Original Article

Random forest-based QSAR modeling for predicting the potency of neprilysin inhibitors using Mordred molecular descriptors

Nizam Albar[1*], Derren DCH. Rampengan[2], Saiful Azhari[3], Mahmudi Mahmudi[3], Farrah Fahdhienie[4], Anggi Susilawati[5] and Muhammad Habiburrahman[6*]

[1]Department of Computer Engineering, Faculty of Engineering, Universitas Serambi Mekkah, Banda Aceh, Indonesia; [2]Faculty of Medicine, Universitas Sam Ratulangi, Manado, Indonesia; [3]Department of Pharmacy, STIKES Assyifa Aceh, Banda Aceh, Indonesia; [4]Faculty of Public Health, Universitas Muhammadiyah Aceh, Banda Aceh, Indonesia; [5]Faculty of Teacher Training and Education, Universitas Bina Bangsa Getsempena, Banda Aceh, Indonesia; [6]Faculty of Medicine, Imperial College London, London, United Kingdom

*Correspondence: nizam.mte@gmail.com (NA) and h.habiburrahman23@imperial.ac.uk (MH)

Abstract

Neprilysin (NEP) is a zinc-dependent metallopeptidase, considered a key therapeutic target in heart failure management. Efficient identification of potent NEP inhibitors remains a challenge in drug discovery. The aim of this study was to develop a quantitative structure–activity relationship (QSAR) model using 2D Mordred molecular descriptors and Random Forest algorithms to predict the inhibitory potency (pIC_{50}) of drug candidates. A curated dataset of compounds with experimentally determined IC_{50} values (in nM) against NEP was preprocessed and converted to pIC_{50}. Mordred was used to calculate 2D molecular descriptors, and descriptors with missing values were excluded. The dataset was split into training, internal validation, and external test sets. A Random Forest regression model was trained using 500 estimators, and model performance was evaluated using R^2, root mean square error (RMSE), mean absolute error (MAE), and concordance correlation coefficient (CCC), while a binary classification model was also constructed. Feature importance, residual analysis, and chemical space visualization were conducted to assess model interpretability and reliability. The regression model demonstrated moderate to strong predictive performance, with R^2 of 0.286, RMSE of 0.949, MAE of 0.723, and CCC of 0.532 in the internal validation. External validation showed improved generalization, with R^2=0.659, RMSE=0.858, MAE=0.630, and CCC=0.763. Binary classification revealed an accuracy of 0.953, precision of 1.000, recall of 0.943, and an F1-score of 0.971, indicating strong discriminative ability in classifying inhibitory versus non-inhibitory compounds. Top contributing descriptors included ATSC2p (feature importance=0.0505), GATS2p (0.0408), and SaasC (0.0317). Principal component analysis (PCA) and Williams plots confirmed that test compounds lie within the model's applicability domain, with no major outliers in leverage or residual distribution. The developed Random Forest-based QSAR model demonstrates strong predictive power and interpretability for identifying NEP inhibitors. This study provides a valuable tool for virtual screening and highlights the relevance of 2D structural features in governing NEP inhibitory activity. It is the first dedicated QSAR analysis of neprilysin inhibition using Mordred descriptors with rigorous internal and external validation.

Keywords: Cardiology drug, drug discovery, heart failure, machine learning, quantitative structure-activity relationship

Received: December 1, 2025 - Revised: February 3, 2026 - Accepted: March 27, 2026 - Published online: April 1, 2026

Introduction

Neprilysin (NEP), also known as membrane metalloendopeptidase (MME; EC 3.4.24.11), is a zinc-dependent enzyme abundantly expressed in renal, pulmonary, and cardiovascular tissues. It plays a central role in maintaining cardiovascular homeostasis by catalyzing the degradation of vasoactive peptides such as natriuretic peptides, bradykinin, adrenomedullin, and substance P—key regulators of vascular tone, fluid balance, and cardiac function [1,2]. NEP activity ensures the balanced turnover of these peptides under normal physiological conditions [3]. However, in chronic heart failure, elevated NEP activity leads to excessive breakdown of beneficial natriuretic peptides, contributing to fluid retention, vasoconstriction, and progressive cardiac dysfunction [1,4].

Targeting NEP has emerged as a compelling therapeutic approach [1,4]. The most notable advance in this domain is the development of sacubitril/valsartan, an angiotensin receptor–neprilysin inhibitor (ARNi) that combines NEP inhibition with blockade of the renin–angiotensin system [5-7]. Clinical evidence from the PARADIGM-HF and PARAGON-HF trials demonstrated that ARNi therapy significantly reduces mortality and hospitalization rates in patients with heart failure with reduced ejection fraction (HFrEF), outperforming traditional RAAS inhibitors [8,9]. These findings have firmly established NEP inhibition as a cornerstone in contemporary heart failure management and have catalyzed ongoing efforts to develop novel NEP inhibitors with enhanced potency, selectivity, and pharmacokinetic profiles. In this light, there is a growing demand for efficient methods to identify and optimize potential inhibitors [2,10]. Computational strategies such as structure-based docking, pharmacophore modeling, and ligand-based quantitative structure–activity relationship (QSAR) modeling have become essential tools in modern drug discovery workflows [11]. QSAR modeling provides a rapid and cost-effective approach for predicting biological activity based on chemical structure, enabling the prioritization of promising candidates before synthesis or biological testing [12,13].

In particular, the use of two-dimensional (2D) molecular descriptors has gained popularity due to their interpretability and computational efficiency. Mordred, a versatile Python-based platform, can generate a wide array of 2D descriptors, offering a rich chemical representation suitable for high-throughput QSAR modeling [14,15]. Despite its broad applicability, the systematic use of Mordred descriptors in QSAR models specifically targeting NEP inhibitors remains largely unexplored. Although recent studies have applied multi-target QSAR (mt-QSAR) models to NEP alongside other enzymes, these efforts often rely on structural fingerprints and prioritize polypharmacology over target-specific interpretability [16,17]. Moreover, such studies frequently lack rigorous external validation or comprehensive classification analysis, limiting their practical utility [16,17].

