Original Article

Check for updates

Available online at
www.heca-analitika.com/malacca_pharmaceutics

Malacca Pharmaceutics

Vol. 3, No. 2, 2025

Interpretable Machine Learning QSAR Models for Classification and Screening of VEGFR-2 Inhibitors in Anticancer Drug Discovery

Teuku Rizky Noviandy [1] and Rinaldi Idroes [2,*]

[1] Department of Information Systems, Faculty of Engineering, Universitas Abulyatama, Aceh Besar 23372, Indonesia; rizky_si@abulyatama.ac.id (T.R.N.)
[2] School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 29 June 2025
Revised 10 September 2025
Accepted 22 September 2025
Available Online 28 September 2025

Keywords:
VEGFR-2 inhibitors
Machine learning
QSAR modeling
Drug discovery

Abstract

Cancer remains a major global health burden, with angiogenesis playing a central role in tumor growth and progression. Vascular Endothelial Growth Factor Receptor-2 (VEGFR-2) is a key mediator of angiogenesis and an attractive therapeutic target, but existing inhibitors are limited by reduced efficacy, toxicity, and resistance, creating a need for more effective predictive models in drug discovery. In this study, an interpretable machine learning based QSAR approach was developed using a curated dataset of 10,221 VEGFR-2 inhibitors from ChEMBL represented by 164 molecular descriptors. Four algorithms, kNN, AdaBoost, Random Forest, and XGBoost, were compared, and XGBoost achieved the best results with an accuracy of 83.67 percent, sensitivity of 91.38 percent, specificity of 71.73 percent, F1-score of 87.17 percent, and AUC of 0.9009. Model interpretation with LIME identified molecular descriptors related to hydrogen bonding, electrostatics, and lipophilicity as key contributors to activity. These results indicate that interpretable ensemble models can combine strong predictive performance with mechanistic insights, supporting rational design and optimization of novel VEGFR-2 inhibitors for anticancer therapy.

1. Introduction

Cancer remains one of the leading causes of mortality worldwide, accounting for millions of deaths each year [1]. Despite considerable advances in early detection, surgical interventions, and systemic therapies, effective treatment options are still limited for many types of cancer [2–4]. The development of new anticancer drugs is a complex and resource-intensive process, often requiring years of research and substantial financial investment [5]. Among the therapeutic targets identified for cancer therapy, angiogenesis, the formation of new blood vessels that supply nutrients and oxygen to tumors, plays a critical role in tumor progression and metastasis [6]. Targeting angiogenesis has therefore emerged as a promising strategy in modern oncology.

Vascular Endothelial Growth Factor Receptor-2 (VEGFR-2) is a key mediator of angiogenesis, making it an attractive therapeutic target for anticancer drug discovery [7]. Inhibiting VEGFR-2 activity can effectively block tumor vascularization and suppress cancer growth. Although several VEGFR-2 inhibitors have been developed, challenges persist due to issues such as off-target toxicity, limited efficacy, and resistance mechanisms. Traditional drug discovery approaches, while invaluable, are often unable to efficiently capture the complex molecular interactions underlying inhibitor activity [8]. Thus, there is a pressing need for robust and predictive

computational models to accelerate the design and optimization of VEGFR-2 inhibitors.

In recent years, machine learning has become an important tool in pharmaceutical research for modeling complex biological systems and predicting drug properties [9, 10]. Machine learning refers to computational methods that learn patterns from data and make predictions without being explicitly programmed, allowing them to handle large and multidimensional datasets more effectively than conventional statistical approaches [11, 12]. In drug discovery, Quantitative Structure–Activity Relationship (QSAR) modeling is widely used to relate chemical structures to biological activity, and the integration of machine learning into QSAR studies has significantly improved predictive accuracy [13].

Interpretability has become an essential feature of machine learning models in drug discovery. While black-box models can achieve high accuracy, their lack of transparency limits their practical value for guiding medicinal chemists [14, 15]. Advances in interpretable machine learning now make it possible not only to predict VEGFR-2 inhibitory activity but also to explain the molecular descriptors and structural features that influence it [16]. This level of understanding helps identify key determinants of activity, prioritize promising scaffolds, and support the rational design of more effective and selective anticancer agents. By combining accurate prediction with clear mechanistic insight, interpretable QSAR models provide a stronger foundation for VEGFR-2 inhibitor development.

This study aims to develop interpretable machine learning–based QSAR models for predicting the activity of VEGFR-2 inhibitors. The focus is on building models that combine predictive accuracy with interpretability, ensuring both reliable activity prediction and clearer understanding of the molecular features that drive VEGFR-2 inhibition. Interpretable machine learning approaches are used to create a rational framework that can guide the design and optimization of new VEGFR-2 inhibitors, ultimately supporting the discovery of improved anticancer therapies.

2. Materials and Methods

2.1. Data Collection

The dataset employed in this study was obtained from the ChEMBL database, a well-established open-access bioactivity repository widely used in cheminformatics and drug discovery research [17]. Specifically, compounds associated with the vascular endothelial growth factor receptor-2 (VEGFR-2, ChEMBL ID: CHEMBL279) were retrieved, including their half-maximal inhibitory concentration (IC_{50}) values. $IC_{50}$is a standard quantitative measure of a compound's potency, representing the concentration required to inhibit 50% of the biological activity of the target receptor.

