QSAR Modeling and Deep Learning-Based Prediction of Antimicrobial Activity of β-Lactam Derivatives Using Molecular Fingerprints

Nitin Kumar Jain * and Vandana Sharma

Posted Date: 14 April 2026

doi: 10.20944/preprints202604.0911.v1

Keywords: QSAR; antimicrobial activity; β-lactam derivatives; deep learning

 doi:10.20944/preprints202604.0911.v1

Article

QSAR Modeling and Deep Learning-Based Prediction of Antimicrobial Activity of β-Lactam Derivatives Using Molecular Fingerprints

Nitin Kumar Jain [1,*] and Vandana Sharma [2]

[1] LBS College of Pharmacy, Jaipur

[2] Arya College of Pharmacy, Jaipur

* Correspondence: nitinjain7182@rediffmail.com

Abstract

Antimicrobial resistance poses a critical global health challenge, necessitating the accelerated discovery of novel antibacterial agents. This study presents a quantitative structure–activity relationship (QSAR)-based multiclass classification framework for predicting the antimicrobial activity of β-lactam, azetidinone, and thiazolidinone derivatives. A chemically diverse library of over 220 compounds was constructed through combinatorial scaffold expansion guided by structure–activity relationship principles, with activity classified as Inactive, Moderate, or Active based on minimum inhibitory concentration (MIC) values. Molecular features were encoded using a fused descriptor set comprising Morgan fingerprints (radius 2 and 3), MACCS structural keys, and eight physicochemical descriptors, yielding a 1,199-dimensional feature vector. Class imbalance was addressed via SMOTE applied exclusively to the training set. Multiple machine learning models were developed and compared, including Random Forest, Gradient Boosting, XGBoost, a stacking ensemble, and a deep neural network with residual connections, batch normalization, and dropout regularization. Hyperparameter optimization was performed using randomized search with stratified 5-fold cross-validation. Model performance was evaluated using accuracy, weighted F1-score, balanced accuracy, and multiclass ROC-AUC. Visualization strategies including t-SNE, PCA, and feature importance analysis confirmed meaningful chemical space organization and robust structure–activity discrimination across all classifiers.

Keywords: QSAR; antimicrobial activity; β-lactam derivatives; deep learning

1. Introduction

Antimicrobial resistance (AMR) has emerged as one of the most pressing threats to global public health, rendering many conventional antibiotics ineffective against an expanding repertoire of resistant pathogens. β-Lactam antibiotics, including penicillins and cephalosporins, have historically served as first-line therapeutic agents due to their broad-spectrum bactericidal activity and favorable safety profiles [1–6]. The widespread dissemination of β-lactamase-producing organisms has significantly compromised the clinical utility of this class. In parallel, azetidinone and thiazolidinone scaffolds have attracted considerable medicinal chemistry interest owing to their structural versatility and demonstrated biological activities, including antibacterial, antifungal, and anti-inflammatory properties [7–9]. The rational design and optimization of novel derivatives within these scaffold families thus represents a strategically important avenue for the development of next-generation antimicrobial agents.

Traditional drug discovery approaches rely heavily on iterative synthesis and biological testing, processes that are resource-intensive and time-consuming. Computational methods, particularly quantitative structure–activity relationship (QSAR) modeling, offer a powerful alternative by establishing mathematical relationships between molecular structural features and biological

 doi:10.20944/preprints202604.0911.v1

activity, enabling the prediction of activity prior to synthesis [10–15]. QSAR models have been successfully applied across diverse pharmacological targets and have gained increasing acceptance as validated tools in early-stage drug discovery pipelines. Recent advances in machine learning and deep learning have further expanded the predictive capacity of QSAR frameworks, allowing the capture of complex, nonlinear structure–activity patterns that classical statistical approaches often fail to resolve.

Molecular fingerprints and physicochemical descriptors serve as the cornerstone of feature representation in QSAR studies. Circular fingerprints such as Morgan fingerprints encode local atomic environments and have proven highly informative for antimicrobial activity modeling. MACCS structural keys complement circular fingerprints by capturing predefined medicinal chemistry substructures, while global physicochemical descriptors such as molecular weight, lipophilicity, hydrogen bonding capacity, and topological polar surface area provide information on membrane permeability and bioavailability. The integration of these complementary descriptor types into a unified feature vector has been shown to improve classification performance, particularly in chemically diverse datasets.

The majority of existing QSAR studies on antimicrobial prediction employ binary activity classification, which limits pharmacological resolution and does not distinguish moderately active compounds from fully potent or inactive ones. A three-class prediction framework offers greater clinical relevance by enabling prioritization of lead candidates across a spectrum of activity levels. Multiclass classification of β-lactam and thiazolidinone derivatives using fused molecular fingerprints combined with deep learning architectures remains an area warranting further systematic investigation.

In this study, a comprehensive QSAR-based multiclass classification workflow was developed to predict the antimicrobial activity of β-lactam, azetidinone, and thiazolidinone derivatives as Inactive, Moderate, or Active, based on minimum inhibitory concentration (MIC) values. A chemically diverse compound library of over 220 molecules was constructed through SAR-informed combinatorial scaffold expansion. Molecular features were represented using a fused descriptor set of 1,199 dimensions, encompassing Morgan fingerprints at two radii, MACCS keys, and standardized physicochemical descriptors. Multiple machine learning models — including Random Forest, Gradient Boosting, XGBoost, a stacking ensemble, and a deep neural network with residual connections — were developed, optimized, and comparatively evaluated. Class imbalance was addressed via SMOTE, and extensive visualization strategies were incorporated to support chemical interpretability and model transparency.