In this study, the aim was to fill this gap by developing a dedicated QSAR model to predict the inhibitory potency (pIC_{50}) of NEP inhibitor candidates using 2D Mordred descriptors and a Random Forest regression algorithm. The model was constructed using ChEMBL-derived datasets. Random Forest was selected for its ability to capture nonlinear structure–activity relationships, identify complex substructure patterns, and maintain robustness against noise with lower risk of overfitting [18,19]. Mordred was employed because it provides more than 1,800 molecular descriptors with seamless integration into RDKit-based, fully reproducible QSAR workflows [20]. A binary classification framework based on activity thresholds and assessing model performance through extensive internal and external validation was further implemented. Feature importance ranking, chemical space visualization, and applicability domain analysis are also performed to enhance model interpretability and reliability, as recommended previously [20]. To our knowledge, this is the first study focusing exclusively on QSAR modeling of NEP inhibition using this methodological combination.

Methods

Dataset preparation and descriptor generation

A dataset of compounds targeting NEP was compiled from the ChEMBL database (as of July 22, 2025), a publicly available repository of bioactivities. The dataset was filtered to retain only

entries with IC_{50} values reported in nanomolar (nM) units and measured using exact equality relations ("="). SMILES strings were sanitized and converted into valid molecular objects using RDKit. IC_{50} values were then transformed into molar units and converted to pIC_{50} values ($-\log_{10}[IC_{50}$ in M]). Two-dimensional molecular descriptors were generated using Mordred, and compounds that failed SMILES parsing or contained missing descriptor values were excluded. The final dataset comprised N valid compounds, each associated with clean descriptor profiles and corresponding pIC_{50} values. The complete preprocessing and modeling workflow is presented in **Figure 1**.

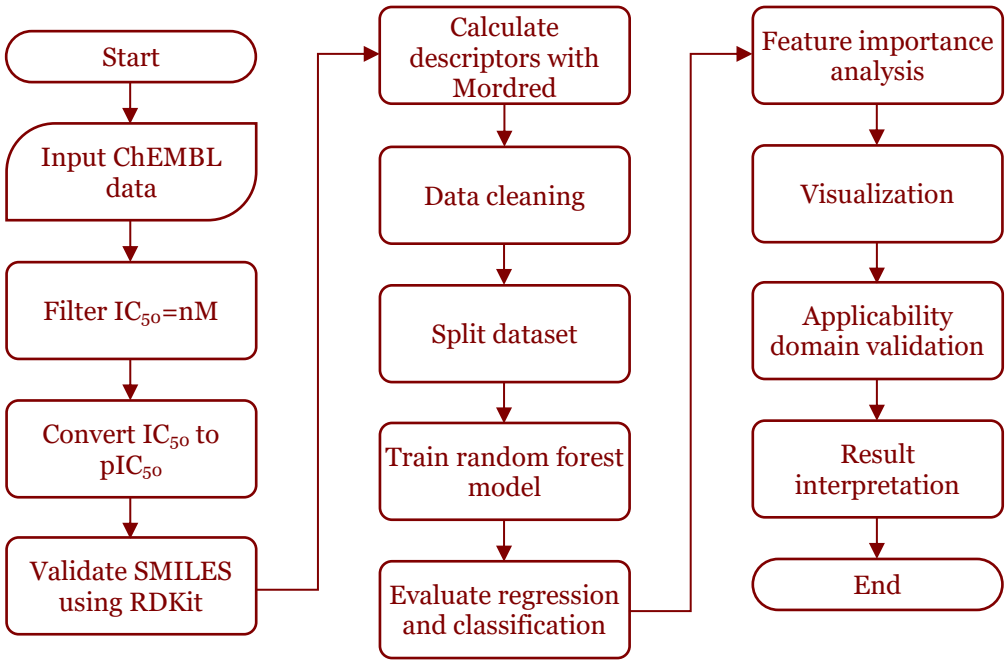

Figure 1. Workflow of the QSAR modeling pipeline to predict neprilysin inhibitors using random Forest model.

Model development and evaluation

Numerical descriptor columns (excluding IC_{50}-related columns) were used as input features. The dataset was split into training (81%), validation (9%), and external test (10%) subsets. A Random Forest Regressor (500 trees) was trained on the training set. Missing values were imputed with the column means. Model performance was assessed on the validation and external sets using R^2, root mean square error (RMSE), mean absolute error (MAE), and concordance correlation coefficient (CCC). Parity plots, residual plots, and histograms were used to visualize prediction quality.

Feature importance and classification modeling

Top 20 most important Mordred descriptors were ranked using feature importance scores from the regression model and visualized on linear and logarithmic scales. For classification, compounds were labeled as active ($pIC_{50} \geq 6$) or inactive. This threshold corresponds to an IC^{50} value of 1 µM, a commonly accepted cutoff in medicinal chemistry to define biologically relevant inhibition [21]. The cutoff was also selected to maintain a balanced class distribution, minimizing model bias toward the majority class and improving the reliability of classifier performance, in line with a published QSAR guideline [22]. A Random Forest Classifier was trained and evaluated using accuracy, precision, recall, F1-score, and confusion matrix visualization. A bar plot of classification metrics and a labeled confusion matrix were also provided.

