in the base model and fingerprint importance in the meta-model. A streamlit web application was developed to demonstrate prediction and feature importance visualization. No fingerprints showed superiority over the others. Concordance analysis showed high agreement with CCC values above 0.98, reflecting high prediction reproducibility. Meta models combining Morgan6 and other fingerprints outperformed individual models in 7 protein targets. SHAP analysis revealed that fingerprint importance is context-dependent on the target proteins. The web application demonstrated the importance of identifying critical substructures using a case study in Donepezil. Although fingerprints modeled different aspects of the molecule, they have similar performance. The fingerprints showed high predictive reproducibility and agreement. This study demonstrates that while individual molecular fingerprints offer comparable predictive performance, integrating them through meta-modeling enhances prediction accuracy and interpretability. Combined with SHAP-based fingerprint importance analysis, the study provides a reproducible and explainable QSAR framework that advances data-driven drug discovery.

Keywords: Drug discovery, Machine learning, Molecular fingerprinting, Quantitative structure-activity relationship (QSAR).

Introduction

Ligand-based drug design methods, including quantitative structure–activity relationship (QSAR) modeling, have been recognized as fundamental components of virtual screening strategies in drug discovery, enabling the efficient evaluation of large molecular libraries. This approach was first introduced 50 years ago by Hansch and Fujita.[1]

***Corresponding author. E mail: ardo.sanjaya@med.maranatha.edu
Tel: +62 859-1066-09851

Citation: Sanjaya A, Ratnawati H, Mianto NA, Camillo KV, Tedjo A, Kusuma WA. Enhancing Quantitative Structure–Activity Relationship Predictive Power and Explainability: Meta-Modeling and Shapley Additive Explanations Feature Importance Analysis for Drug Discovery. Trop J Nat Prod Res. 2025; 9(8): 3784 – 3793 https://doi.org/10.26538/tjnpr/v9i8.38

Official Journal of Natural Product Research Group, Faculty of Pharmacy, University of Benin, Benin City, Nigeria

At its core, QSAR modeling is employed to generate mathematical models capable of predicting biological and toxicological properties, such as IC_{50}, EC_{50}, and various toxicity endpoints, through the application of regression and classification techniques.[2,3] These mathematical models utilize molecular descriptors and fingerprints to characterize chemical structures.[4–6] Such descriptors and fingerprints serve as inputs for various algorithms designed to correlate structural information with biological properties. Consequently, the predictive performance of these models is partly dependent on the molecular descriptors or fingerprints used as inputs, as well as the quality of the underlying data.[7] Numerous fingerprint types are currently available, each representing a distinct set of structural information for a given molecule.[7,8] Although QSAR modeling has become an indispensable tool for screening extensive molecular libraries, the selection of molecular fingerprints plays a pivotal role in influencing the predictive accuracy of the model. In recent decades, QSAR modeling has progressed from simple linear regression to sophisticated machine-learning-based approaches, which offer the advantage of handling large datasets.[9–12] These methodological advances have facilitated the development of more precise models but have also introduced challenges, such as reproducibility and input selection. Various

molecular fingerprints have been developed to capture unique structural and chemical characteristics. These fingerprints are generally classified into structural keys, such as MACCS[13] and PubChem fingerprints,[14] and the so-called true fingerprints, including Atom Pairs[15] and Circular Fingerprints.[16] Structural keys typically rely on predefined patterns and are highly context-specific, often depending on the databases from which they are derived. Conversely, true fingerprints can provide comprehensive representations of molecular substructures.[8,17,18]

Despite the diversity of available fingerprints, a comprehensive comparison has yet to be conducted to evaluate their performance in predicting potency parameters, such as IC_{50}. Studies have predominantly focused on classification tasks within pharmacokinetics and have shown limited generalizability to QSAR and drug potency prediction.[19–22] A recent study by Orosz et al.[18] compared several fingerprints and descriptors for ADME-Tox predictions. However, this study did not assess the ability of these descriptors or fingerprints to predict potency. Furthermore, the reproducibility of machine-learning-based QSAR models remains a concern, as it is unclear whether consistent evaluation metrics translate into comparable predictive outcomes. This uncertainty presents a challenge when selecting the most appropriate fingerprint for a specific application.

Given that these fingerprints represent distinct structural representations of molecules, combining them may offer a more comprehensive molecular representation, thereby enhancing predictive performance. This rationale is supported by previous research that has improved traditional QSAR models using transfer or ensemble learning.[10,17] Transfer learning involves the use of models derived from large, publicly available datasets and the subsequent application—or "transfer"—of learned knowledge to a new, typically smaller dataset. Cai et al. (2020) investigated this approach and emphasized its utility for datasets characterized by limited sample sizes.[23] Another strategy, ensemble learning, integrates predictions from multiple base models, each incorporating distinct fingerprint representations to yield a final prediction. This technique has demonstrated the potential to improve predictive accuracy across diverse real-world scenarios. Kwon et al. applied ensemble learning to enhance the classification performance on several real-world toxicological datasets.[24–26] Their findings indicated that combining machine-learning models led to significant improvements over individual base models, although statistical validation was not conducted to confirm these differences.

The predictive performance of various fingerprints in QSAR modeling, using IC_{50} as the primary endpoint, was systematically assessed to address these research gaps. Multiple machine-learning algorithms were implemented, and a concordance analysis was conducted to evaluate the level of agreement among the models derived from different fingerprints, thereby assessing fingerprint reproducibility. In addition, models integrating predictions from multiple fingerprints were examined for their potential to improve predictive performance.

The contribution of each fingerprint is quantified using feature importance analysis, providing a comprehensive evaluation of fingerprint effectiveness in the context of drug discovery. It is anticipated that the findings of this study will support the identification of the most suitable fingerprints for QSAR modeling, facilitate the advancement of virtual screening strategies, and enhance the overall drug discovery workflow.

Materials and Methods

Biological activity data were obtained from the ChEMBL database (https://www.ebi.ac.uk/chembl/), focusing on proteins selected to represent various protein classes. These proteins were selected as representative potential drug targets in the human body. The list of proteins is shown in Table 1.

Table 1: Protein Targets used in this research.

No	ChEMBL ID	Protein Name	Classification
1	CHEMBL262	Glycogen synthase kinase-3 beta	Protein Kinase
2	CHEMBL2971	Tyrosine-protein kinase JAK2	Protein Kinase
3	CHEMBL220	Acetylcholinesterase	Enzymes - Hydrolase
4	CHEMBL230	Cyclooxygenase-2	Enzymes - Oxidoreductase
5	CHEMBL217	Dopamine D2 receptor	Family A G Protein-coupled Receptor
6	CHEMBL210	Beta-2 adrenergic receptor	Family A G Protein-coupled Receptor
7	CHEMBL228	Serotonin transporter	Transporter
8	CHEMBL3884	Sodium/glucose cotransporter 2	Transporter
9	CHEMBL233	Mu opioid receptor	Family A G Protein-coupled Receptor
10	CHEMBL325	Histone deacetylase 1	Eraser

Fingerprint Generation

Fingerprints were generated using the Python package pyFingerprint version 2.0[27] and RDKit version 2023.03.2.[28] The fingerprints produced included Atom Pairs,[15] Topological Torsions,[29] ECFP/Morgan,[16] MACCS keys,[13] and PubChem fingerprints.[14] The RDKit package was employed to generate the Atom Pair[15] and Topological Torsion[29] fingerprints using default settings with a bit length of 2048 bits. ECFP fingerprints were also generated using RDKit by applying a radius of two for ECFP4 and three for ECFP6, both with a bit length of 2048 bits. The pyFingerprint package [30] was used to generate MACCS keys with a bit length of 166 bits and PubChem fingerprints with a bit length of 881 bits. https://doi.org/10.3390/toxics11090785

Data Preprocessing and Cleaning

Data retrieval was conducted using the Python ChEMBL WebResource Client version 0.10.8. Data entries lacking canonical SMILES or IC_{50} values were excluded from the analysis. Duplicate entries were merged by inferring the IC_{50} value from the median. IC_{50} values were log-transformed to pIC_{50} prior to model training and prediction. The coefficient of determination (R^2) was used as the primary evaluation metric in all models.