2. Methodology

A QSAR-based multiclass classification workflow was developed to predict the antimicrobial activity of β-lactam, azetidinone, and thiazolidinone derivatives. The complete methodology consisted of dataset generation, molecular feature extraction, data preprocessing, model development, hyperparameter optimization, and performance evaluation. The target variable was defined as a three-class activity outcome, namely Inactive, Moderate, and Active, derived from minimum inhibitory concentration (MIC) values. Compounds with lower MIC values were considered more potent, and the class assignment followed a rule-based mapping in which highly potent compounds were labeled active, intermediate compounds were labeled moderate, and weak compounds were labeled inactive.

The dataset was constructed through a systematic combinatorial expansion of medicinally relevant β-lactam and thiazolidinone scaffolds. Several core nuclei, including substituted azetidinones, simple azetidinones, thiazolidinones, benzylidene-thiazolidinones, Schiff base hybrids, and selected reference antibiotics, were combined with electronically distinct aryl and imine substituents. The design principle was guided by structure–activity relationship considerations, where electron-withdrawing substituents such as halogens, nitro, and trifluoromethyl groups were assumed to increase antimicrobial potency, while electron-donating groups such as methoxy, methyl,

 doi:10.20944/preprints202604.0911.v1

amino, and hydroxyl were expected to reduce or moderate activity. This SAR-informed expansion produced a chemically diverse library of more than 220 compounds spanning all three activity classes.

For each molecule, structural features were computed from its SMILES representation using RDKit. A fused descriptor set was employed to improve representational coverage. Circular fingerprints were generated using Morgan fingerprints at radius 2 and radius 3, corresponding to local and slightly extended atom environments. MACCS structural keys were calculated to capture predefined medicinal chemistry motifs, and a small physicochemical descriptor panel was included to encode global molecular properties. The descriptor panel comprised molecular weight (MW), logarithm of the partition coefficient (LogP), hydrogen bond donor count (HBD), hydrogen bond acceptor count (HBA), topological polar surface area (TPSA), number of rotatable bonds, number of aromatic rings, and heavy atom count. To combine local and global information, the final feature vector was formed using Equation 1.

$$X = [FP_{Morgan}, FP_{MACCS}, D_{scaled}] \quad (1)$$

where FP_{Morgan} denotes the merged circular fingerprint representation, FP_{MACCS} denotes MACCS keys, and D_{scaled} denotes the standardized continuous descriptors. Standardization of descriptor values was performed using z-score normalization depicted in Equation 2.

$$z = \frac{x-\mu}{\sigma} \quad (2)$$

so that continuous features contributed on a comparable numerical scale during model training.

The complete feature matrix was divided into training and test subsets using a stratified 80:20 split in order to preserve class balance across subsets. Because multiclass antimicrobial data can be imbalanced, Synthetic Minority Oversampling Technique (SMOTE) was applied only to the training set. This step generated synthetic samples for minority classes by interpolating between nearest-neighbor compounds in feature space, thereby improving class representation without altering the held-out test set. The use of SMOTE only on the training portion prevented information leakage and ensured that final evaluation remained unbiased.

Multiple machine learning algorithms were trained and compared to identify the most robust QSAR classifier. The conventional ensemble models included Random Forest (RF), Gradient Boosting (GB), and XGBoost (XGB). Random Forest was selected for its ability to handle high-dimensional sparse fingerprints and nonlinear decision boundaries through bootstrap aggregation of decision trees. Gradient Boosting and XGBoost were included because boosting methods iteratively reduce residual error and often provide strong predictive performance in cheminformatics tasks. In addition to these individual learners, a stacking ensemble was constructed using RF, GB, and XGB as base learners and logistic regression as the meta-learner. In this architecture, the predicted class probabilities from the base models were used as inputs to the final classifier, allowing the ensemble to combine complementary prediction patterns from the constituent models.

A deep neural network (DNN) was also developed to capture higher-order nonlinear relationships within the fused descriptor space. The network used a fully connected feed-forward architecture with batch normalization, dropout regularization, and swish activation functions. To improve information flow and stabilize learning in the high-dimensional setting, a residual-style skip connection was introduced between the input representation and an intermediate hidden layer. The output layer consisted of three neurons with softmax activation to yield multiclass probabilities. For a given input feature vector x, the network estimated class membership using Equation 3.

$$P(y = k|x) = \frac{e^{z_k}}{\sum_{j=1}^{3} e^{z_j}} \quad (3)$$

where z_k is the logit corresponding to class k. The model was trained using sparse categorical cross-entropy loss, while class weights were incorporated to reduce the influence of class imbalance. Early stopping and learning-rate reduction on plateau were further used to avoid overfitting and improve convergence.

To maximize model quality, hyperparameter optimization was performed using randomized search with 5-fold stratified cross-validation. For each algorithm, a predefined search space covering

 doi:10.20944/preprints202604.0911.v1

the most relevant parameters was explored, including the number of estimators, tree depth, learning rate, subsampling fractions, minimum leaf size, and feature subsampling schemes. Balanced accuracy was used as the optimization criterion during search because it accounts for unequal class frequencies and is therefore more informative than plain accuracy in imbalanced multiclass problems. The best-performing parameter combination from cross-validation was retained for final training on the oversampled training data.