For this work, the IC_{50} values were preprocessed and transformed into a binary classification task to facilitate the development of predictive machine learning models. Compounds with IC_{50} values below 1000 nM were labeled as active inhibitors, while compounds with IC_{50} values greater than or equal to 1000 nM were categorized as inactive [18]. This threshold was selected in alignment with commonly adopted criteria in QSAR and chemoinformatics studies to distinguish potent from weak inhibitors.

The final curated dataset consisted of 10,221 compounds, categorized into two classes based on their IC_{50} values: 6,346 active inhibitors (62.09%) and 3,875 inactive inhibitors (37.91%). This distribution reveals a moderate class imbalance, with active compounds forming the majority. Representative examples of the chemical structures included in the dataset are illustrated in Figure 1. Such an imbalance was carefully considered during model development to ensure robust predictive performance and to mitigate potential bias toward the dominant active class.

2.2. Molecular Descriptors Calculation

Molecular descriptors are quantitative representations of chemical structures that capture physicochemical, topological, electronic, and structural properties of compounds [19]. These descriptors serve as features for machine learning models, enabling the establishment of (QSAR models that link chemical structure to biological activity.

In this study, molecular descriptors were computed using the Mordred descriptor calculation package, a widely used open-source tool capable of generating a comprehensive set descriptors [20]. Initially, a large pool of descriptors was generated for all compounds in the curated VEGFR-2 dataset. To improve model performance and reduce redundancy, feature selection techniques were subsequently applied. Descriptors with near-constant variance (threshold < 0.1) were first removed, followed by the elimination of highly correlated descriptors based on a multicollinearity threshold of 0.80 [21].

After preprocessing and feature refinement, a final set of 164 molecular descriptors was retained for use in QSAR model construction. This curated descriptor set provided a balanced representation of the structural and

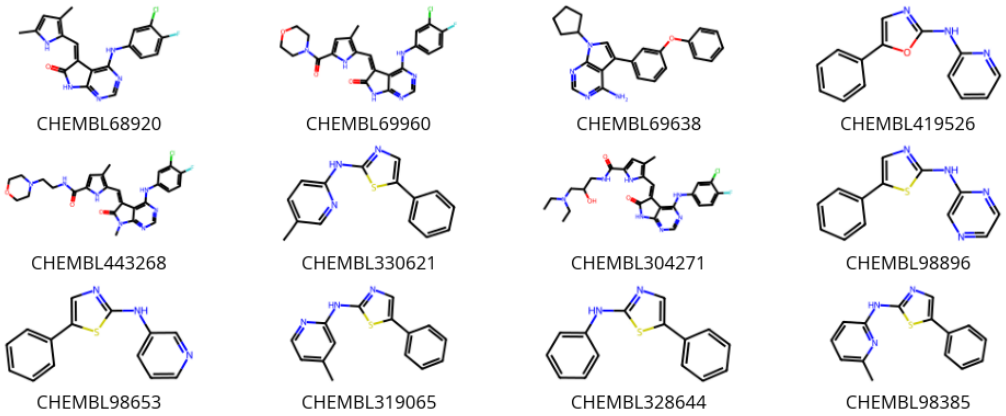

Figure 1. Representative chemical structures of selected VEGFR-2 inhibitors from the ChEMBL database.

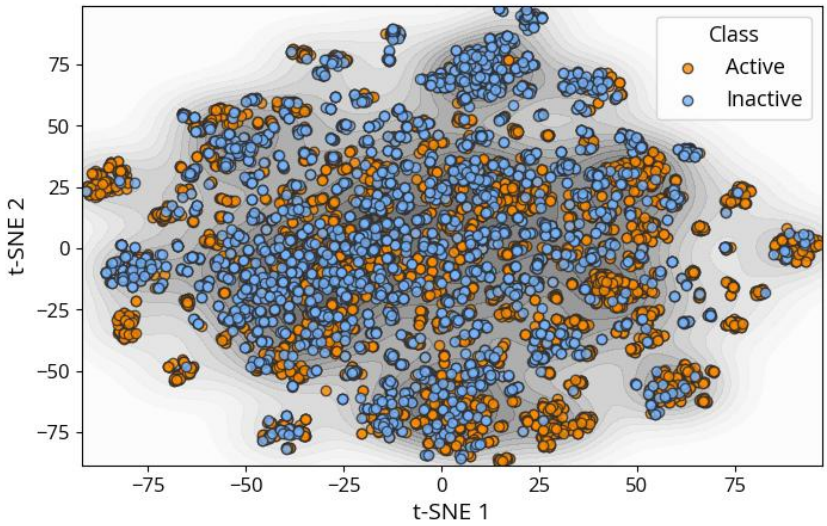

Figure 2. t-SNE visualization of active and inactive VEGFR-2 inhibitors in chemical space.

physicochemical diversity of the compounds, while minimizing noise and redundancy in the data.