Chemical space and applicability domain analysis

Principal Component Analysis (PCA) was conducted using standardized Mordred descriptors to visualize the chemical space and assess the distribution of training and test compounds. This allowed evaluation of whether test compounds were projected within the learned chemical space

of the training data. A Williams plot was constructed using standardized residuals and leverage values, calculated from the standardized feature matrix, to define the model's applicability domain and identify potential outliers or influential compounds. Additionally, a descriptor correlation heatmap was generated to examine patterns of multicollinearity among the numerical features.

Software and reproducibility

All analyses were conducted using Python 3.9. Libraries included rdkit for molecular handling, mordred for descriptor calculation, scikit-learn for machine learning, and matplotlib/seaborn for visualization. All code and outputs are available upon request to ensure reproducibility.

Results

Model performance

The parity plots for internal and external validation illustrating the relationship between observed and predicted pIC_{50} values are presented in **Figure 2A** and **2B**. In internal validation, the model yielded an R^2 of 0.286, RMSE of 0.949, MAE of 0.723, and a CCC of 0.532—indicating modest predictive accuracy within the training data. The regression line deviated notably from the identity line, suggesting limited precision and greater dispersion in predictions. External validation, however, showed improved performance with an R^2 of 0.659, RMSE of 0.858, MAE of 0.630, and CCC of 0.763. While still showing deviation at certain pIC_{50} ranges, predictions tended to follow the observed trend, indicating moderate external predictivity. The confusion matrix of the model in separating active and less active NEP inhibitor candidates is presented in **Figure 3**. Full performance metrics, including classification task results (accuracy=0.953, precision=1.000, recall=0.943, F1-score=0.971), are presented in **Table 1**.

Table 1. Performance metrics of the Random Forest models for regression and classification tasks

Model type and parameters	Value
Regression	
Internal	
R^2	0.286
Root mean square error (RMSE)	0.949
Mean absolute error (MAE)	0.723
Concordance correlation coefficient (CCC)	0.532
External	
R^2	0.659
RMSE	0.858
MAE	0.630
CCC	0.763
Classification	
Accuracy	0.953
Precision	1.000
Recall	0.943
F1-score	0.971

Feature importance analysis

The top 20 Mordred descriptors, illustrated in **Figure 4**, were ranked based on their relative importance values, which ranged from 0.0052 to 0.0505. The most influential descriptor was ATSC2p (centered Moreau–Broto autocorrelation of lag 2 weighted by polarizability) with an importance score of 0.0505, followed by GATS2p (Geary autocorrelation of lag 2 weighted by polarizability, 0.0408) and SaasC (atom-type E-state: count of C(sp2) connected to two single bonds, 0.0317). Other relevant contributors included MINssCH2 (minimum atom-type E-state index for $-CH_2-$, 0.0203), VSA_EState3 (sum of electrotopological state values in a particular VSA bin, 0.0198), and GATS3p (0.0176). These findings highlight the importance of topological autocorrelation, electronic distribution, and atom-type characteristics in driving the inhibitory activity against NEP. Collectively, these descriptors suggest that both electronic properties and molecular topology play key roles in determining compound potency.

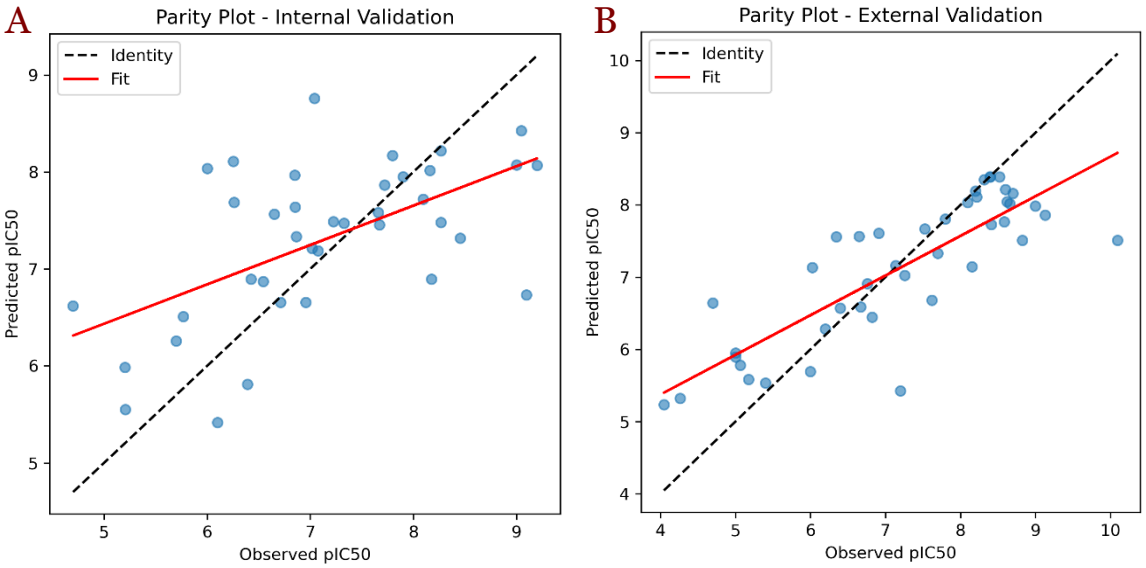

Figure 2. Parity plots showing predicted versus observed pIC50 values for internal (A) and external (B) validation sets. The dashed line represents the ideal identity line (perfect prediction), while the red line indicates the linear regression fit.