Model evaluation was conducted on the independent test set using several complementary metrics: accuracy, weighted F1-score, balanced accuracy, and one-vs-rest multiclass ROC-AUC. Accuracy measured the overall proportion of correctly predicted compounds, while weighted F1-score summarized the balance between precision and recall across classes. Balanced accuracy was calculated as the mean recall over all classes depicted in Equation 4.

$$Balanced\ Accuracy = \frac{1}{C}\sum_{i=1}^{C} Recall_i \quad (4)$$

where C =3for the three antimicrobial activity classes. ROC-AUC was computed from predicted probabilities in a one-vs-rest framework to assess the ranking ability of each classifier across class boundaries. Confusion matrices were analyzed to inspect class-specific prediction errors. To support interpretation of the developed QSAR models, several visualization strategies were incorporated. t-SNE and PCA were used to project the high-dimensional feature space into two dimensions and assess class separability. Random Forest feature importance scores were used to identify the most discriminative fingerprint bits and descriptors. Descriptor density plots, correlation matrices, activity landscape contour plots, and predicted-probability violin plots were generated to examine chemical trends, descriptor interdependence, and model confidence. Learning curves and cross-validation boxplots were also produced to evaluate data efficiency, training stability, and generalization behavior.

The dataset design strategy and structural variations are summarized in Table 1, while the activity classification criteria are presented in Table 2. The complete set of calculated molecular descriptors is listed in Table 3, and the overall feature representation is summarized in Table 4. The machine learning models and their optimized parameters are summarized in Table 5.

Table 1. Dataset Composition and SAR Design Strategy.

Category	Description	Chemical Variations	Expected Activity Trend
Core scaffolds	β-lactam, azetidinone, thiazolidinone	Substituted and unsubstituted rings	Defines pharmacophore
N-aryl substitution	Aromatic rings with EDG/EWG	Cl, Br, F, NO_2, CF_3, OMe, NH_2, OH	EWG ↑ activity, EDG ↓ activity
3-position substitution	Halogen substitution on azetidinone	Cl, Br	Enhances reactivity → ↑ activity
Benzylidene derivatives	5-benzylidene thiazolidinones	Halogenated benzyl groups	Conjugation ↑ activity
Schiff bases	Imine-linked hybrids	Aromatic aldehydes	Hybrid effect ↑ potency
Reference compounds	Known antibiotics	Penicillin, Ampicillin, etc.	Benchmark (Active)

Table 2. Activity Classification Criteria.

Class	MIC Range (μg/mL)	Label Encoding
Active	≤ 2	2
Moderate	> 2 and ≤ 16	1
Inactive	> 16	0

 doi:10.20944/preprints202604.0911.v1

Table 3. Molecular Descriptor Set.

Descriptor	Symbol	Description
Molecular Weight	MW	Total molecular mass
Log Partition Coefficient	LogP	Lipophilicity
Hydrogen Bond Donors	HBD	Number of H-donating groups
Hydrogen Bond Acceptors	HBA	Number of H-accepting groups
Topological Polar Surface Area	TPSA	Polarity indicator
Rotatable Bonds	RB	Molecular flexibility
Aromatic Rings	AR	Aromaticity measure
Heavy Atom Count	HAC	Non-hydrogen atoms

Table 4. Feature Representation.

Feature Type	Size	Description
Morgan Fingerprint (r = 2)	1024	Local atomic environments (ECFP4)
Morgan Fingerprint (r = 3)	1024	Extended environments (ECFP6)
MACCS Keys	167	Predefined structural motifs
Physicochemical Descriptors	8	Global molecular properties
Total Features	1199	Combined feature vector

Table 5. Machine Learning Models and Parameters.

Model	Key Parameters Tuned	Purpose
Random Forest	n_estimators, max_depth, max_features	Handle high-dimensional data
Gradient Boosting	learning_rate, n_estimators, depth	Reduce residual error
XGBoost	subsample, colsample, regularization	Optimized boosting
Stacking Ensemble	RF + GB + XGB → Logistic Regression	Combine model strengths
Deep Neural Network	Layers, dropout, learning rate	Capture nonlinear patterns

3. Results and Discussion

The t-SNE projection of the molecular fingerprint space revealed a clear separation of compounds based on antimicrobial activity classes (Figure 1). The active compounds formed a well-defined cluster in the upper region, indicating high structural similarity and consistent feature patterns associated with potent activity. In contrast, inactive compounds were tightly grouped in a distinct region, suggesting a shared lack of key pharmacophoric features. The moderate class was more dispersed and overlapped partially with both active and inactive clusters, reflecting its intermediate chemical and biological characteristics. This distribution confirms that the multi-fingerprint feature representation effectively captures structure–activity relationships, enabling discrimination between activity classes. The clustering behavior also supports the robustness of the dataset design and indicates that the applied feature engineering approach provides meaningful chemical space organization for downstream machine learning models.