The dataset was split at an 80:20 ratio to train the base and meta-models. An additional split was applied for training and validation using a 75:25 ratio for both the base and meta-models. These preprocessing steps were similar to the study by Khan et al. [31] and were designed to ensure the absence of data leakage during model development.

Base Model Creation
Machine-learning models were created using the Python library scikit-learn version 1.2.2.[32] Three algorithms were selected as base models: Random Forest,[33] Support Vector Machine (SVM),[34] and Extreme Gradient Boosting.[35] These models were chosen for their consistency and high performance, each representing a distinct algorithmic paradigm in machine-learning-based prediction. [36]
Hyperparameter optimization was performed using the Python package Optuna version 3.3.0,[37] employing the mean squared error as the objective function, with a maximum of 200 trials per model.

Concordance Analysis
The Concordance Correlation Coefficient (CCC) was employed as a statistical measure to evaluate the agreement between predicted values derived from different fingerprints. [38] This metric was selected because of its ability to assess both the accuracy and degree of concordance between the outputs of the two models. A custom function was implemented to compute the CCC using the following formula:

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}$$

Where:
ρ_c = CCC
ρ = Pearson correlation coefficient between two sets of predicted values
σ_x = Standard deviation of the predicted values from the first model
σ_y = Standard deviation of the predicted values from the second model
μ_x = Mean of the predicted values from the first model
μ_y = Mean of the predicted values from the second model

The CCC values were computed for each fingerprint combination pair. These values were visualized as heat maps using the Python package Seaborn (version 0.13.2) to enable intuitive comparisons across different fingerprint combinations. CCC analyses were conducted separately for each of the three machine-learning algorithms, resulting in three sets of CCC results.

Meta-Model Creation
Meta-models were constructed using three algorithms: SVM, Ridge Regression, and Extreme Gradient Boosting. For the meta-model development, the prediction outputs from the Random Forest-based models were utilized as input features. The Random Forest was selected because of its consistently superior performance across all base models, achieving the highest accuracy metrics. [31]
The fingerprints used in the meta-modeling were paired with ECFP6, a fingerprint commonly utilized in QSAR applications because of its robust performance and widespread adoption. [31] Meta-models were trained on unseen data from the hold-out set using a training–validation split of 80:20.
The Python package Optuna was employed for hyperparameter optimization, with a maximum of 50 trials per meta-model. The mean squared error served as the objective function during tuning. The R^2 was used as the evaluation metric for assessing meta-model performance. [38]

Feature Importance Analysis
To evaluate the contributions of different fingerprints to the predictive performance of the meta-models, a Shapley Additive Explanations (SHAP) analysis was conducted using the Python package SHAP (version 0.42.1). This method was selected because of its effectiveness in interpreting the outputs of machine-learning models, [31,36] particularly when multiple models are used, as in the meta-modeling process.
The SHAP values quantify the contribution of each fingerprint to the predictions, where larger absolute values (positive or negative) indicate a greater influence on the model output. To ensure a model-agnostic and comprehensive assessment of fingerprint relevance, the absolute SHAP values were aggregated across different samples and ChEMBL IDs for each fingerprint type.

Statistical Testing and Data Visualization
Statistical analysis was conducted using the Python packages SciPy version 1.10.1[39] and Statsmodels version 0.14.0[40] to compute p-values for multiple statistical comparisons. The Friedman test, a non-parametric alternative to repeated measures ANOVA, was applied to compare fingerprint performance in the base models. Where significant differences were observed, Nemenyi post hoc tests were conducted for further analysis. Additionally, paired t-tests and Wilcoxon signed-rank tests, adjusted using the Benjamini–Hochberg correction for multiple testing, were used to assess performance differences between the base and meta-models. Friedman's test was also applied to the SHAP results, including both aggregated SHAP values and individual fingerprint SHAP values for each tested protein, to assess statistical significance. Data visualization was performed using the Python packages matplotlib (version 3.7.1)[41] and seaborn version 0.12.2.[42]

Comparison of Performance Metrics for Individual Proteins
The R^2 values were analyzed by aggregating results according to specific proteins or fingerprints. [6] For each combination of fingerprint and machine-learning algorithm, 10 R^2 values were calculated for both the base and meta-models, reflecting QSAR performance per protein. These R^2 values were used to compare each protein's performance between its base model (using a single fingerprint) and its meta-model counterpart (utilizing multiple fingerprints).

Comparison of Performance Metrics for Individual Fingerprints
An additional set of R^2 values was compiled by aggregating data based on fingerprints across all tested proteins. This included R^2 metrics from both the base models and meta-models trained using each fingerprint. Five R^2 metrics—one for each fingerprint—were generated, enabling comparisons of single-fingerprint performance across all proteins with those derived from multi-fingerprint meta-models. [6]
As all meta-models were constructed using ECFP6, no direct comparison was made between the performance of this fingerprint in base modeling and its role in meta-modeling.

Web-App Development
As a proof of concept, a web application was developed using the Streamlit package (version 1.39.0) to predict pIC_{50} values by integrating two distinct molecular fingerprints: Morgan6 and PubChem. These fingerprints represent complementary encoding strategies and were selected to provide diverse structural information.
The application accepts SMILES strings as input, which are converted into the two selected fingerprints before being processed by the pretrained Random Forest models. The resulting predictions are then combined using a meta-model to improve accuracy.
In addition to the prediction outputs, the application includes the results of a feature importance analysis conducted using SHAP. The top 10 most influential features or bits from each base model and fingerprint type were calculated. To enhance interpretability, the important features of the tested compounds were directly visualized using 2D molecular structures (for Morgan6) or substructure definitions (PubChem).
This web application highlights the advantages of combining different molecular descriptors and emphasizes the role of model explainability in QSAR modeling. This study provides valuable insights into drug development and screening workflows. The application is publicly accessible on the Hugging Face platform (https://huggingface.co/spaces/faith8/meta-qsar).

Results and Discussion

Comparing Predictive Performance Across Different Fingerprints
This study systematically assessed the predictive validity of various molecular fingerprints for QSAR modeling using IC_{50} values. A total of 180 base models were constructed using 10 proteins, six fingerprints, and three machine-learning algorithms (Figure 1). The resulting R^2 values were averaged to facilitate comparisons across the different algorithms.