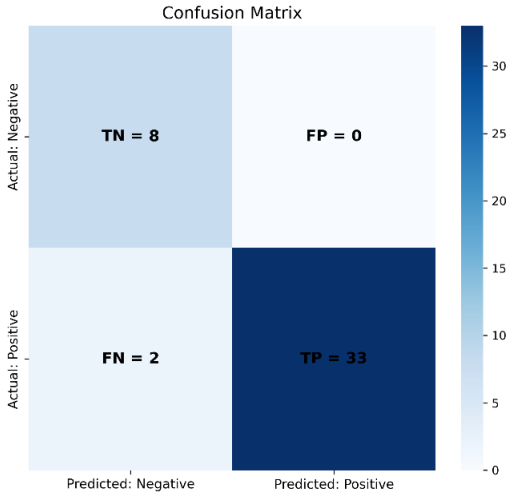

Figure 3. Labeled confusion matrix showing classification outcomes for active ($pIC_{50} \geq 6$) and inactive ($pIC_{50} < 6$) neprilysin inhibitors in the external test set.

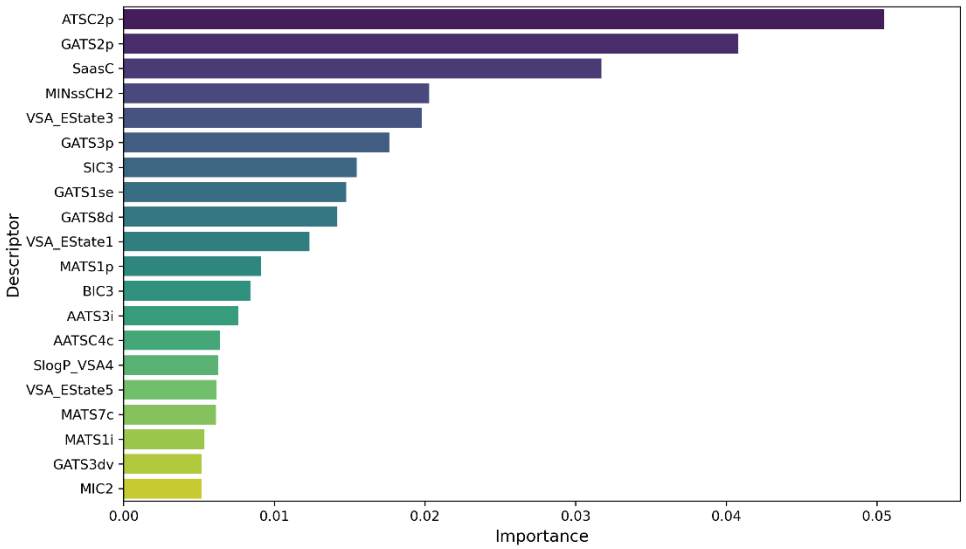

Figure 4. Top 20 most important Mordred descriptors ranked by feature importance in the Random Forest regression model for pIC_{50} prediction. Descriptors related to structural complexity, molecular symmetry, and polarizability showed the highest contributions to model performance.

Descriptor correlation analysis

A Pearson correlation heatmap was generated using all numerical Mordred descriptors included in the model to assess potential redundancy among molecular descriptors (**Figure 5**). The heatmap revealed multiple clusters of highly correlated descriptors, particularly within the same descriptor families (e.g., autocorrelation, topological, and EState indices), suggesting internal structural coherence. While many descriptor pairs showed moderate to strong positive or negative correlations, others appeared largely uncorrelated, indicating a diverse feature set. This pattern underscores the importance of including a robust machine learning algorithm—such as Random Forest—that is resilient to multicollinearity during model training. The visualization also highlights opportunities for future dimensionality reduction or descriptor pruning to enhance model interpretability (**Figure 5**).