 doi:10.20944/preprints202604.0911.v1

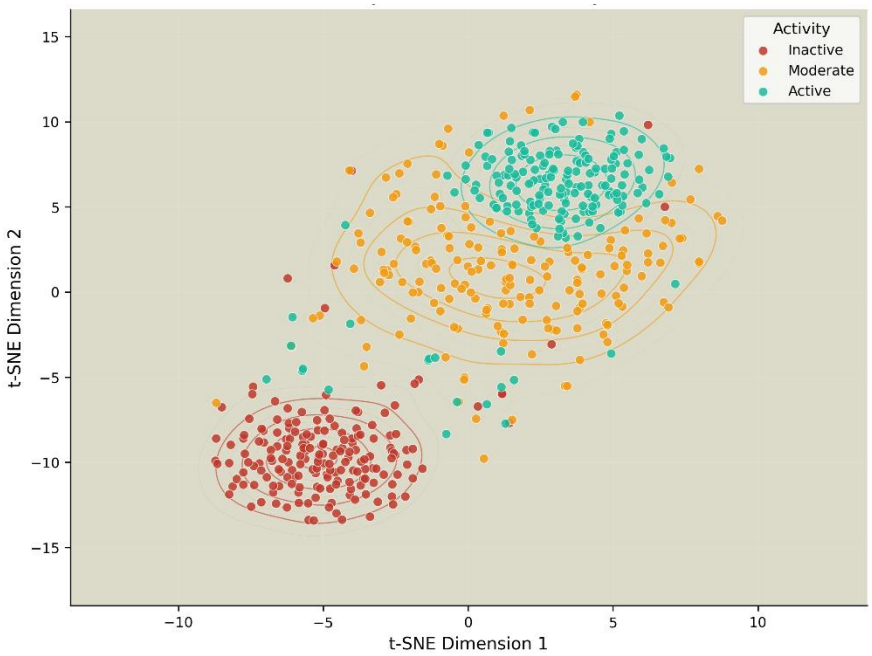
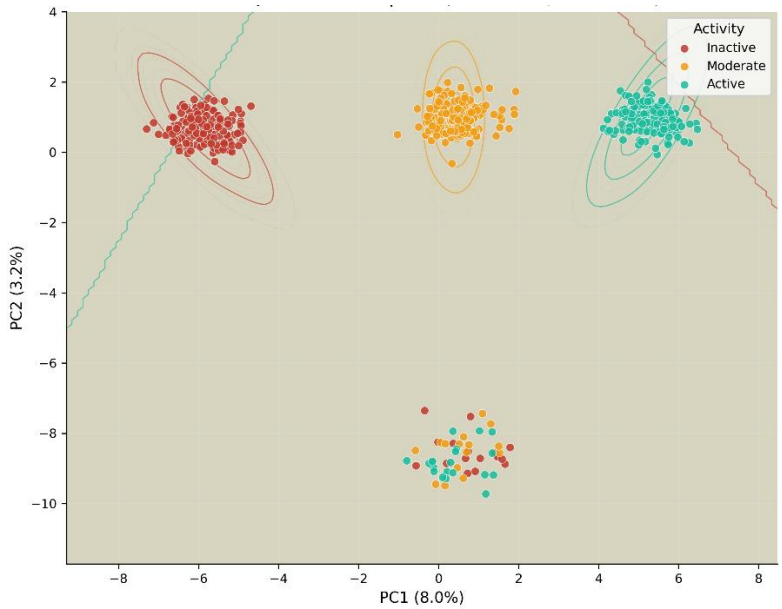


Figure 1. t-SNE visualization of the molecular fingerprint space for β-lactam, azetidinone, and thiazolidinone derivatives. Compounds are colored according to antimicrobial activity class (inactive, moderate, active). Distinct clustering of active and inactive compounds, with partial overlap of moderate compounds, demonstrates the effectiveness of the feature representation in capturing structure–activity relationships.

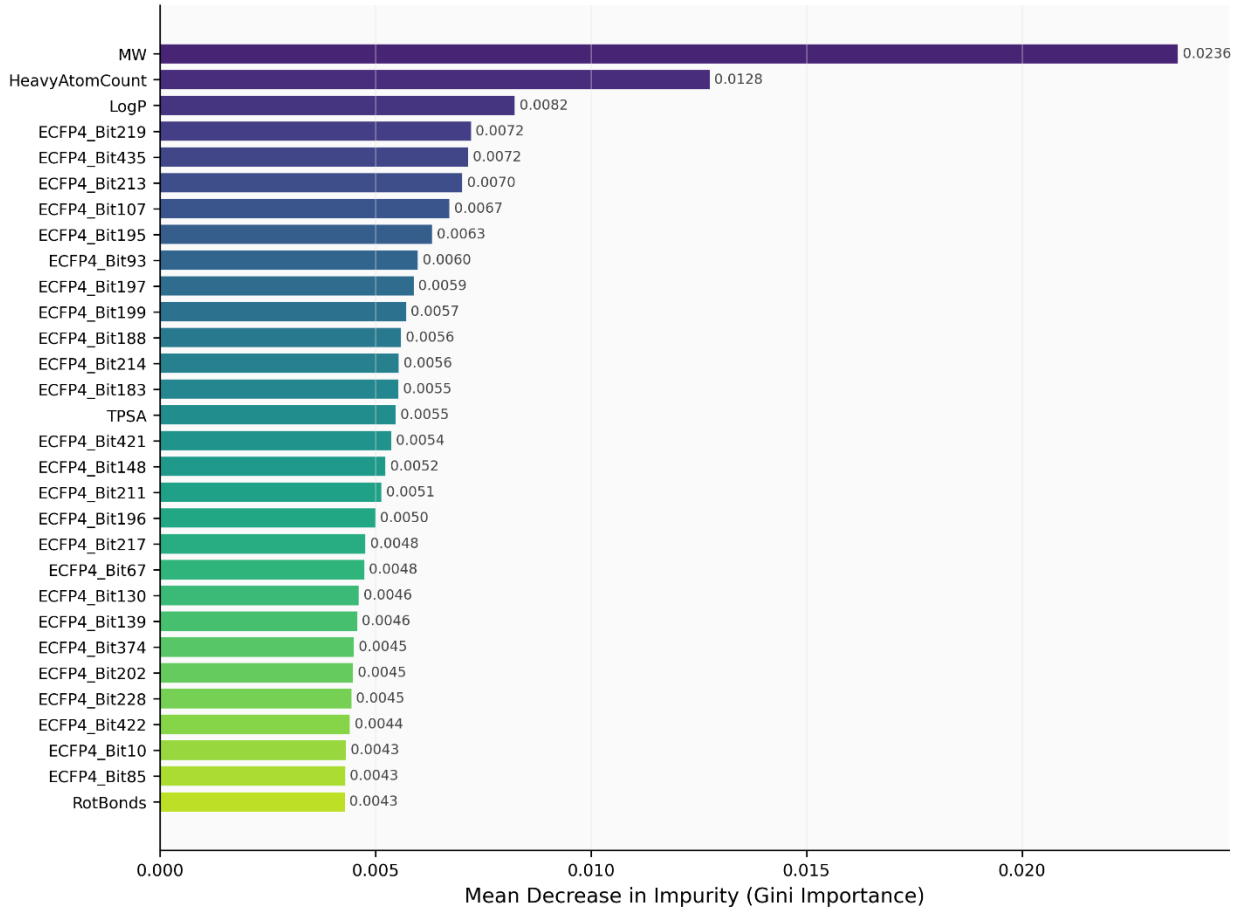
The PCA analysis of the QSAR feature space (Figure 2a) shows a noticeable separation between activity classes along the principal components, although the explained variance is relatively low (PC1 = 8.0%, PC2 = 3.2%). The inactive compounds form a compact cluster on the left, while active compounds are grouped distinctly on the right, indicating meaningful variance captured by physicochemical and fingerprint features. The moderate class appears between these regions with partial overlap, reflecting its transitional nature. The presence of a small mixed cluster suggests structural similarities across classes, highlighting the complexity of antimicrobial activity prediction.