2.3. Dataset Preparation

To enable the development and evaluation of predictive models, the curated dataset was partitioned into training and test sets using an 80:20 split ratio [22]. The training set was utilized for model building and internal validation, while the independent test set served to assess the generalization performance of the models. Before model training, all molecular descriptors were subjected to standardization to ensure uniform feature scaling. This step involved transforming each descriptor to have a mean of zero and a standard deviation of one, thereby preventing descriptors with larger numerical ranges from disproportionately influencing the learning process [23].

To visualize the distribution of active and inactive compounds in chemical space, t-SNE dimensionality reduction was applied to the descriptor set. As shown in Figure 2, active and inactive compounds are broadly interspersed, reflecting the structural diversity of the dataset and underscoring the challenge of accurately distinguishing the two classes.

2.4. Machine Learning Algorithms

To assess the predictive performance of QSAR models for VEGFR-2 inhibitors, four machine learning algorithms were selected: k-Nearest Neighbors (kNN), AdaBoost, Random Forest (RF), and Extreme Gradient Boosting (XGBoost). These methods were chosen to represent a broad spectrum of learning strategies, ranging from simple instance-based classification to advanced ensemble techniques, allowing for a balanced and comprehensive comparison.

The kNN algorithm is a non-parametric approach that classifies compounds according to the majority activity of their nearest neighbors in the descriptor space. Its effectiveness depends strongly on the distance metric used and the local structure of the data, making it straightforward yet sensitive to data distribution [24]. AdaBoost, in contrast, is an adaptive ensemble method

that combines multiple weak learners, most often shallow decision trees, into a stronger model. By iteratively reweighting misclassified samples, AdaBoost directs more attention to difficult cases, thereby enhancing overall classification accuracy [25].

Random Forest, another ensemble-based technique, builds multiple decision trees through bootstrap sampling and feature randomness. The aggregation of predictions through majority voting not only improves robustness but also helps to reduce the risk of overfitting, a common concern in high-dimensional QSAR data [26]. XGBoost extends the idea of boosting by sequentially constructing trees, with each new tree correcting the errors of its predecessors [27]. This gradient boosting framework is recognized for its computational efficiency, scalability, and high predictive accuracy, making it especially popular in cheminformatics and bioinformatics applications.

In the present study, all four algorithms were applied using their default hyperparameters to establish baseline performance. This strategy ensures a fair and unbiased comparison while offering an initial evaluation of their suitability for QSAR modeling of VEGFR-2 inhibitors. Subsequent analysis focused on both predictive accuracy and interpretability, laying the groundwork for identifying the most effective approach to support drug discovery.

2.5. Performance Evaluation

The performance of the machine learning models was assessed using a set of widely accepted classification metrics [28, 29].

Specifically, accuracy, precision, sensitivity, specificity, and F1-score were computed to provide a comprehensive evaluation of model performance. Accuracy measures the overall proportion of correctly classified compounds, while precision quantifies the proportion of predicted active compounds that are truly active. Sensitivity (recall) captures the ability of the model to correctly identify active compounds, whereas specificity reflects the ability to correctly identify inactive compounds. The F1-score, as the harmonic mean of precision and recall, provides a balanced assessment, especially in cases of class imbalance.

In addition to these metrics, model performance was further examined through visual diagnostic tools. Confusion matrices were generated to illustrate the distribution of true positives, false positives, true negatives, and false negatives for each model. Furthermore, the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under the Curve (AUC) values were used to evaluate the discriminative ability of the models across various classification thresholds [30]. Finally, the precision–recall (PR) curve was plotted to highlight model performance under different precision-recall trade-offs, which is particularly informative in imbalanced datasets.

2.6. Model Interpretation

To gain mechanistic insight into the structural features influencing VEGFR-2 inhibition, the best-performing machine learning model was interpreted using Local Interpretable Model-agnostic Explanations (LIME). By generating locally faithful surrogate models, LIME enabled the identification of molecular descriptors that most strongly influenced predictions for individual compounds, providing both global and compound-specific interpretability. The analysis highlighted descriptors related to electrostatic interactions, molecular topology, hydrogen bonding potential, and lipophilicity as key determinants of activity, consistent with known pharmacophoric requirements of kinase inhibitors. These interpretable insights reinforce the reliability of the predictive model and offer rational guidance for the design of novel VEGFR-2 inhibitors.

3. Results and Discussion

The predictive performance of the four machine learning models, kNN, AdaBoost, RF, and XGBoost, was evaluated on the test set using accuracy, precision, sensitivity, specificity, and F1-score (Table 1). Among the models tested, ensemble-based approaches again demonstrated superior predictive capabilities compared to the instance-based kNN classifier. XGBoost achieved the highest overall accuracy (83.67%), followed by Random Forest (82.15%). Both models also yielded the highest F1-scores (87.17% and 86.21%, respectively), indicating a strong balance between precision and recall. By contrast, kNN and AdaBoost produced comparatively lower accuracies (81.12% and 72.08%) and F1-scores (85.17% and 78.94%), reflecting their relatively weaker capacity to capture the complex structure–activity relationships within the dataset.