ISSN 2616-0684 (Print)
ISSN 2616-0692 (Electronic)

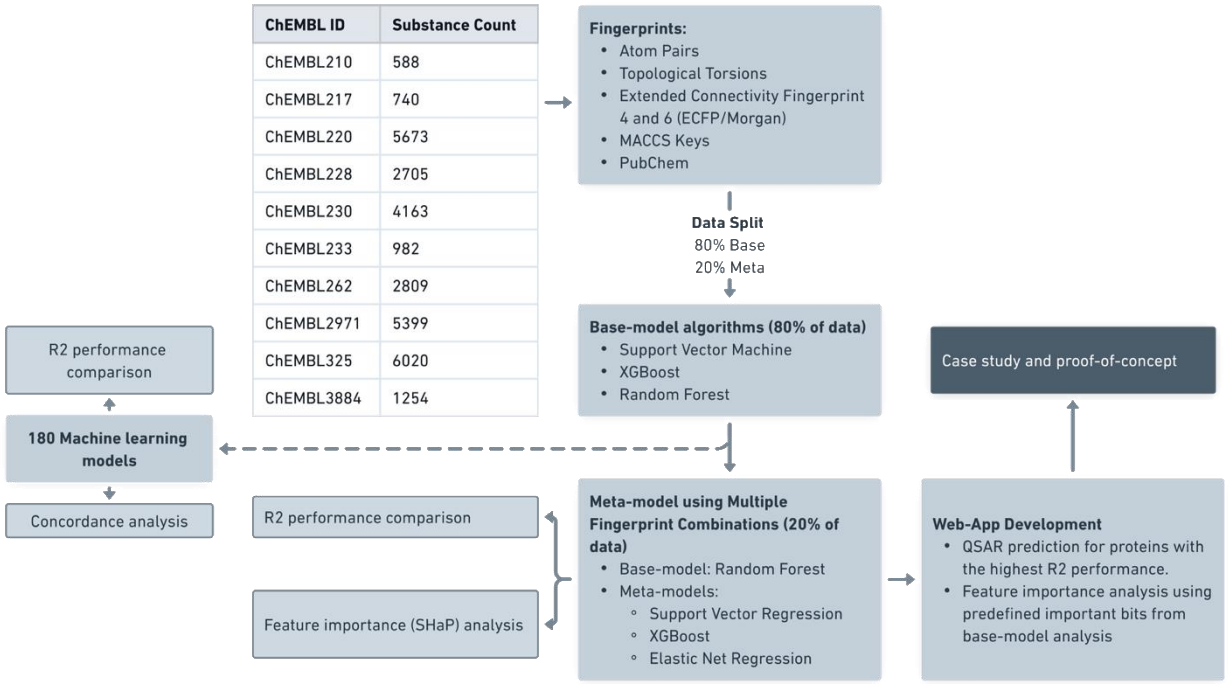


Figure 1: Overview of the QSAR modeling pipeline used in this study.

Specifically, 180 models were generated using a combination of 10 proteins, six fingerprints, and three algorithms. The R^2 values were averaged across the algorithms to assess fingerprint performance, producing R^2 metrics that were both protein- and model-agnostic (Figure 2A). Visual inspection revealed that the metrics of the tested fingerprints were similar. However, statistical analysis using the Friedman test indicated a significant difference in performance between at least one pair of fingerprints (Figure 2B). Follow-up post hoc testing (Figure 2C) revealed that the PubChem fingerprint underperformed significantly compared with MACCS (0.50 vs 0.57; $p < 0.05$), Morgan4 (0.50 vs 0.57; $p < 0.05$), and Morgan6 (0.50 vs 0.56; $p < 0.05$). These results suggest that the performance of the PubChem fingerprint was consistently lower than that of the other fingerprints.

Although these differences were statistically significant, they may not be substantial enough to indicate major disparities in overall fingerprint utility. Nevertheless, the relatively poor performance of the PubChem fingerprint suggests that it may not be optimal for high-stakes applications. This underperformance may be attributed to the key-based structural approach used by PubChem, which may lack the capacity to fully represent molecular complexity compared with circular fingerprints, such as ECFP.

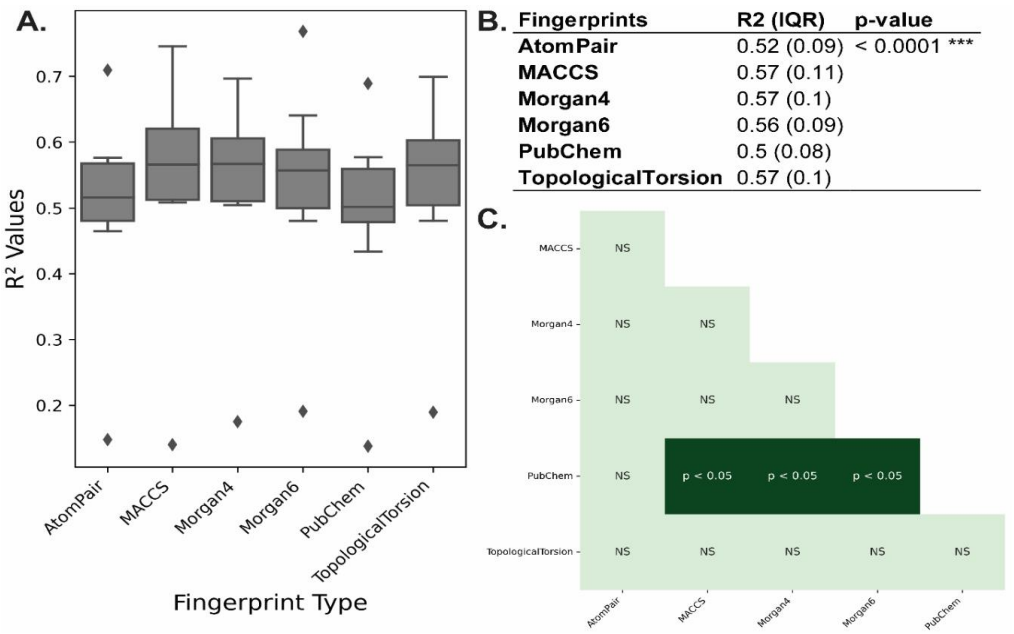


Figure 2: Comparative analysis of QSAR model performance of different molecular fingerprints.

Moreover, the smaller feature space of the PubChem fingerprint increases the risk of bit collisions, where identical bit representations correspond to different compounds.[43] This limitation can reduce the fingerprint's ability to capture key molecular features essential for

predicting compound activity, especially for complex or novel structures. This is further compounded by PubChem's focus on database-driven substructures that lack specificity and do not adapt well to chemically diverse datasets.[14] By contrast, ECFP fingerprints offer flexibility and high resolution by encoding chemical structures without relying on predefined features.[16,44]

An additional noteworthy finding is that the MACCS fingerprint consistently outperformed PubChem, despite both being based on key structural systems. This is remarkable given that MACCS uses a smaller feature space (166 bits) than that of PubChem (881 bits).[13,14] Furthermore, MACCS fingerprints are susceptible to bit collisions, sometimes representing up to eight distinct compounds with a single bit.[43] However, this disparity in performance may be attributable to the optimization of the MACCS fingerprint for bioactivity modeling, whereas PubChem is primarily designed for chemical similarity and substructure searching.[13] Several studies have reported varying findings regarding the predictive performance of MACCS and PubChem, often depending on the specific protein target or toxicological endpoint under investigation.[18,45] Thus, the inconsistencies observed in this study may reflect differences in how each fingerprint encodes functionally relevant molecular features. Some fingerprints are inherently better suited for specific biological or chemical prediction tasks.