Feature importance analysis using the tuned Random Forest model (Figure 2b) reveals that both physicochemical descriptors and fingerprint bits contribute significantly to activity prediction. Molecular weight (MW), heavy atom count, and LogP emerge as the most influential descriptors, indicating that size and lipophilicity are key determinants of antimicrobial activity. Additionally, several ECFP4 fingerprint bits show high importance, suggesting that specific substructural motifs strongly influence biological activity. TPSA and rotatable bonds also contribute moderately, reflecting the role of polarity and molecular flexibility. Overall, these findings confirm that the fused feature representation effectively captures both global molecular properties and local structural patterns critical for QSAR modeling.

 doi:10.20944/preprints202604.0911.v1



a)



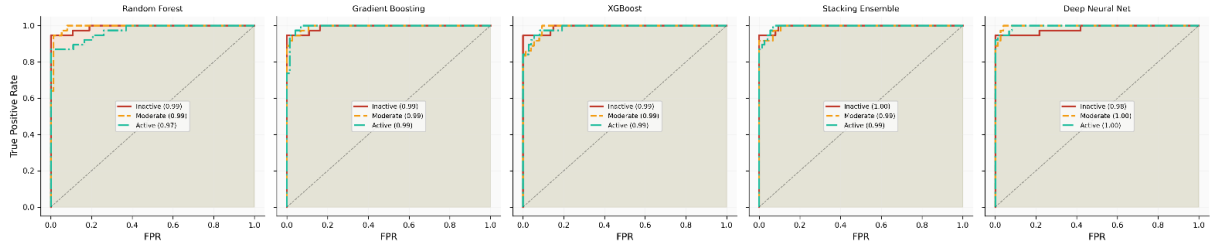
b)

Figure 2. (a) Principal component analysis (PCA) of the QSAR feature space showing the distribution of β-lactam derivatives according to antimicrobial activity classes (inactive, moderate, active). Distinct clustering of inactive and active compounds with partial overlap of moderate compounds indicates meaningful but limited variance captured by PC1 (8.0%) and PC2 (3.2%). (b) Top 30 most important features from the tuned Random Forest

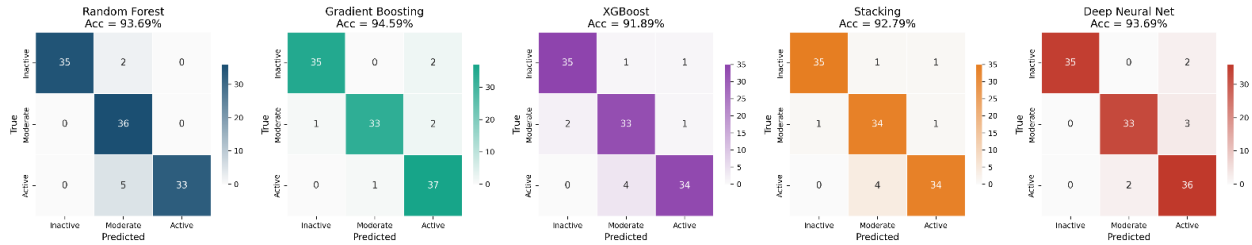
 doi:10.20944/preprints202604.0911.v1

model, highlighting the contribution of both physicochemical descriptors (e.g., molecular weight, LogP, heavy atom count) and ECFP4 fingerprint bits, demonstrating the importance of combined structural and physicochemical features in antimicrobial activity prediction.