Sensitivity, which reflects the ability to correctly identify active VEGFR-2 inhibitors, was highest for Random Forest (91.87%) and XGBoost (91.38%), suggesting that these models were particularly effective at minimizing false negatives. This is a critical attribute in drug discovery, where overlooking potentially active compounds could lead to missed opportunities. However, AdaBoost's strength in sensitivity (86.15%) came at the cost of reduced specificity (50.31%), indicating difficulty in distinguishing inactive inhibitors.

Table 1. Performance of four machine learning models on the test set.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
KNN	81.12	81.47	89.21	68.62	85.17
AdaBoost	72.08	72.84	86.15	50.31	78.94
Random Forest	82.15	81.21	91.87	67.12	86.21
XGBoost	83.67	83.33	91.38	71.73	87.17

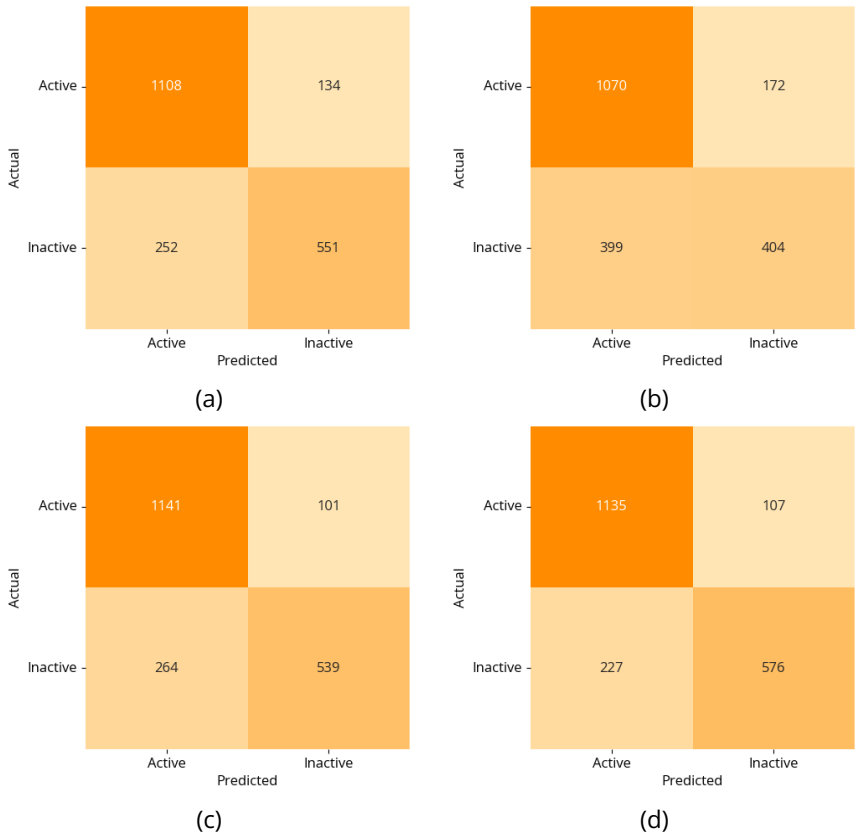

Figure 3. Confusion matrices of machine learning models for VEGFR-2 inhibitor classification: (a) kNN, (b) AdaBoost, (c) Random Forest, and (d) XGBoost.

Conversely, Random Forest and XGBoost offered a more favorable trade-off between sensitivity and specificity. Both maintained high sensitivity (91.87% and 91.38%, respectively) while also delivering stronger specificity values (67.12% and 71.73%) compared to kNN (68.62%) and AdaBoost. This balance is particularly important given the moderate class imbalance in the dataset, ensuring that inactive compounds are not disproportionately misclassified as actives.

Precision analysis revealed that XGBoost (83.33%) and kNN (81.47%) slightly outperformed the other models in correctly identifying true active inhibitors among predicted positives. The higher precision values of these models reduce the likelihood of false positives, which is advantageous when prioritizing compounds for experimental validation. In contrast, while Random Forest and AdaBoost maintained acceptable precision values (81.21% and 72.84%, respectively), their overall predictive reliability was limited by weaker specificity and accuracy.

Taken together, the results highlight the advantage of tree-based ensemble learning methods over simpler classifiers for QSAR modeling of VEGFR-2 inhibitors. The superior performance of Random Forest and XGBoost can be attributed to their ability to capture non-linear relationships and interactions among molecular descriptors, while also mitigating overfitting through ensemble averaging and boosting strategies. In contrast, kNN's reliance on local similarity measures may have limited its ability to generalize across the chemically diverse dataset. Similarly, AdaBoost's iterative reweighting process, though effective at boosting sensitivity, introduced bias toward the majority class at the expense of specificity.