Despite these differences, nearly all fingerprints tested demonstrated strong predictive performance. This reinforces the importance of selecting the most appropriate fingerprint for each task to maximize predictive capability across various protein targets.

Analyzing Reproducibility: Consistency in Predictions Across Fingerprints

Different molecular fingerprints encode distinct structural information and may, therefore, yield varying predictive outcomes. The agreement across fingerprints was evaluated to assess the consistency of these predictions. The results demonstrated that the prediction outputs across the various models—Random Forest (Figure 3A), SVM (Figure 3B), and Extreme Gradient Boosting (Figure 3C)—were nearly identical among fingerprint pairs.

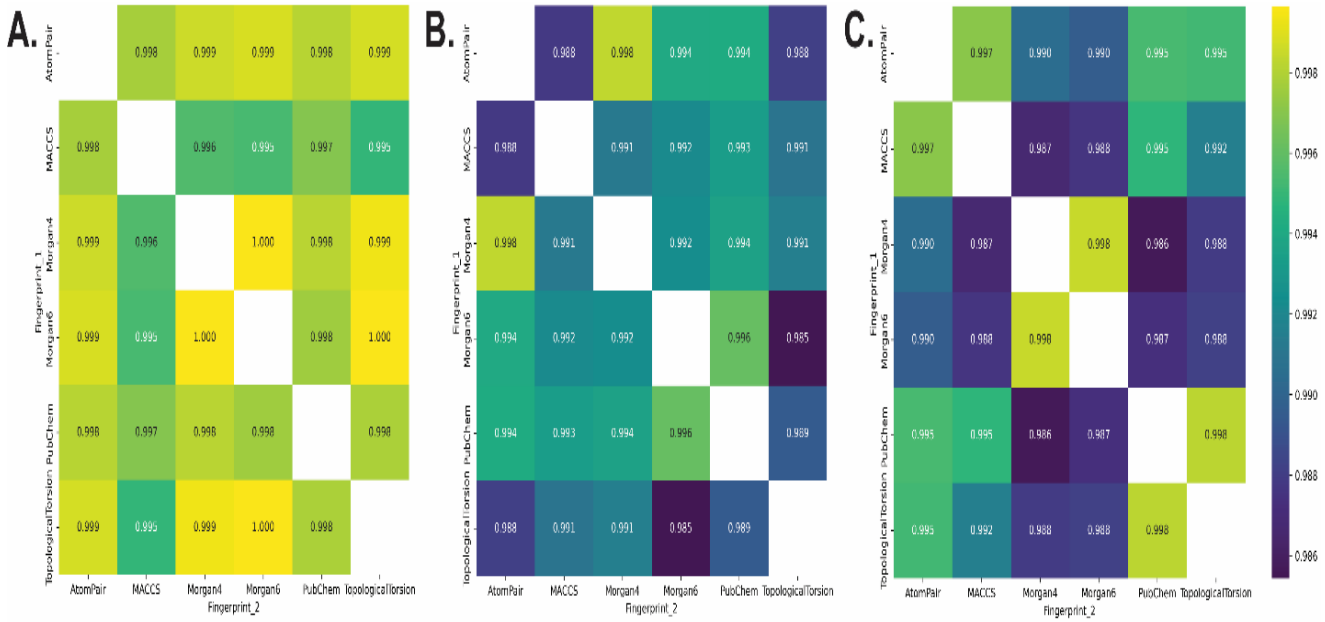


Figure 3: Concordance analysis of molecular fingerprints displayed using heatmap matrices.

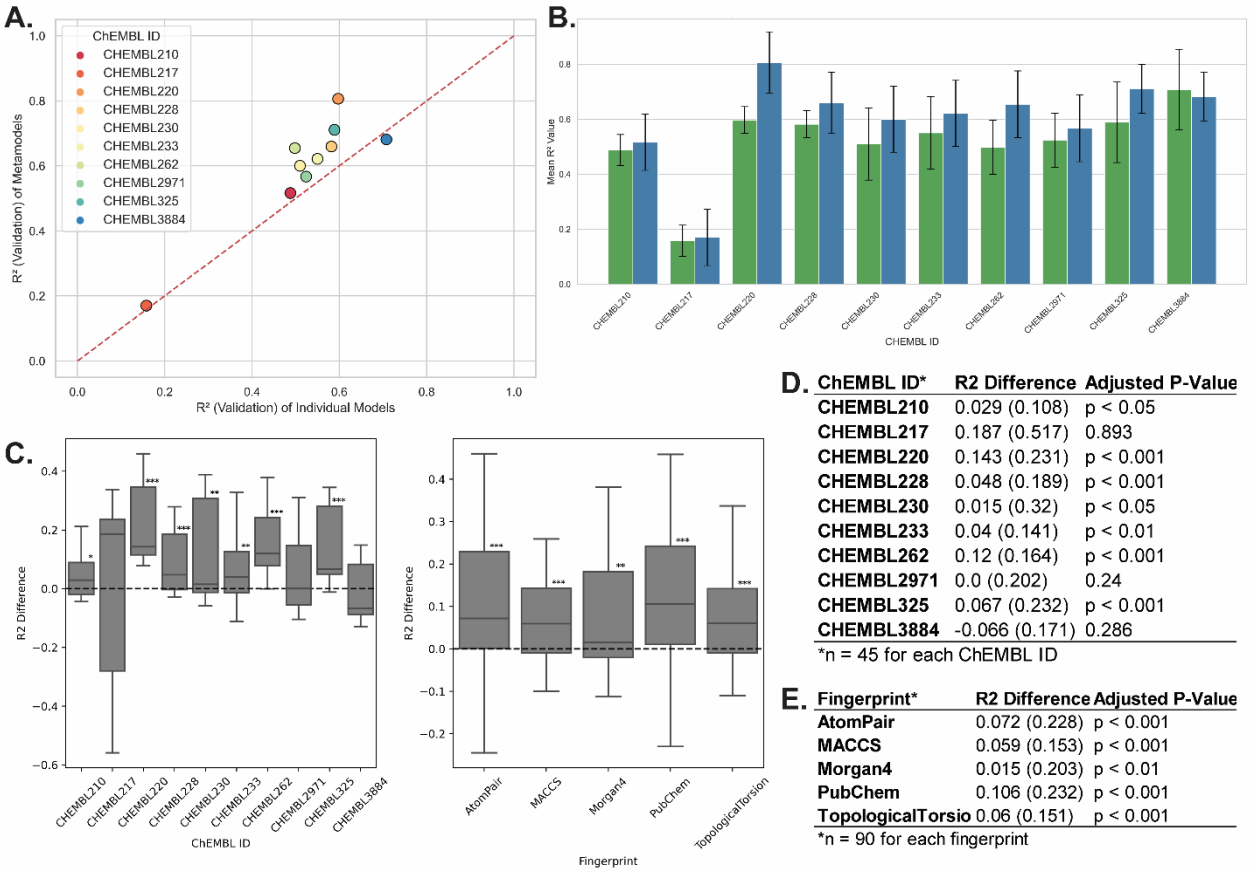
The lowest CCC was 0.985 between the Morgan6 and Topological Torsion fingerprints when using the SVM model. In general, the Random Forest model produced the highest CCC values across fingerprint combinations, indicating the greatest prediction consistency. These findings suggest that the fingerprints tested across these models yielded highly reproducible and consistent results.