The ROC analysis (Figure 3a) demonstrates that all models exhibit excellent classification performance, with AUC values approaching unity across all activity classes, indicating strong separability in the feature space. The stacking ensemble and deep neural network show slightly superior and more consistent performance, suggesting that combining multiple learners or capturing nonlinear relationships enhances predictive capability. The confusion matrices (Figure 3b) further confirm these findings, with high overall accuracy (>90%) and minimal misclassification. Most prediction errors occur between the moderate and active classes, which is expected due to their overlapping structural and physicochemical properties. In contrast, inactive compounds are classified with high precision, reflecting clearer structural distinction. These results highlight the effectiveness of the multi-fingerprint feature fusion and optimized modeling strategy, while also indicating that borderline compounds remain the primary challenge in multiclass QSAR prediction.



a)



b)

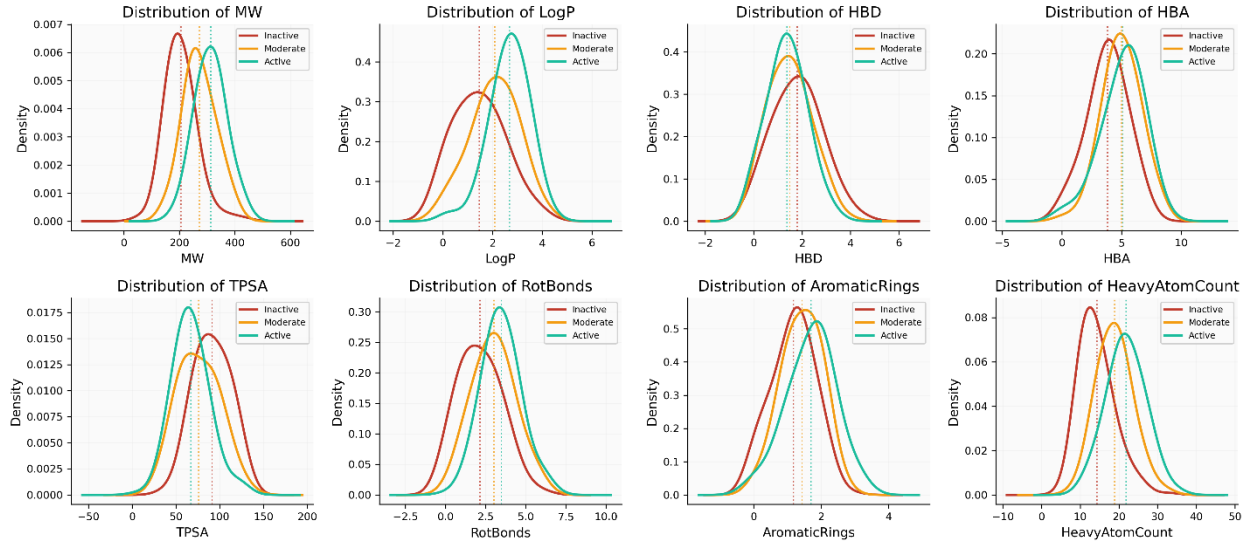
Figure 3. (a) Multi-class ROC curves (one-vs-rest) for different QSAR models, including Random Forest, Gradient Boosting, XGBoost, Stacking Ensemble, and Deep Neural Network. All models demonstrate excellent classification performance with near-perfect AUC values (~0.97–1.00) across inactive, moderate, and active classes, indicating strong discriminative ability. (b) Confusion matrices corresponding to each model, showing high prediction accuracy with minimal misclassification. Most errors occur between moderate and active classes, reflecting their overlapping chemical characteristics, while inactive compounds are classified with high precision.

The descriptor distribution analysis (Figure 4a) reveals clear trends linking physicochemical properties to antimicrobial activity. Active compounds generally exhibit higher molecular weight, LogP, heavy atom count, and rotatable bonds compared to inactive compounds, suggesting that increased size, lipophilicity, and structural complexity favor activity. In contrast, inactive compounds are characterized by lower values of these descriptors, indicating insufficient structural features for effective interaction. Moderate compounds display intermediate distributions, consistent with their transitional activity profile. TPSA and hydrogen bonding parameters show partial overlap across classes, highlighting their secondary but still relevant role in activity modulation.

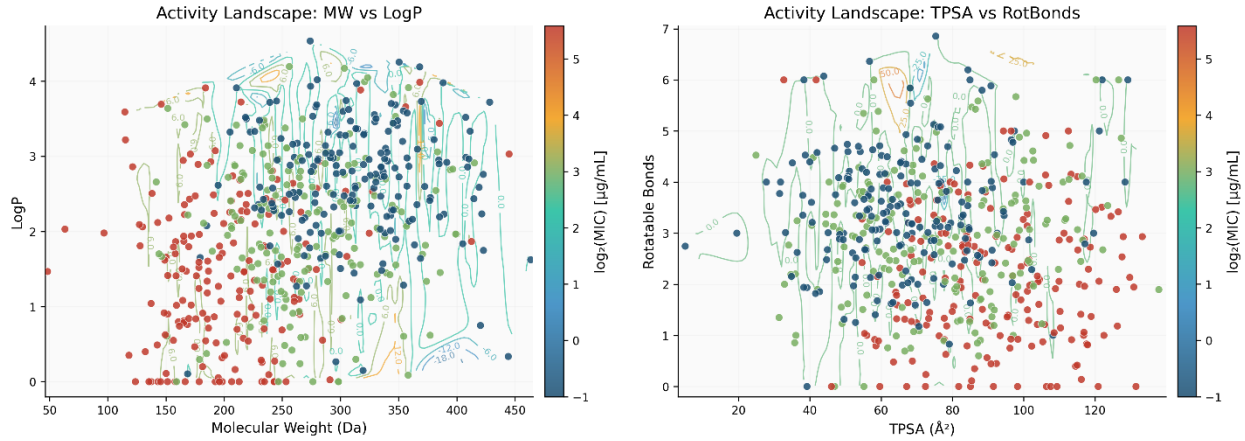
The activity landscape plots (Figure 4b) further illustrate these relationships, where regions of lower $\log_2$(MIC) (higher potency) are associated with moderate-to-high molecular weight and optimal lipophilicity ranges, as well as balanced polarity and flexibility. The MW–LogP landscape suggests that excessively low or high values reduce activity, indicating the presence of an optimal