To further examine classification behavior, confusion matrices were generated for each machine learning model (Figure 3). These matrices provide detailed insight into the distribution of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The kNN model achieved a high number of true positives

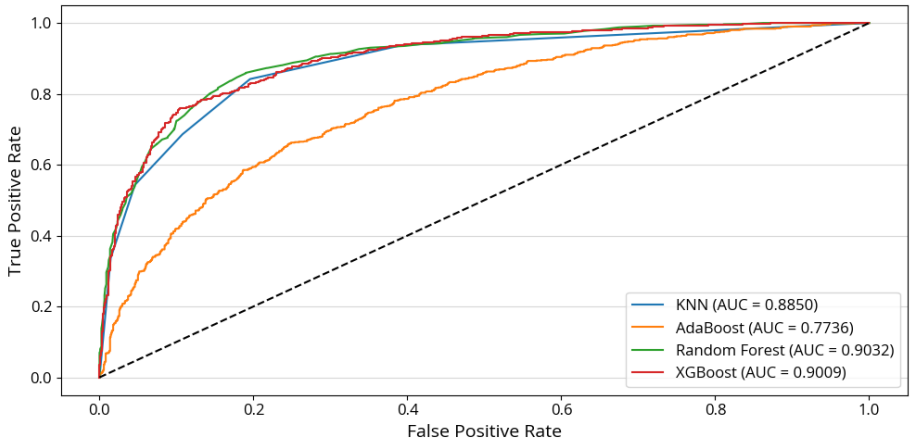

Figure 4. ROC curves of kNN, AdaBoost, Random Forest, and XGBoost models with corresponding AUC values.

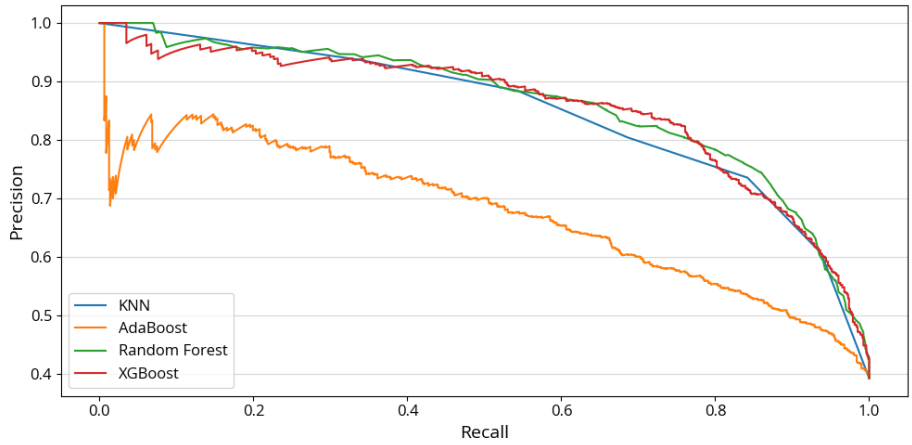

Figure 5. Precision–Recall (PR) curves of kNN, AdaBoost, Random Forest, and XGBoost models.

(1108) but also produced 134 false negatives and a comparatively low true negative count (551), reflecting weaker specificity relative to ensemble methods. AdaBoost performed less effectively, with higher false negatives (172) and false positives (399), yielding only 404 true negatives and confirming its bias toward predicting active inhibitors at the expense of specificity. Random Forest, by contrast, demonstrated strong performance with the highest true positive count (1141), few false negatives (101), and 539 true negatives, achieving a balanced trade-off between sensitivity and specificity. Similarly, XGBoost delivered robust results with 1135 true positives, 107 false negatives, and the highest true negative count (576), underscoring its superior balance between sensitivity and specificity and reinforcing its overall reliability.

The discriminative performance of the machine learning models was further assessed using ROC curves and their corresponding AUC values (Figure 4). ROC analysis provides a threshold-independent measure of classification ability, offering insights into the trade-off between sensitivity and specificity across different decision thresholds. The Random Forest model achieved the highest AUC value (0.9032), indicating excellent discriminative power in distinguishing active from inactive VEGFR-2 inhibitors. The XGBoost model closely followed with an AUC of 0.9009, also demonstrating robust classification capability with only a marginal difference from Random Forest. The kNN model showed strong performance as well, with an AUC of 0.8850, reinforcing its relatively high sensitivity despite weaker specificity. In contrast, the AdaBoost model (AUC = 0.7736) exhibited a noticeably weaker ROC profile, suggesting less reliable performance across varying thresholds and confirming its earlier limitations in specificity and overall predictive balance.

Given the moderate class imbalance in the dataset (62.09% actives vs. 37.91% inactives), Precision–Recall (PR) curves were examined to better assess model performance under imbalanced conditions (Figure 5). PR analysis is particularly informative in such cases, as it directly reflects the trade-off between precision (minimizing false positives) and recall (minimizing false negatives). Both Random Forest and XGBoost consistently maintained high precision across a wide range of recall values, confirming their strong ability to