CCC has previously been employed to quantify the agreement between model predictions and experimental data, thereby validating the model's generalizability.[24,46] In this study, CCC was used to evaluate whether the predictions based on different fingerprints were consistent. The strong concordance observed indicates that QSAR models constructed using varying fingerprints can achieve comparable predictive performance, despite differences in the structural information encoded by each fingerprint.

These results highlight the high reproducibility of the models used in this study. The strongest concordance was observed in models developed using the Random Forest algorithm, which showed a near-perfect correlation. This underscores the capacity of tree-based approaches to handle diverse molecular representations and model non-linear relationships in data.[47,48] The findings also reinforce the widespread use and strong performance of Random Forests in QSAR modeling.[48–50] Overall, these results further support the robustness and reproducibility of Random Forests when multiple fingerprints are used for QSAR predictions. Based on these results, Random Forest was selected as the base model for further evaluation and integration with meta-modeling approaches.

Unified Meta-Models Improve Prediction Accuracy

The integration of multiple molecular fingerprints was explored to enhance predictive performance. Meta-models incorporating multiple fingerprints consistently outperformed base models that relied on a single fingerprint (Figure 4A and B). Statistical comparison of base and meta-models across individual proteins revealed a statistically significant improvement in seven of the 10 proteins tested (Figure 4C). The median increase in R^2 (Figure 4D) ranged from 0.015 (CHEMBL230) to 0.143 (CHEMBL217).



D.

ChEMBL ID*	R2 Difference	Adjusted P-Value
CHEMBL210	0.029 (0.108)	p < 0.05
CHEMBL217	0.187 (0.517)	0.893
CHEMBL220	0.143 (0.231)	p < 0.001
CHEMBL228	0.048 (0.189)	p < 0.001
CHEMBL230	0.015 (0.32)	p < 0.05
CHEMBL233	0.04 (0.141)	p < 0.01
CHEMBL262	0.12 (0.164)	p < 0.001
CHEMBL2971	0.0 (0.202)	0.24
CHEMBL325	0.067 (0.232)	p < 0.001
CHEMBL3884	-0.066 (0.171)	0.286

*n = 45 for each ChEMBL ID

E.

Fingerprint*	R2 Difference	Adjusted P-Value
AtomPair	0.072 (0.228)	p < 0.001
MACCS	0.059 (0.153)	p < 0.001
Morgan4	0.015 (0.203)	p < 0.01
PubChem	0.106 (0.232)	p < 0.001
TopologicalTorsio	0.06 (0.151)	p < 0.001

*n = 90 for each fingerprint

Figure 4: Comparison of individual model and meta-model performance across different proteins and fingerprints.

Across all fingerprints tested, significant performance gains were recorded when they were used in meta-models in combination with Morgan6 (Figure 4E). The smallest improvement was observed for the Morgan4 fingerprint (R^2 difference = 0.015; $p < 0.01$), whereas the largest was observed for the PubChem fingerprint (R^2 difference = 0.106; $p < 0.001$). These findings indicate that meta-models effectively leverage the complementary characteristics of diverse fingerprints to improve prediction accuracy across a majority of protein targets.

The meta-modeling strategy was based on the premise that combining multiple molecular fingerprints could exploit their individual strengths. For instance, circular fingerprints, such as ECFP/Morgan, encode detailed structural information regarding atom connectivity,[16] which complements the broader, less granular, and more interpretable features encoded by key-based fingerprints, such as MACCS or PubChem.[13,14] This diversity enables meta-models to detect molecular features that may be overlooked when using individual fingerprints alone, thereby enhancing predictive capability.

Prior research supports this approach, demonstrating that ensemble and meta-modeling techniques often outperform single-model strategies.[49,51,52] This increased predictive power is particularly valuable in the early stages of drug discovery, when even minor improvements can significantly influence candidate identification outcomes. Moreover, combining fingerprints with different scopes provides a versatile modeling framework. For example, modeling chemically diverse compounds, such as phytochemicals, may require specialized fingerprints to adequately represent their structural variability.[45] Utilizing a combination of well-established fingerprints, such as ECFP, along with diverse structural representations may yield superior QSAR performance without compromising interpretability or computational efficiency. In this study, the meta-modeling framework was algorithm-agnostic; that is, the same procedure was applied across all algorithms used to construct the base models. Emphasis was placed on evaluating the overall predictive improvement derived from meta-model integration rather than on differences between algorithm-specific advantages.

Fingerprint Contributions to Predictive Modeling Depend on the Target Substance

Feature importance analysis using SHAP values was conducted to investigate whether consistent patterns existed in how different fingerprints contributed to the meta-model predictions. Figure 5A presents a comparison of SHAP value distributions across fingerprints. Morgan4 exhibited the highest median absolute SHAP value; however, this difference was not statistically significant (Figure 5B).

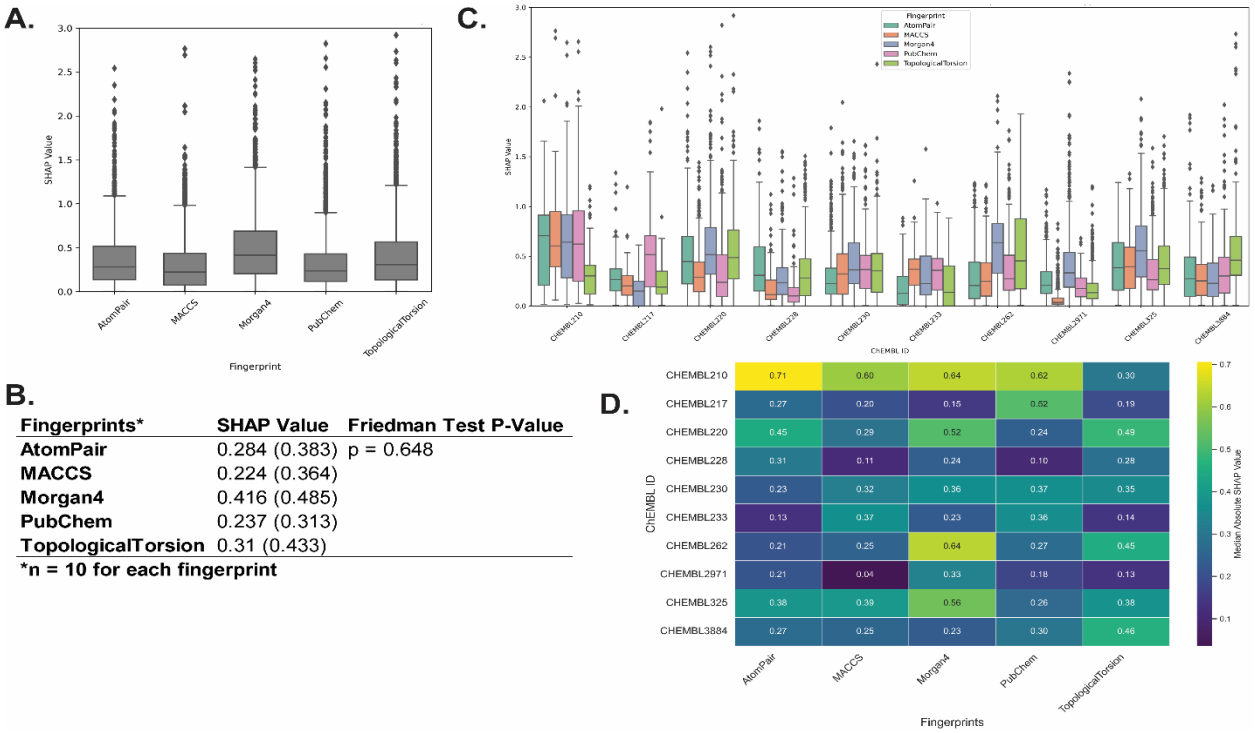
Figure 5C compares the mean absolute SHAP values for each fingerprint grouped by protein. This figure highlights considerable heterogeneity in SHAP values across different proteins, which is further emphasized by the variations in median SHAP values shown in Figure 5D. Specific protein-level analyses were not the focus of this study; rather, the objective was to determine whether any individual fingerprint consistently contributed significantly to the overall predictive performance of the model.