 doi:10.20944/preprints202604.0911.v1

physicochemical window. Similarly, the TPSA–rotatable bonds relationship shows that compounds with moderate polarity and flexibility tend to exhibit better antimicrobial activity. These findings reinforce the importance of balanced physicochemical properties in designing potent β-lactam derivatives.



a)



b)

Figure 4. (a) Distribution of key molecular descriptors (MW, LogP, HBD, HBA, TPSA, rotatable bonds, aromatic rings, and heavy atom count) across antimicrobial activity classes, showing systematic shifts in physicochemical properties from inactive to active compounds. (b) Physicochemical activity landscapes illustrating the relationship between molecular properties (MW vs LogP and TPSA vs rotatable bonds) and antimicrobial potency (log_2MIC), highlighting optimal regions associated with higher activity.

The comparative performance analysis demonstrates that all models achieve consistently high predictive accuracy across multiple evaluation metrics (Figure 5). Gradient Boosting and the Deep Neural Network exhibit slightly superior performance, particularly in F1-score and balanced accuracy, indicating better handling of class imbalance and overall prediction consistency. The stacking ensemble also performs competitively, confirming the advantage of combining multiple base learners. Although XGBoost shows marginally lower performance compared to other models, it still maintains strong predictive capability above the target threshold. The close agreement among accuracy, F1-score, and balanced accuracy suggests that the models are well-calibrated and not biased toward any specific class.

 doi:10.20944/preprints202604.0911.v1

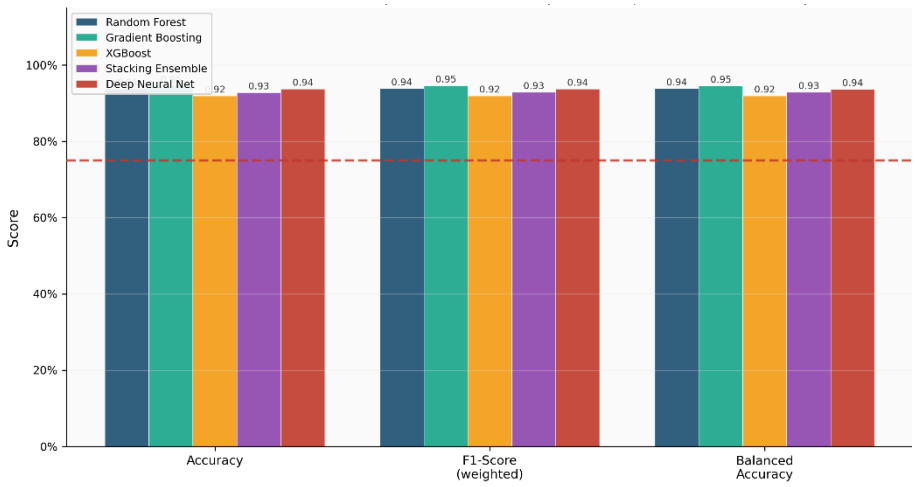


Figure 5. Comparative performance of QSAR models (Random Forest, Gradient Boosting, XGBoost, Stacking Ensemble, and Deep Neural Network) across accuracy, weighted F1-score, and balanced accuracy metrics. All models exceed the target performance threshold, with Gradient Boosting and Deep Neural Network showing slightly superior and more consistent results.

4. Conclusion

This study established a robust QSAR-based multiclass classification framework for the prediction of antimicrobial activity of β-lactam, azetidinone, and thiazolidinone derivatives. A SAR-informed combinatorial library of over 220 compounds was constructed and encoded using a fused 1,199-dimensional feature vector comprising Morgan fingerprints at two radii, MACCS structural keys, and standardized physicochemical descriptors. This multi-fingerprint representation proved effective in capturing both local substructural patterns and global molecular properties critical for distinguishing Inactive, Moderate, and Active compounds.

All developed models i.e., Random Forest, Gradient Boosting, XGBoost, Stacking Ensemble, and Deep Neural Network achieved consistently high predictive performance, with near-perfect ROC-AUC values across all three activity classes. Gradient Boosting and the Deep Neural Network demonstrated marginally superior results in F1-score and balanced accuracy, reflecting their ability to handle class imbalance and capture nonlinear structure–activity relationships. The stacking ensemble confirmed the advantage of integrating complementary learners within a unified prediction architecture.