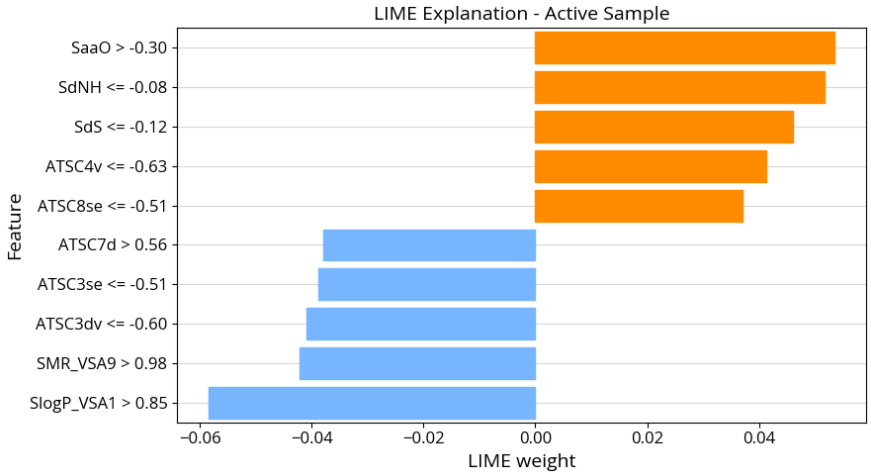

Figure 6. LIME explanation for an active sample.

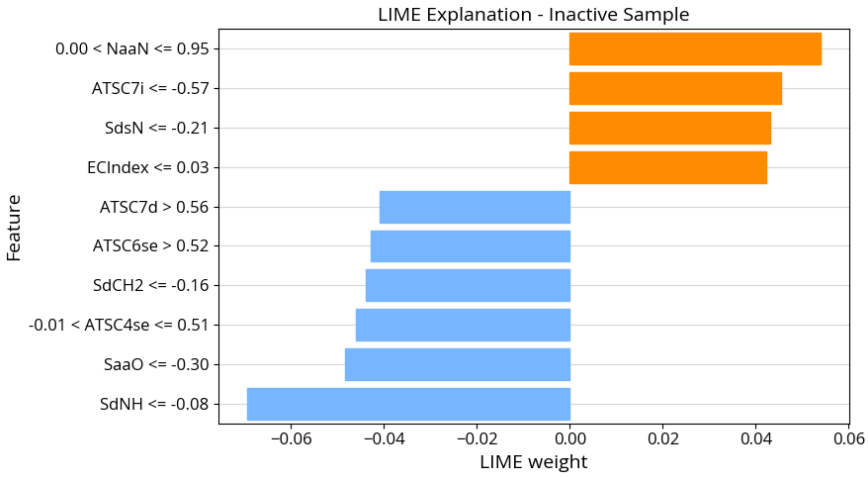

Figure 7. LIME explanation for an inactive sample.

reliably identify active inhibitors while preserving classification accuracy. XGBoost showed slightly more stability at higher recall thresholds, while Random Forest exhibited comparable performance overall. The kNN model also performed relatively well, though its precision declined more steadily as recall increased, indicating reduced reliability when attempting to capture a larger fraction of true positives. By contrast, AdaBoost performed the weakest, with a pronounced drop in precision throughout the recall range, reflecting its tendency to overpredict active compounds even when confidence was low.

Among the models evaluated, XGBoost emerged as the most effective approach for VEGFR-2 inhibitor classification. It consistently achieved a strong balance between sensitivity and specificity, showed excellent discriminative ability in ROC analysis, and maintained high precision across a broad range of recall values in PR evaluation. The confusion matrix further confirmed its reliability in correctly identifying both active and inactive compounds. Overall, these results highlight XGBoost's robustness and suitability for QSAR modeling in drug discovery, where minimizing both false positives and false negatives is critical.

Since XGBoost was identified as the best-performing model, we further explored its interpretability using LIME to understand the molecular features driving classification decisions. Figure 6 shows the LIME explanation for an active sample, highlighting structural and physicochemical descriptors that contributed positively or negatively to the prediction. Features such as SaaO, SdNH, and SdS had strong positive weights, supporting the classification of the compound as active, while descriptors like SlogP_VSA1 and SMR_VSA9 contributed negatively, opposing the active label. This local interpretability analysis provides valuable insights into how XGBoost leverages molecular descriptors to differentiate active inhibitors from inactives, bridging predictive accuracy with chemical interpretability.

To further probe model interpretability, LIME was also applied to an inactive sample (Figure 7). The analysis

highlights features such as NaaN, ATSC7i, and SdSN with strong positive contributions supporting the inactive classification, while descriptors including SdNH, SaaO, and ATSC4se exerted negative influence, opposing the inactive label. Together, these insights demonstrate how XGBoost integrates structural and physicochemical information to distinguish inactive compounds, complementing the earlier active-sample analysis and reinforcing the model's capacity for interpretable, compound-level predictions.

This study demonstrates the potential of interpretable machine learning models for QSAR analysis of VEGFR-2 inhibitors, but several limitations should be noted. First, the dataset, while large and diverse, was drawn only from the ChEMBL database. As a result, the models may not fully capture the wider chemical space relevant to VEGFR-2 inhibition, which could limit their ability to generalize to novel scaffolds. Incorporating additional experimental or proprietary data could improve robustness. Second, the binary classification approach, based on a fixed IC_{50} threshold, simplified model development but did not reflect the full range of inhibitory potency. Regression-based modeling of continuous IC_{50} values could provide a more detailed view of activity and help prioritize the most potent compounds. Third, model interpretability was explored using LIME, which provided useful local explanations for individual predictions. However, these insights are sample-specific. Combining such local methods with global approaches, such as feature importance or SHAP values, may give a more complete understanding of how molecular descriptors influence predictions across different inhibitor classes. Finally, the models were trained using default hyperparameters to allow baseline comparisons. Although XGBoost and Random Forest performed well, systematic hyperparameter optimization and ensemble strategies could further enhance accuracy and stability.