The SHAP analysis conducted in this study provides insights into the relative contributions of each fingerprint to the model predictions. Given the absence of statistically significant differences among fingerprint contributions, it can be inferred that the impact of each fingerprint on meta-model performance—particularly when paired with Morgan6—is largely comparable. This similar contribution across all fingerprints may explain why meta-modeling effectively improves predictive performance: it integrates each fingerprint's unique representation of ligand molecular features.

Further analysis revealed that fingerprint contributions vary across specific proteins, indicating that each fingerprint may have context-dependent importance. This suggests that the predictive relevance of a fingerprint depends on the molecular characteristics of the target protein.

These findings are consistent with prior research demonstrating the value of using multiple fingerprints to capture chemical diversity. Several studies have reported that custom-tailored fingerprints often outperform standard industrial fingerprints, such as ECFP4, when optimized for particular datasets or target classes.[6,44,45,53] This

underscores the importance of selecting fingerprints based on the chemical space and biological context of the QSAR models.
The results of this study reinforce the need to tailor fingerprint selection strategies to the characteristics of the data and goals of the application. Future research should explore the integration of alternative descriptors, molecular fingerprints, and physicochemical features to further enhance model robustness and accuracy.



B.

Fingerprints*	SHAP Value	Friedman Test P-Value
AtomPair	0.284 (0.383)	p = 0.648
MACCS	0.224 (0.364)	
Morgan4	0.416 (0.485)	
PubChem	0.237 (0.313)	
TopologicalTorsion	0.31 (0.433)	

***n = 10 for each fingerprint**

Figure 5: Analysis of SHAP values derived from the meta-models across different fingerprints and proteins

Web Application Development for Prediction and Feature Importance Analysis

The combined analysis demonstrated that individual fingerprints contributed uniquely to predictive performance depending on the target protein. Meta-modeling has been shown to enhance predictive power. To validate this approach, a web-based application was developed to extend prior QSAR modeling work, focusing on two proteins—CHEMBL220 and CHEMBL226—which exhibited the greatest improvement in predictive performance following meta-modeling.

The application was implemented using Streamlit and designed to predict pIC_{50} values from SMILES input. The current version supports up to 10 molecules per submission and computes both Morgan6 and PubChem fingerprints. These fingerprints are then passed into Random Forest base models, and their predictions are aggregated using a meta-model to yield final pIC_{50} outputs. The accuracy gain from using this meta-model is notable, with an approximate increase of 0.10 in R^2 values (Table 2).

Table 2: Comparison of Base Model vs Meta Model Performances

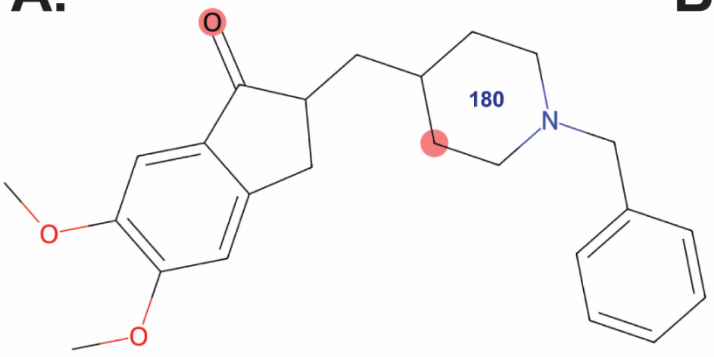
Proteins	Base Models [a]	Model Performance (R2)	Meta Models	Meta Model Performance (R2)
CHEMBL 220	PubChem Fingerprints	0.685	Support Vector Regression [b]	0.814
	Morgan6 – Fingerprints	0.693	XGBoost	0.795
			Elastic Net Regression	0.808
CHEMBL 262	PubChem Fingerprints	0.565	Support Vector Machine	0.671
	Morgan6 – Fingerprints	0.596	XGBoost	0.673
			Elastic Net Regression [b]	0.685

[a] All algorithms used by the base model are Random Forest. [b] The machine learning algorithm used to create the meta-models

The final output includes predicted pIC_{50} values, 2D molecular visualizations annotated with important features, and explanations highlighting key substructures identified by the PubChem fingerprint. This approach was designed to enhance transparency in model prediction and to provide insights into the specific contributions of each fingerprint to the predicted bioactivity.
Combining Morgan6 and PubChem fingerprints enabled a complementary structural perspective. Circular fingerprints, including Morgan6, effectively capture atomic-level structural relationships (including bond-specific contributions), whereas PubChem fingerprints, which are based on predefined features, offer broader but more generalizable representations. Collectively, these findings reflect a balance between specificity and interpretability, which are key advantages in drug design applications.
The functionality of the tool was further evaluated by applying a trained meta-model to predict activity against CHEMBL220 (acetylcholinesterase). Donepezil, a known AChE inhibitor, was used to assess whether key bioactive substructures were identified by the model.
The model identified three positively contributing substructures and several negatively contributing ones. As the model was trained on pIC_{50} values, positive contributions reflect increased inhibitory potency. Specifically, the model attributed the nitrogen atom in the piperidine ring (Figure 6A) as a positive feature, as supported by both Morgan6

and PubChem fingerprints (Bit 180). This result aligns with prior findings by Sugimoto et al., who demonstrated that replacing nitrogen substantially decreases potency.[54] Additionally, both fingerprints indicated that the carbonyl group of the indanone moiety negatively influenced predicted potency (Figure 6B). This is notable because experimental evidence has previously shown that this group is essential for biological activity, as replacements reduce potency. The analysis also revealed that a methoxy group on the indanone moiety was associated with decreased predicted potency, contradicting earlier observations.

While this negative contribution may not imply an actual loss in potency, it highlights a site where further molecular optimization may enhance activity. These case study results illustrate the critical role of model explainability in QSAR applications and underscore the importance of experimental validation as a vital component of model deployment. This example highlights the valuea of interpretable modeling in guiding hypothesis generation for drug discovery.



B.

Bit Location	Substructure	Contribution
Bit: 684	O=C-C-C-C-C	Negative
Bit: 623	O=C-C:C-C	Positive
Bit: 180	>= 1 saturated or aromatic nitrogen-containing ring size 6	Positive

Figure 6: Feature importance analysis of Donepezil.

Conclusion

Our study demonstrated the effectiveness of integrating various molecular fingerprints using meta-models to predict drug bioactivity. The results showed that combining Morgan6 and PubChem fingerprints enhanced predictive performance compared to using individual fingerprints, with statistically significant improvements in R^2 values across multiple proteins. Furthermore, SHAP analysis provided a deeper understanding of fingerprint contributions, revealing context-dependent importance based on target proteins. The utility of this approach in drug discovery was demonstrated by developing a web-based application that integrated predictions with feature importance visualization. This approach adds explainability to QSAR modeling and offers actionable insights into how molecular substructures contribute to bioactivity predictions.