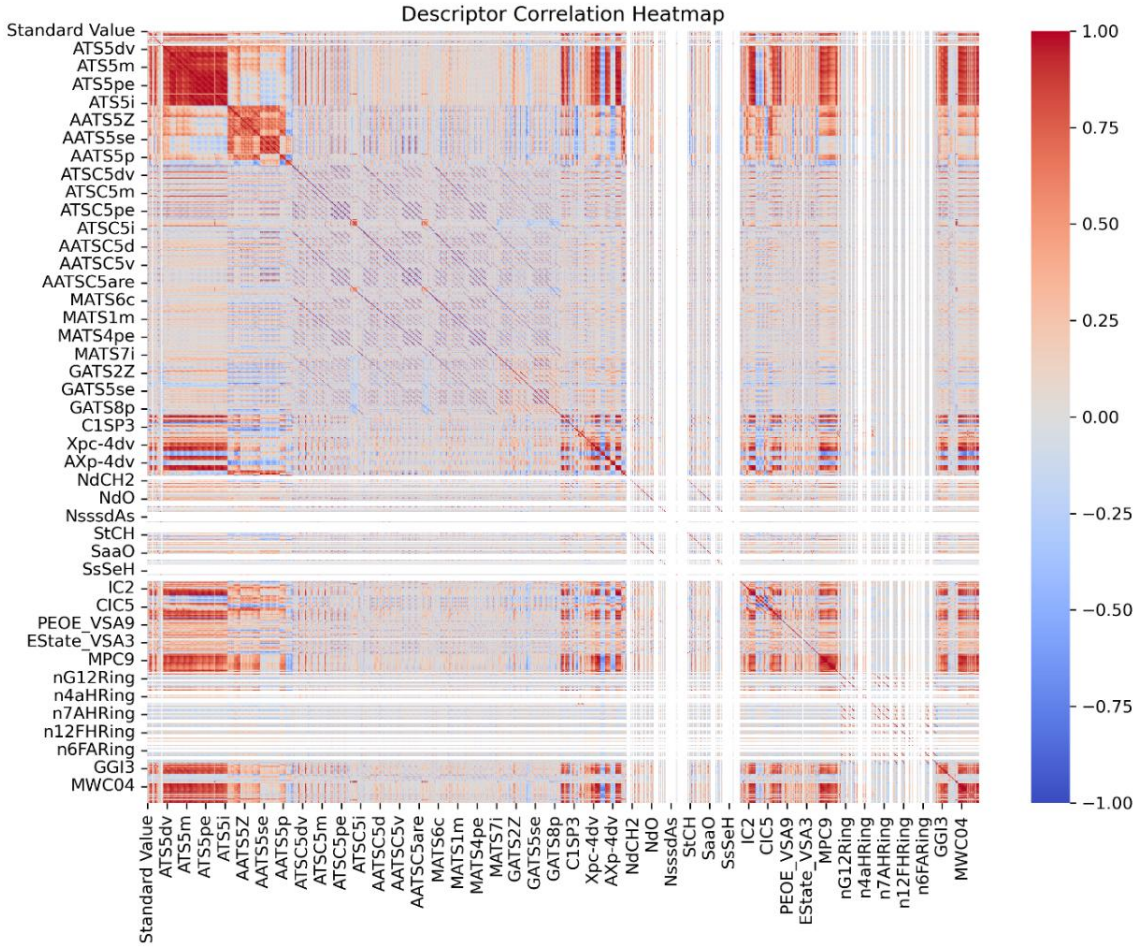

Figure 5. Pearson correlation heatmap of Mordred-derived molecular descriptors used in the QSAR model. Red and blue indicate strong positive and negative correlations, respectively, while white reflects weak or no correlation. Distinct correlation blocks correspond to descriptor families, highlighting both redundancy and variability across the feature space.

Residual analysis

Results from residual analysis conducted for both internal and external validation sets are presented in **Figure 6**. Histogram plots demonstrated that residuals were generally centered around zero, with the external validation residuals exhibiting a slightly broader distribution. The internal validation histogram showed moderate skewness and dispersion, while the external residuals followed a more symmetric, bell-shaped pattern, supporting the model's generalization. Residual scatter plots against observed pIC_{50} values revealed no pronounced heteroscedasticity in either dataset. In the internal set, residuals showed a relatively wide spread across the pIC_{50} range, consistent with the observed underperformance metrics (R^2=0.286; RMSE=0.949; CCC=0.532; **Table 1**). In contrast, the external set showed a tighter clustering of residuals near

zero, particularly in mid-to-high pIC_{50} ranges, aligning with improved predictive strength (R^2=0.659; RMSE=0.858; CCC=0.763; **Table 1**).

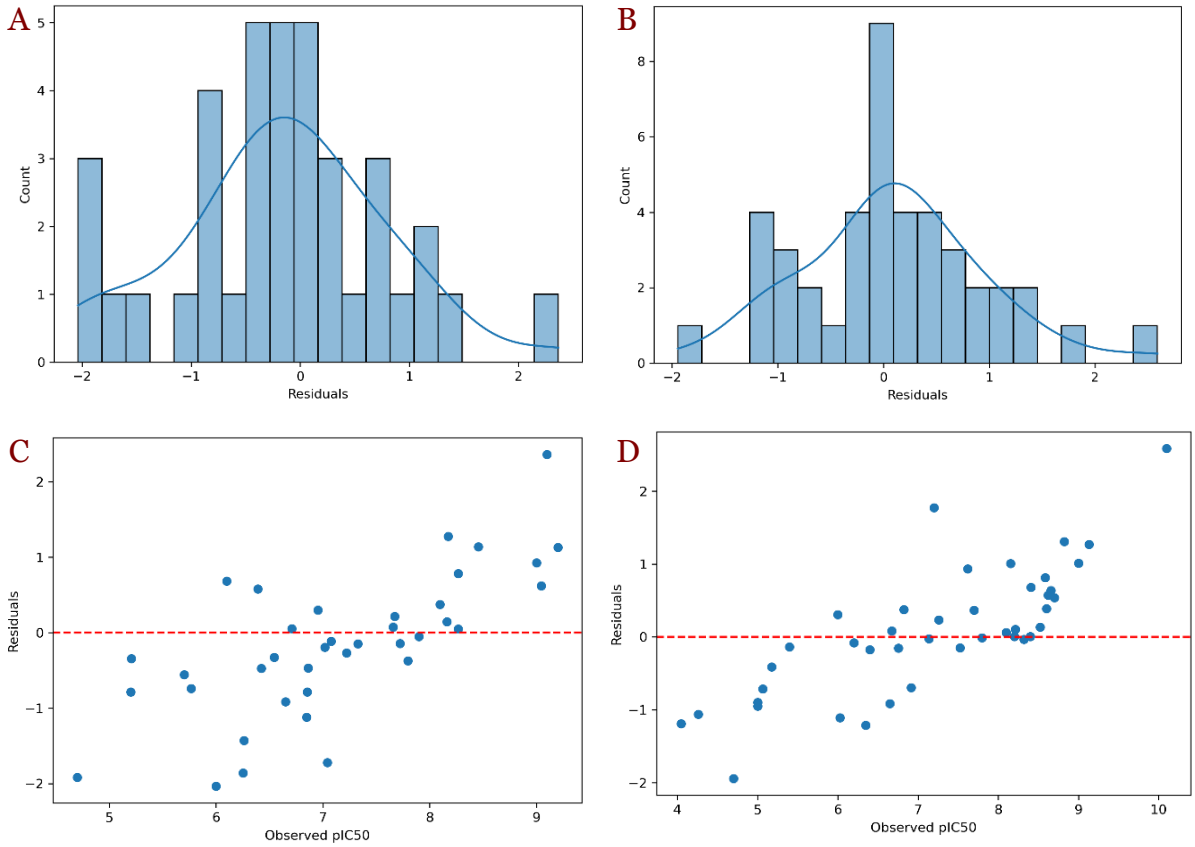

Figure 6. Residual analysis of the Random Forest regression model. (A) Histogram of residuals in the internal validation set shows a narrow, symmetric distribution. (B) Histogram of residuals in the external validation set indicates slight dispersion but remains centered. C) Scatter plot of residuals vs. observed pIC_{50} in the internal set shows no trend or bias. (D) External residuals exhibit minor outliers at the lower and upper ends but remain predominantly close to zero.