Future work could focus on integrating these models into a virtual screening pipeline that includes molecular docking, pharmacokinetic evaluation, and toxicity prediction, creating a more comprehensive workflow for anticancer drug discovery. Extending the approach to multi-target modeling may also help address resistance by identifying inhibitors with broader kinase selectivity.

4. Conclusions

This study demonstrates that interpretable machine learning–based QSAR modeling offers a powerful approach for predicting and rationalizing the activity of VEGFR-2 inhibitors in anticancer drug discovery. Among the tested algorithms, XGBoost emerged as the most effective, achieving the highest accuracy, F1-score, and balanced trade-off between sensitivity and specificity, while also providing robust interpretability through LIME analyses. The insights gained into key molecular descriptors underscore the potential of interpretable models not only to enhance predictive performance but also to guide medicinal chemists in scaffold optimization and rational drug design. Although limitations remain such as dataset scope, binary classification thresholds, and the need for hyperparameter optimization, the framework established here highlights the value of combining predictive accuracy with interpretability. Future integration with virtual screening, docking, and pharmacokinetic evaluations may accelerate the discovery of potent, selective, and clinically viable VEGFR-2 inhibitors, ultimately advancing therapeutic strategies in oncology.

Author Contributions: Conceptualization, T.R.N. and R.I.; methodology, R.I.; software, T.R.N.; validation, R.I.; formal analysis, T.R.N.; investigation, T.R.N.; resources, R.I.; data curation, R.I.; writing—original draft preparation, T.R.N.; writing—review and editing, R.I.; visualization, T.R.N.; supervision, R.I.; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript."

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Lin, H.-Y., and Park, J. Y. (2023). Epidemiology of Cancer, *Anesthesia for Oncological Surgery*, Springer International Publishing, Cham, 11–16. doi:10.1007/978-3-031-50977-3_2.
2. Liu, B., Zhou, H., Tan, L., Siu, K. T. H., and Guan, X.-Y. (2024). Exploring Treatment Options in Cancer: Tumor Treatment Strategies, *Signal Transduction and Targeted Therapy*, Vol. 9, No. 1, 175. doi:10.1038/s41392-024-01856-7.
3. Bertolaccini, L., Casiraghi, M., Uslenghi, C., Maiorca, S., and Spaggiari, L. (2024). Recent Advances in Lung Cancer Research: Unravelling the Future of Treatment, *Updates in Surgery*, Vol. 76, No. 6, 2129–2140. doi:10.1007/s13304-024-01841-3.
4. Fitzgerald, R. C., Antoniou, A. C., Fruk, L., and Rosenfeld, N. (2022). The Future of Early Cancer Detection, *Nature Medicine*, Vol. 28, No. 4, 666–677. doi:10.1038/s41591-022-01746-x.
5. Carugo, A., and Draetta, G. F. (2019). Academic Discovery of Anticancer Drugs: Historic and Future Perspectives, *Annual Review of Cancer Biology*, Vol. 3, No. 1, 385–408. doi:10.1146/annurev-cancerbio-030518-055645.
6. Liu, Z.-L., Chen, H.-H., Zheng, L.-L., Sun, L.-P., and Shi, L. (2023). Angiogenic Signaling Pathways and Anti-Angiogenic Therapy for Cancer, *Signal Transduction and Targeted Therapy*, Vol. 8, No. 1, 198. doi:10.1038/s41392-023-01460-1.

7. Shah, A. A., Kamal, M. A., and Akhtar, S. (2021). Tumor Angiogenesis and VEGFR-2: Mechanism, Pathways and Current Biological Therapeutic Interventions, *Current Drug Metabolism*, Vol. 22, No. 1, 50–59. doi:10.2174/1389200221666201019143252.

8. Dhudum, R., Ganeshpurkar, A., and Pawar, A. (2024). Revolutionizing Drug Discovery: A Comprehensive Review of AI Applications, *Drugs and Drug Candidates*, Vol. 3, No. 1, 148–171. doi:10.3390/ddc3010009.

9. Choudhuri, S., Yendluri, M., Poddar, S., Li, A., Mallick, K., Mallik, S., and Ghosh, B. (2023). Recent Advancements in Computational Drug Design Algorithms through Machine Learning and Optimization, *Kinases and Phosphatases*, Vol. 1, No. 2, 117–140. doi:10.3390/kinasesphosphatases1020008.

10. Noviandy, T. R., Nisa, K., Idroes, G. M., Hardi, I., and Sasmita, N. R. (2024). Classifying Beta-Secretase 1 Inhibitor Activity for Alzheimer's Drug Discovery with LightGBM, *Journal of Computing Theories and Applications*, Vol. 2, No. 2, 138–147. doi:10.62411/jcta.10129.

11. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. B., Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.