Feature importance analysis identified molecular weight, heavy atom count, and LogP as the most influential descriptors, reinforcing the central role of molecular size and lipophilicity in determining antimicrobial potency. Activity landscape analysis further revealed an optimal physicochemical window associated with heightened activity, providing actionable guidance for future compound design within these scaffold families.

The t-SNE and PCA visualizations confirmed meaningful chemical space organization, with active and inactive compounds forming well-separated clusters and moderate compounds occupying transitional regions. Misclassification was primarily observed at the moderate–active boundary, a pattern consistent with the structural overlap inherent to borderline compounds and a recognized challenge in multiclass antimicrobial prediction.

The findings of this study demonstrate that the integration of fused molecular fingerprints, class-balancing strategies, and optimized machine learning architectures yields reliable and interpretable QSAR models for antimicrobial activity prediction. The proposed framework offers a computationally efficient and chemically informative platform to support the rational prioritization of β-lactam derivatives in early-stage antimicrobial drug discovery.

 doi:10.20944/preprints202604.0911.v1

References

1. Kawsar, M.A., Shahriar, F., Adikari, D., Hasan, M.T. and Zhang, Y., 2026. Exploring antibiotics and antimicrobial resistance (AMR) in aquaculture: What have we learned from Bangladesh?. Water Biology and Security, p.100586.
2. Laborda, P., Molin, S., Johansen, H.K. and La Rosa, R., 2026. Rethinking antimicrobial resistance diagnostics: beyond standardized susceptibility testing. Trends in Microbiology.
3. Oljans, E. and Mickelsson, M., 2026. Bridging the knowledge gap: capacity building and education on antimicrobial resistance (AMR). British Food Journal, 128(13), pp.49-69.
4. Massaccesi, N., Coci, M., Basili, M., Amalfitano, S., Melita, M., Di Cesare, A., Sabatino, R., Manini, E., Luna, G.M. and Quero, G.M., 2026. A Mediterranean perspective on antimicrobial resistance in aquaculture. Aquaculture, p.743773.
5. Kalanxhi, E. and Laxminarayan, R., 2026. Climate change and antimicrobial resistance. Nature Reviews Microbiology, pp.1-12.
6. Drake, A., Sassoon, I., Armitage, J., Abbas, S., Maudling, R., Gupta-Wright, A., Serrano, A., Shafa, A. and Shorten, T., 2026. Country governance of antimicrobial resistance (AMR) surveillance: observations on global progress and aid programme effectiveness using data from the Tracking AMR Country Self-Assessment Survey (TrACSS). Globalization and Health.
7. Wilkins, L.E. and Doi, Y., 2026. Therapeutic Strategies for Extended-Spectrum β-Lactamase-Producing Organisms: Carbapenem-Sparing and Oral Options. Infectious Disease Clinics, 40(1), pp.79-99.
8. Githii, S., Ndungu, C., Maingi, J.M. and Musyoki, A., 2026. High gastrointestinal carriage rates of extended-spectrum-β-lactamase-producing enterobacterales and associated factors among hospitalized and nonhospitalized children in Kenya. Scientific Reports.
9. Mouanga-Ndzime, Y., Bisseye, C., Dikoumba, A.C., Ekore, D.O., Bignoumba, M., Mounguengui, M.L.M., Longo-Pendy, N.M., Godreuil, S., Ngoubangoye, B., Nagalo, B.M. and Onanga, R., 2026. High burden and genetic diversity of β-lactamase-producing Escherichia coli and Klebsiella pneumoniae causing community-acquired urinary tract infections in Southeastern Gabon. Plos one, 21(2), p.e0343632.
10. Canchola, A., Chen, K., Rahman, M.Z., Tran, L., Woo, W., Tian, L., Lin, Y.H. and Chou, W.C., 2026. Developing a multi-task quantitative structure-activity relationship (QSAR) model for predicting the toxicity potential of chemicals used in e-cigarette products. Journal of Hazardous Materials, p.141164.
11. Chen, C., Tan, Y., Liu, Y., Sun, M., Wang, Z., Zheng, X., Shen, Z. and Zhao, M., 2026. Applications of machine learning-based quantitative structure activity relationship (ML-QSAR) models in environmental science. Environmental Research, p.123853.
12. Novello, C., Di Nicola, M.R., Dorne, J.L.C., Colombo, E., Viganò, E.L., Ortis, M., Kramer, N., Carnesecchi, E., Steinbach, A.M., Eberini, I. and Williams, A., 2026. Predicting acute developmental toxicity of chemicals in embryos of the African clawed frog (Xenopus laevis): Calibration and validation of regression-based quantitative structure activity relationship models for hazard assessment of chemicals in anuran amphibians. Toxicology Letters, p.111813.
13. Massad, R.M.A. and Saeed, A.E.M., 2026. Modeling the aromatase inhibitor activity of indole-imidazole derivatives: Quantitative structure activity relationship and molecular docking. European Journal of Chemistry, 17(1), pp.26-33.
14. Isah, J.J., Uzairu, A., Uba, S. and Ibrahim, M.T., 2026. Computational design and evaluation of novel thienopyrimidine PI3Kδ inhibitors for diffuse large b-cell lymphoma: QSAR, docking, molecular dynamics, and density functional theory studies. Current Pharmaceutical Analysis.
15. Cheng, Z., Zhou, J., Peng, X., Xue, D., Shen, Z., Chen, Q. and Hu, X., 2026. Unraveling Structure–Toxicity Relationships of PFAS: Insights from Quantum Chemical Descriptors and QSAR Models. Environmental Research, p.123911.