Applicability domain and chemical space analysis

The applicability domain and generalizability of the developed QSAR model based on the PCA and Williams plot are presented in **Figure 7**. PCA projection onto the first two principal components revealed a dense and overlapping distribution of training (blue) and test (red) compounds. The test set was broadly and evenly embedded within the chemical space occupied by the training set, indicating that predictions were made within a well-represented region of descriptor space.

No clustering of external compounds in peripheral or sparsely populated areas was observed, suggesting minimal risk of extrapolation beyond the model's learned domain. Moreover, the Williams plot showed that nearly all compounds fell within the conventional boundaries of applicability domain, where the standardized residuals were bounded between ±3. Only one compound exhibited high leverage, and none were classified as both high-leverage and high-residual outliers.

These findings confirm that the QSAR model operates within a well-defined and chemically meaningful domain, reinforcing its utility for screening novel NEP inhibitors that are structurally similar to those in the training dataset.

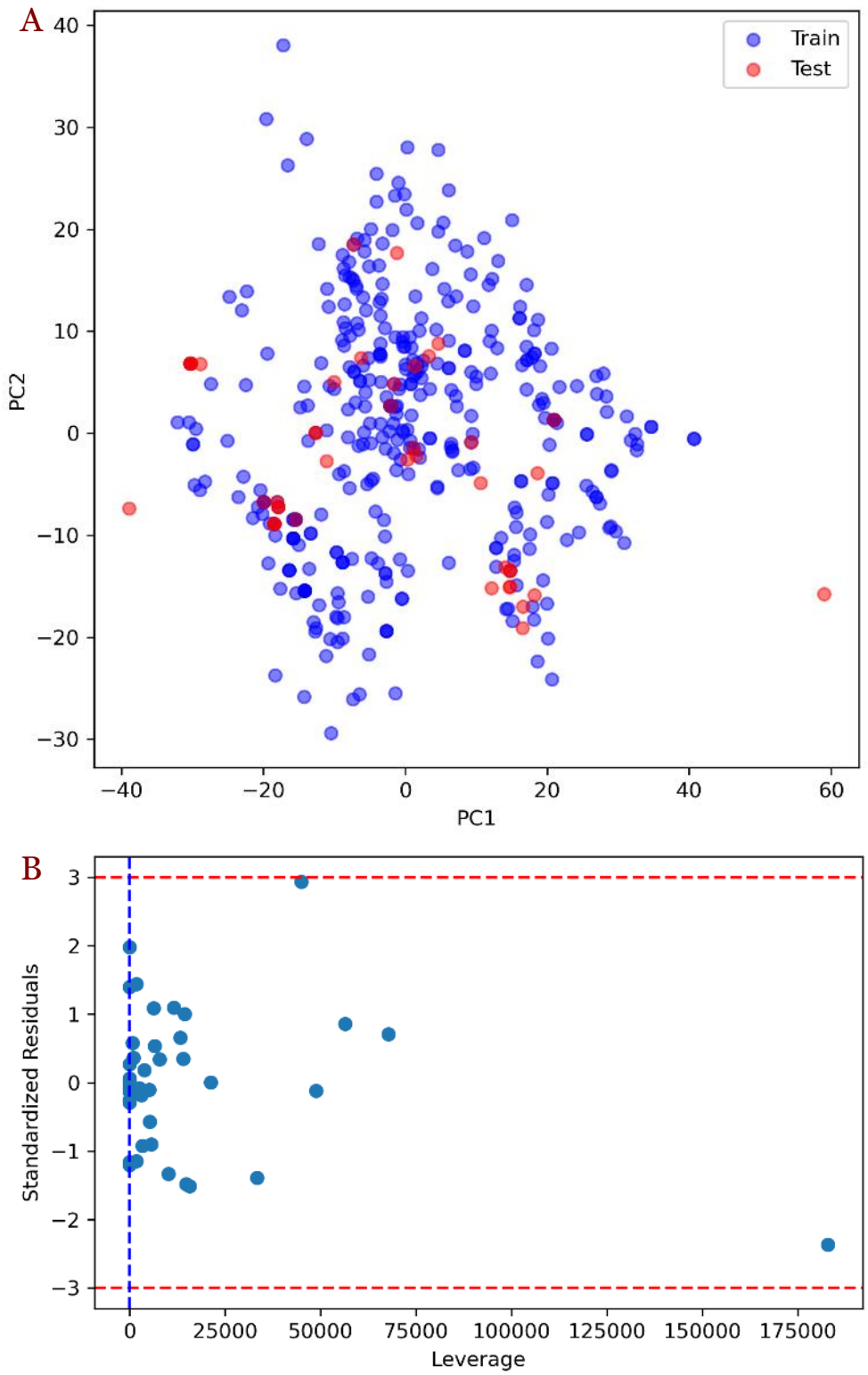

Figure 7. Principal component analysis (A) and Williams plots (V) depicting the applicability domain and chemical space visualization of the neprilysin QSAR model.