12. Supriatna, D. J. I., Saputra, H., and Hasan, K. (2023). Enhancing the Red Wine Quality Classification Using Ensemble Voting Classifiers, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 42–47. doi:10.60084/ijds.v1i2.95.

13. Noviandy, T. R., Maulana, A., Idroes, G. M., Irvanizam, I., Subianto, M., and Idroes, R. (2023). QSAR-Based Stacked Ensemble Classifier for Hepatitis C NS5B Inhibitor Prediction, *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, 220–225. doi:10.1109/COSITE60233.2023.10250039.

14. Noviandy, T. R., Maulana, A., Idroes, G. M., Suhendra, R., Afidh, R. P. F., and Idroes, R. (2024). An Explainable Multi-Model Stacked Classifier Approach for Predicting Hepatitis C Drug Candidates, *Sci*, Vol. 6, No. 4, 81. doi:10.3390/sci6040081.

15. Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., and Ranjan, R. (2023). Explainable AI (XAI): Core Ideas, Techniques, and Solutions, *ACM Computing Surveys*, Vol. 55, No. 9, 1–33. doi:10.1145/3561048.

16. Noviandy, T. R., Maulana, A., Irvanizam, I., Idroes, G. M., Maulydia, N. B., Tallei, T. E., Subianto, M., and Idroes, R. (2025). Interpretable Machine Learning Approach to Predict Hepatitis C Virus NS5B Inhibitor Activity Using Voting-Based LightGBM and SHAP, *Intelligent Systems with Applications*, Vol. 25, 200481. doi:10.1016/j.iswa.2025.200481.

17. Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overington, J. P. (2012). ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery, *Nucleic Acids Research*, Vol. 40, No. D1, D1100–D1107. doi:10.1093/nar/gkr777.

18. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Machine Learning Approach to Predict AXL Kinase Inhibitor Activity for Cancer Drug Discovery Using XGBoost and Bayesian Optimization, *Journal of Soft Computing and Data Mining*, Vol. 5, No. 1, 46–56.

19. Grisoni, F., Consonni, V., and Todeschini, R. (2018). Impact of Molecular Descriptors on Computational Models, 171–209. doi:10.1007/978-1-4939-8639-2_5.

20. Moriwaki, H., Tian, Y. S., Kawashita, N., and Takagi, T. (2018). Mordred: A Molecular Descriptor Calculator, *Journal of Cheminformatics*, Vol. 10, No. 1, 1–14. doi:10.1186/s13321-018-0258-y.

21. Tropsha, A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation, *Molecular Informatics*, Vol. 29, Nos. 6–7, 476–488. doi:10.1002/minf.201000061.

22. Joseph, V. R. (2022). Optimal Ratio for Data Splitting, *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Vol. 15, No. 4, 531–538. doi:10.1002/sam.11583.

23. Suhendra, R., Husdayanti, N., Suryadi, S., Juliwardi, I., Sanusi, S., Ridho, A., Ardiansyah, M., Murhaban, M., and Ikhsan, I. (2023). Cardiovascular Disease Prediction Using Gradient Boosting Classifier, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 56–62. doi:10.60084/ijds.v1i2.131.

24. Singh, R., Ganeshpurkar, A., Ghosh, P., Pokle, A. V., Kumar, D., Singh, R. bhushan, Singh, S. K., and Kumar, A. (2021). Classification of Beta-site Amyloid Precursor Protein Cleaving Enzyme 1 Inhibitors by Using Machine Learning Methods, *Chemical Biology & Drug Design*, Vol. 98, No. 6, 1079–1097. doi:10.1111/cbdd.13965.

25. Noviandy, T. R., Idroes, G. M., Hardi, I., Afjal, M., and Ray, S. (2024). A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry, *Infolitika Journal of Data Science*, Vol. 2, No. 1, 34–44. doi:10.60084/ijds.v2i1.199.

26. Maulana, A., Afidh, R. P. F., Maulydia, N. B., Idroes, G. M., and Rahimah, S. (2024). Predicting Obesity Levels with High Accuracy: Insights from a CatBoost Machine Learning Model, *Infolitika Journal of Data Science*, Vol. 2, No. 1, 17–27. doi:10.60084/ijds.v2i1.195.

27. Hidayat, T., Hadinata, E., Damanik, I. S., Vikki, Z., and Irvanizam, I. (2023). Implementation of Hybrid CNN-XGBoost Method for Leukemia Detection Problem, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 15–21. doi:10.60084/ijds.v1i1.87.

28. Ahsan, M. M., Luna, S. A., and Siddique, Z. (2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review, *Healthcare*, Vol. 10, No. 3, 541. doi:10.3390/healthcare10030541.

29. Noviandy, T. R., Idroes, G. M., Syukri, M., and Idroes, R. (2024). Interpretable Machine Learning for Chronic Kidney Disease Diagnosis: A Gaussian Processes Approach, *Indonesian Journal of Case Reports*, Vol. 2, No. 1, 24–32. doi:10.60084/ijcr.v2i1.204.

30. Tharwat, A. (2021). Classification Assessment Methods, *Applied Computing and Informatics*, Vol. 17, No. 1, 168–192. doi:10.1016/j.aci.2018.08.003.