Our study has several notable limitations. First, it focused only on two fingerprints (Morgan6 and PubChem), which may not capture the full complexity of molecular structures. This choice may only partially represent datasets with more complex chemical diversity, and a wider selection of fingerprints must be utilized for each QSAR dataset being modeled. The web-based application was also designed as a proof of concept for small datasets. Although scalability may be a concern, its performance on other unseen datasets remains to be tested. Future improvements should focus on incorporating additional fingerprints, such as 3D fingerprints or other molecular descriptors, to enhance robustness and flexibility across various applications. The validation of large and diverse datasets, covering different molecular types and target proteins, is also required to ensure the generalizability and reliability of model predictions. This study highlights the value of combining fingerprints for robust QSAR modeling and interpretable feature analysis to improve the drug discovery workflow.

Conflict of Interests

The authors declare no conflict of interest

Authors' Declaration

The authors hereby declare that the work presented in this article is original and that any liability for claims relating to the content of this article will be borne by them.

Acknowledgments

The authors thanked Maranatha Christian University for funding and related facilities in this research.

Funding

This research was funded by the Internal Research Grant from Maranatha Christian University (023/SK/ADD/UKM/V/2024) to Ardo Sanjaya and Hana Ratnawati.

References

1. Hansch Corwin, Fujita Toshio. p -σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. J Am Chem Soc. 1964;86(8):1616-1626. doi:10.1021/ja01062a035
2. Divina E, Erlina L, Anbar Istiadi K, Yasa Putri ES, Fadilah F. QSAR and Anticancer effect of Alkyl salicylate on Breast cancer T47D Cells Lines. Res J Pharm Technol. 2022:4607-4613. doi:10.52711/0974-360X.2022.00773
3. Hamada Hakim A. Predictive QSAR models for the toxicity of Phenols. Asian J Res Chem. 2022:433-438. doi:10.52711/0974-4150.2022.00076
4. Prema V, Sivaramakrishnan M, Rabiya M. A Concise Review on role of QSAR in Drug Design. Asian J Res Chem. 2023:459-466. doi:10.52711/0974-4150.2023.00076
5. K P, Fathima F, Ray R, Kumar L, Verma R. Atom based 3D QSAR and Fingerprint based 2D QSAR of Novel Molecules as MmpL3 receptor inhibitors for *Mycobacterium tuberculosis*. Res J Pharm Technol. 2021:6321-6329. doi:10.52711/0974-360X.2021.01093
6. Boldini D, Ballabio D, Consonni V, Todeschini R, Grisoni F, Sieber SA. Effectiveness of molecular fingerprints for exploring the chemical space of natural products. J Cheminformatics. 2024;16(1):35. doi:10.1186/s13321-024-00830-3
7. De P, Kar S, Ambure P, Roy K. Prediction reliability of QSAR models: an overview of various validation tools. Arch Toxicol. 2022;96(5):1279-1295. doi:10.1007/s00204-022-03252-y

8. Ucak UV, Ashyrmamatov I, Lee J. Reconstruction of lossless molecular representations from fingerprints. J Cheminformatics. 2023;15(1):26. doi:10.1186/s13321-023-00693-0
9. Ehimen A. Erazua, Sunday A. Akintelu, Joy M. Adelowo, Simon N. Odoemene, Olubunmi M. Josiah, Sakirat Funke Raheem, Dayo F. Latona, Moriam D. Adeoye, Akintomiwa O. Esan, Abel K. Oyebamiji. QSAR and Molecular Docking Studies on Nitro (Triazole/Imidazole)-Based Compounds as Anti-Tubercular Agents. Trop J Nat Prod Res. 2021;5(11):2022-2029. doi:10.26538/tjnpr/v5i11.22
10. Niazi SK, Mariam Z. Recent Advances in Machine-Learning-Based Chemoinformatics: A Comprehensive Review. Int J Mol Sci. 2023;24(14):11488. doi:10.3390/ijms241411488
11. Wijianto B, Ritmaleni, Purnomo H, Nurrochmad A. Quantitative structure activity relationship (qsar) study and biological evaluation on mono-ketone analogs of curcumin as antioxidant. Res J Pharm Technol. 2020;13(10):4829-4829. doi:10.5958/0974-360X.2020.00850.1
12. Reisky M. Tammu, Lastiar R. Sitompul, Rahel Simamora, Didik H. Utomo, Eka D. Putri. Revealing the Potential of New Immunomodulatory Agents from Katokkon Pepper as a Native Toraja Plant. Trop J Nat Prod Res. 2025;9(1). doi:10.26538/tjnpr/v9i1.36
13. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL Keys for Use in Drug Discovery. J Chem Inf Comput Sci. 2002;42(6):1273-1280. doi:10.1021/ci010132r
14. National Library of Medicine. PubChem Substructure Fingerprint. National Institutes of Health (NIH); 2009. https://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.pdf
15. Carhart RE, Smith DH, Venkataraghavan R. Atom pairs as molecular features in structure-activity studies: definition and applications. J Chem Inf Comput Sci. 1985;25(2):64-73. doi:10.1021/ci00046a002
16. Rogers D, Hahn M. Extended-Connectivity Fingerprints. J Chem Inf Model. 2010;50(5):742-754. doi:10.1021/ci100050t
17. Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, Oprea TI, Baskin II, Varnek A, Roitberg A, Isayev O, Curtalolo S, Fourches D, Cohen Y, Aspuru-Guzik A, Winkler DA, Agrafiotis D, Cherkasov A, Tropsha A. QSAR without borders. Chem Soc Rev. 2020;49(11):3525-3564. doi:10.1039/D0CS00098A
18. Orosz Á, Héberger K, Rácz A. Comparison of Descriptor- and Fingerprint Sets in Machine Learning Models for ADME-Tox Targets. Front Chem. 2022;10:852893. doi:10.3389/fchem.2022.852893
19. Sainy N, Dubey N, Sharma R, Dubey N, Sainy J. 3D QSAR Analysis of Flavones as Antidiabetic agents. Res J Pharm Technol. 2022:1689-1695. doi:10.52711/0974-360X.2022.00283
20. Afif Z, Eddy Santoso MI, Khotimah H, Satriotomo I, Widjajanto E, Rahayu M, Rianawati SB, Nandar Kurniawan S, Rakhmatiar R, Sapti Iskandar D, Hakimah A, Azizah S, Andriani N, Agustina K. *Centella asiatica* improved Insomnia through MAPK/ERK Signaling Pathway: In Silico Study. Res J Pharm Technol. 2023:587-592. doi:10.52711/0974-360X.2023.00100
21. Mahtal A, Toufik H, Akabli T, Lamchouri F. In Silico Screening of New Derivatives as Inhibitors of Enoyl-[Acyl-Carrier-Protein] Reductase From *Staphylococcus Aureus* Via 2D-QSAR Analysis, Molecular Docking and ADME/Tox Prediction. Trop J Nat Prod Res. 2025;9(4):1616. doi:10.26538/tjnpr/v9i4.34
22. Motamedi F, Pérez-Sánchez H, Mehridehnavi A, Fassihi A, Ghasemi F. Accelerating Big Data Analysis through LASSO-Random Forest Algorithm in QSAR Studies. Xu J, ed. Bioinformatics. 2022;38(2):469-475. doi:10.1093/bioinformatics/btab659
23. Cai C, Wang S, Xu Y, Zhang W, Tang K, Ouyang Q, Lai L, Pei J. Transfer Learning for Drug Discovery. J Med Chem. 2020;63(16):8683-8694. doi:10.1021/acs.jmedchem.9b02147
24. Kwon S, Bae H, Jo J, Yoon S. Comprehensive ensemble in QSAR prediction for drug discovery. BMC Bioinformatics. 2019;20(1):521. doi:10.1186/s12859-019-3135-4
25. Xu Y. Deep Neural Networks for QSAR. In: Heifetz A, ed. Artificial Intelligence in Drug Design. Vol 2390. Methods in Molecular Biology. Springer US; 2022:233-260. doi:10.1007/978-1-0716-1787-8_10
26. Gini GC. QSAR: Using the Past to Study the Present. In: Nicolotti O, ed. Computational Toxicology. Vol 2834. Methods in Molecular Biology. Springer US; 2025:3-39. doi:10.1007/978-1-0716-4003-6_1
27. Ji H, Deng H, Lu H, Zhang Z. Predicting a Molecular Fingerprint from an Electron Ionization Mass Spectrum with Deep Neural Networks. Anal Chem. 2020;92(13):8649-8653. doi:10.1021/acs.analchem.0c01450
28. Landrum G, Tosco P, Kelley B, Ric (Riccardo), Sriniker (GitHub: sriniker), Gedeck P, Vianello R, Cosgrove D, Schneider N, Kawashima E, Nealschneider D (Dan N), Dalke A, Jones G, Cole B, Swain M, Turk S, Savelyev A, Vaucher A, Wójcikowski M, Take I, Probst D, Scalfani VF, Ujihara K, Godin G, Pahl A, Berenger F, Varjo JL, Biggs J, [strets123], JP. RDKit: Open-source cheminformatics. 2023. doi:10.5281/zenodo.8053810.
29. Nilakantan R, Bauman N, Dixon JS, Venkataraghavan R. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. J Chem Inf Comput Sci. 1987;27(2):82-85. doi:10.1021/ci00054a008
30. Szafarczyk M, Ludynia P, Kukla P. A Python library for efficient computation of molecular fingerprints. 2024. doi:10.48550/ARXIV.2403.19718
31. Khan MZI, Ren JN, Cao C, Ye HYX, Wang H, Guo YM, Yang JR, Chen JZ. Comprehensive hepatotoxicity prediction: ensemble model integrating machine learning and deep learning. Front Pharmacol. 2024;15:1441587. doi:10.3389/fphar.2024.1441587
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(85):2825-2830. http://jmlr.org/papers/v12/pedregosa11a.html
33. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J Chem Inf Comput Sci. 2003;43(6):1947-1958. doi:10.1021/ci034160g
34. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl. 1998;13(4):18-28. doi:10.1109/5254.708428
35. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016:785-794. doi:10.1145/2939672.2939785
36. Boldini D, Grisoni F, Kuhn D, Friedrich L, Sieber SA. Practical guidelines for the use of gradient boosting for molecular property prediction. J Cheminformatics. 2023;15(1):73. doi:10.1186/s13321-023-00743-7
37. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM; 2019:2623-2631. doi:10.1145/3292500.3330701