Discussion

This study developed an interpretable QSAR model for NEP inhibitors using 2D Mordred molecular descriptors and a Random Forest algorithm. While the classification model performed well (accuracy=0.953; F1 score=0.971), the regression model showed moderate performance, with an R^2 of 0.286 for internal validation and 0.659 for external validation (**Table 1**). The higher concordance correlation coefficient (CCC=0.763) and lower errors in the external set suggest better generalizability to unseen compounds. Residual analyses further supported the model's

external reliability, with residuals generally centered around zero and no major heteroscedastic patterns. Nonetheless, the relatively low internal R^2 and wide residual spread indicate that model stability may be affected by underlying data variability or descriptor redundancy. Comparison with prior QSAR efforts is limited by the sparse availability of NEP-specific models, as most literature focuses on multitarget profiling or combined metalloprotease inhibition [16]. Findings from a previous study support the relevance of structural considerations, reporting that pharmacophore-based screening and molecular dynamics simulations can successfully identify NEP ligands with high binding energy and favorable stability [23]. Another study further highlighted the structural complexity of NEP inhibition, reporting that sacubitril not only binds to the enzyme but also disrupts its tertiary structure, suggesting that inhibitory activity may involve conformational effects [24].

Feature importance analysis revealed that descriptors such as ATSC2p (centered on atomic electronegativity-weighted autocorrelation at lag 2), GATS2p (Geary autocorrelation of polarizability at lag 2), and SaasC (the sum of atom-type E-state indices for aromatic sp^2 carbons) were the most influential in predicting NEP inhibitory activity. These descriptors collectively emphasize the importance of molecular branching, electronic distribution, and atomic environment symmetry. These findings are more informative compared to previous structural study of NEP molecule [25]. ATSC2p and GATS2p highlight how the distribution of electronegativity and polarizability across bonded atoms plays a central role in modulating molecular reactivity and binding propensity. SaasC, on the other hand, implicates the contribution of π-systems and aromatic regions, which may engage in π–π stacking or hydrophobic interactions within the NEP binding cleft. This observation is highly consistent with NEP's well-characterized structural biology. As a zinc-dependent metalloprotease with a deep, hydrophobic, and flexible catalytic pocket, NEP accommodates a wide variety of peptides and inhibitors [2,26]. Whereas prior QSAR study did not incorporate structural insights, the present findings are contextualized within NEP's binding-site architecture, supported by crystallographic and enzymatic data [25]. Its substrate promiscuity arises from an ability to engage with diverse chemical scaffolds through electrostatic interactions, shape complementarity, and surface contact rather than strict lock-and-key binding [27]. Consequently, the significance of descriptors reflecting global molecular topology and electron distribution is mechanistically justified. These physicochemical properties enable ligands to navigate NEP's flexible binding groove and form stabilizing contacts, particularly via polarizable groups and extended π-systems [28,29]. Therefore, the prominence of ATSC2p, GATS2p, and SaasC in the present model not only underscores their statistical influence but also aligns with the biochemical determinants of ligand–NEP affinity observed in crystallographic and enzymatic studies.

The classification framework, based on a pIC_{50} threshold of 6, demonstrated high precision (1.000) and strong overall accuracy (0.953), indicating its suitability for binary decision-making in virtual screening. However, its utility should be limited to compounds within the model's learned chemical space. Initial applicability domain (AD) diagnostics revealed issues with leverage-based outlier detection due to unscaled input features, which were corrected through appropriate feature scaling. Nonetheless, future implementations should adopt more chemically meaningful AD strategies—such as distance-to-model or Mahalanobis distance metrics—to enhance outlier detection and ensure more reliable predictions. While the dataset was carefully curated for assay consistency (IC_{50} values in nM), its restricted chemical diversity may limit the model's ability to generalize across novel or structurally diverse scaffolds.

To improve model robustness and broader applicability, future work should aim to enrich the training set with chemically diverse NEP inhibitors and incorporate more expressive descriptors, such as 3D structural or quantum mechanical features that capture conformational flexibility and electronic properties. Although the Random Forest model performed well, exploring alternative algorithms—such as gradient boosting, support vector regression, or deep learning—may yield complementary or superior performance, particularly for challenging predictions involving outliers or rare chemotypes. Further enhancements should include Y-randomization tests to assess model stability, systematic descriptor pruning to reduce redundancy, and prospective validation against external libraries. Coupling the model with

molecular docking and ADMET profiling would also strengthen its translational relevance in early-stage drug discovery.

Conclusion

This study presents the first NEP-specific QSAR model that combines interpretable 2D descriptors with a dual-mode (regression and classification) framework, yielding high predictive performance. Key descriptors relate to structural complexity and polarizability, consistent with NEP's enzymatic characteristics. While the model is promising for early-stage virtual screening, its current applicability is limited by chemical diversity and the dimensionality of descriptors. Future refinements incorporating broader datasets, advanced descriptor classes, and comparative algorithms are essential to enhance generalizability and mechanistic insight.

Ethics approval

Not required.

Acknowledgments

None to declare.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Derived data supporting the findings of this study are available from the corresponding author on request.

Declaration of artificial intelligence use

Artificial intelligence tools (ChatGPT, OpenAI, San Francisco, CA, USA) were used to assist in language editing, refinement of sentence structure, and improvement of readability. The authors reviewed and verified all content to ensure accuracy and scientific integrity. AI was not used for data analysis, result generation, or interpretation.

How to cite

Albar N, Rampengan DDCH, Azhari S, *et al*. Random forest-based QSAR modeling for predicting the potency of neprilysin inhibitors using Mordred molecular descriptors. Narra X 2026; 4 (1): e242 - http://doi.org/10.52225/narrax.v4i1.242.