38. Shayanfar S, Shayanfar A. Comparison of various methods for validity evaluation of QSAR models. BMC Chem. 2022;16(1):63. doi:10.1186/s13065-022-00856-4
39. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, Van Der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2
40. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. In: ; 2010:92-96. doi:10.25080/Majora-92bf1922-011
41. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9(3):90-95. doi:10.1109/MCSE.2007.55
42. Waskom M. seaborn: statistical data visualization. J Open Source Softw. 2021;6(60):3021. doi:10.21105/joss.03021
43. Kuwahara H, Gao X. Analysis of the effects of related fingerprints on molecular similarity using an eigenvalue entropy approach. J Cheminformatics. 2021;13(1):27. doi:10.1186/s13321-021-00506-2
44. Asahara R, Miyao T. Extended Connectivity Fingerprints as a Chemical Reaction Representation for Enantioselective Organophosphorus-Catalyzed Asymmetric Reaction Prediction. ACS Omega. 2022;7(30):26952-26964. doi:10.1021/acsomega.2c03812
45. Seo M, Shin HK, Myung Y, Hwang S, No KT. Development of Natural Compound Molecular Fingerprint (NC-MFP) with the Dictionary of Natural Products (DNP) for natural product-based drug development. J Cheminformatics. 2020;12(1):6. doi:10.1186/s13321-020-0410-3
46. Chirico N, Gramatica P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. J Chem Inf Model. 2011;51(9):2320-2335. doi:10.1021/ci200211n
47. Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N, Cedrón F, Novoa FJ, Carballal A, Maojo V, Pazos A, Fernandez-Lozano C. A review on machine learning approaches and trends in drug discovery. Comput Struct Biotechnol J. 2021;19:4538-4558. doi:10.1016/j.csbj.2021.08.011
48. Nayarisseri A, Khandelwal R, Tanwar P, Madhavi M, Sharma D, Thakur G, Speck-Planche A, Singh SK. Artificial Intelligence, Big Data and Machine Learning Approaches in Precision Medicine & Drug Discovery. Curr Drug Targets. 2021;22(6):631-655. doi:10.2174/1389450122999210104205732
49. Schlender T, Viljanen M, Van Rijn JN, Mohr F, Peijnenburg WJgm, Hoos HH, Rorije E, Wong A. The Bigger Fish: A Comparison of Meta-Learning QSAR Models on Low-Resourced Aquatic Toxicity Regression Tasks. Environ Sci Technol. 2023;57(46):17818-17830. doi:10.1021/acs.est.3c00334
50. Trinh TX, Seo M, Yoon TH, Kim J. Developing random forest based QSAR models for predicting the mixture toxicity of TiO2 based nano-mixtures to *Daphnia magna*. NanoImpact. 2022;25:100383. doi:10.1016/j.impact.2022.100383
51. Cheng Z, Hwang SS, Bhave M, Rahman T, Chee Wezen X. Combination of QSAR Modeling and Hybrid-Based Consensus Scoring to Identify Dual-Targeting Inhibitors of PLK1 and p38γ. J Chem Inf Model. 2023;63(21):6912-6924. doi:10.1021/acs.jcim.3c01252
52. Padi S, Cardone A, Sriram RD. A Meta-model for ADMET Property Prediction Analysis. 2023. doi:10.1101/2023.12.05.570279
53. Venkatraman V, Gaiser J, Demekas D, Roy A, Xiong R, Wheeler TJ. Do Molecular Fingerprints Identify Diverse Active Drugs in Large-Scale Virtual Screening? (No). Pharmaceuticals. 2024;17(8):992. doi:10.3390/ph17080992
54. Sugimoto H, Yamanish Y, Iimura Y, Kawakami Y. Donepezil Hydrochloride (E2020) and Other Acetylcholinesterase Inhibitors. Curr Med Chem. 2000;7(3):303-339. doi:10.2174/0929867003375191
55. Sanjaya A. Enhancing QSAR Predictive Power and Explainability: Meta-Modeling and SHAP Feature Importance Analysis for Drug Discovery (Python Scripts). 2024:15052624 Bytes. doi:10.6084/M9.FIGSHARE.28014527.V1