References

1. Bozkurt B, Nair AP, Misra A, *et al*. Neprilysin inhibitors in heart failure: The science, mechanism of action, clinical studies, and unanswered questions. JACC Basic Transl Sci 2022;8(1):88-105.
2. Zhang X, Hu C, Tian E, *et al*. Comprehensive review on neprilysin (NEP) inhibitors: Design, structure-activity relationships, and clinical applications. Front Pharmacol 2024;15:1501407.
3. Nakagawa H, Saito Y. Roles of natriuretic peptides and the significance of neprilysin in cardiovascular diseases. Biology 2022;11(7):1017.
4. Mittal S, Harikrishnan S, Gupta A, *et al*. Angiotensin receptor neprilysin inhibitor in chronic heart failure and comorbidity management: Indian consensus statement. Ther Adv Cardiovasc Dis 2024;18:17539447241301959.
5. Cho I-J, Kang S-M. Angiotensin receptor-neprilysin inhibitor in patients with heart failure and chronic kidney disease. Kidney Res Clin Pract 2021;40(4):555.
6. Ramadhan RN, Rampengan DD, Puling IM, *et al*. Efficacy of angiotensin receptor neprilysin inhibitor in hypertension management: A systematic review and meta-analysis of clinical trials. Narra J 2024;4(3):e1247.

7. Heriansyah T, Purnawarman A. ARNI vs ACE inhibitors in improving left ventricular geometry, diastolic function, and cardiac power output in HFrEF patients: A prospective cohort study among Acehnese, Indonesia. J Penelit Pendidik IPA 2025;11(4):48-56.
8. Rohde LE, Claggett BL, Wolsk E, *et al*. Cardiac and noncardiac disease burden and treatment effect of sacubitril/valsartan: Insights from a combined PARAGON-HF and PARADIGM-HF analysis. Cir Heart Fail 2021;14(3):e008052.
9. Lu H, Claggett BL, Packer M, *et al*. Race in heart failure: A pooled participant-level analysis of the global PARADIGM-HF and PARAGON-HF trials. JACC Heart Fail 2025;13(1):58-71.
10. Noviandy TR, Idroes GM, Tallei TE, *et al*. QSAR modeling for predicting beta-secretase 1 inhibitory activity in alzheimer's disease with support vector regression. Malacca Pharm 2024;2(2):79-85.
11. Das NR, Sharma T, Toropov AA, *et al*. Machine-learning technique, QSAR and molecular dynamics for hERG–drug interactions. J Biomol Struct Dyn 2023;41(23):13766-13791.
12. Singh B, Crasto M, Ravi K, *et al*. Pharmaceutical advances: Integrating artificial intelligence in QSAR, combinatorial and green chemistry practices. Intell Pharm 2024;2(5):598-608.
13. Tropsha A, Isayev O, Varnek A, *et al*. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. Nat Rev Drug Discov 2024;23(2):141-155.
14. Sanches IH, Mendonca SS, Alves VM, *et al*. QSAR models for predicting cardiac toxicity of drugs. In: Hong H, editor. QSAR in safety evaluation and risk assessment. Amsterdam: Elsevier; 2024.
15. Potterton A, Heifetz A, Townsend-Nicholson A. Predicting residence time of GPCR ligands with machine learning. In: Heifetz A, editor. Artificial intelligence in drug design. New York: Springer; 2021.
16. Shah SK, Chaple DR, Masand VH, *et al*. Multi-target in-silico modeling strategies to discover novel angiotensin converting enzyme and neprilysin dual inhibitors. Sci Rep 2024;14(1):15991.
17. Fang J, Wang L, Wu T, *et al*. Network pharmacology-based study on the mechanism of action for herbal medicines in Alzheimer treatment. J Ethnopharmacol 2017;196:281-292.
18. Idroes GM, Noviandy TR, Idroes GM, *et al*. Prognostication of differentiated thyroid cancer recurrence: An explainable machine learning approach. Narra X 2024;2(3):e183.
19. Priyanto D, Hairani H, Marzuki K, *et al*. Optimization of random forest for health data classification using PCA and K-Means SMOTE-ENN. Eng Technol Appl Sci Res 2025;15(5):27646-27652.
20. Moriwaki H, Tian Y-S, Kawashita N, *et al*. Mordred: A molecular descriptor calculator. J Cheminform 2018;10(1):4.
21. Iwuala BN, Aliyu AB, Ayo RG-O, *et al*. Quantitative structure-activity relationship (QSAR) studies of dihydroorotate dehydrogenase inhibitors of *Plasmodium falciparum* for the combat of malaria resistance using machine learning models classification approach. In Silico Res Biomed 2025;1:100019.
22. Chong A, Phua S-X, Xiao Y, *et al*. Establishing the foundations for a data-centric AI approach for virtual drug screening through a systematic assessment of the properties of chemical data. eLife 2024;13:RP97821.
23. Thakur S, Sinhari A, Gaikwad AB, *et al*. A structure-based pharmacophore modelling approach to identify and design new neprilysin (NEP) inhibitors: An in silico-based investigation. Arch Biochem Biophys 2024;756:110019.
24. Jovanović JĐ, Antonijević M, Vojinović R, *et al*. In silico study of inhibitory capacity of sacubitril/valsartan toward neprilysin and angiotensin receptor. RSC Adv 2022;12(46):29719-29726.
25. Pope D, Madura JD, Cascio M. β-amyloid and neprilysin computational studies identify critical residues implicated in binding specificity. J Chem Inf Model 2014;54(4):1157-1165.
26. Schiering N, D'Arcy A, Villard F, *et al*. Structure of neprilysin in complex with the active metabolite of sacubitril. Sci Rep 2016;6(1):27909.
27. Moss S, Subramanian V, Acharya KR. Crystal structure of peptide-bound neprilysin reveals key binding interactions. FEBS Lett 2020;594(2):327-336.
28. Sankhe R, Rathi E, Manandhar S, *et al*. Repurposing of existing FDA approved drugs for Neprilysin inhibition: An in-silico study. J Mol Struct 2021;1224:129073.
29. Zheng W, Tian E, Liu Z, *et al*. Small molecule angiotensin converting enzyme inhibitors: A medicinal chemistry perspective. Front Pharmacol 2022;13:968104.