JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5994

PREDICTION OF INHIBITOR BINDING AFFINITY AND MOLECULAR INTERACTIONS IN MPRO DENGUE USING MACHINE LEARNING

Venia Restreva Danestiara[1*]; Marwondo[2]; Nayla Nurul Azkiya[3]

Informatics Study Program[1, 2, 3]
Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia[1, 2, 3]
www.unibi.ac.id[1, 2, 3]
veniarestreva@unibi.ac.id[1*], marwondo@unibi.ac.id[2], naylanurulazkiya23@student.unibi.ac.id[3]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract—*The dengue virus experiences rapid mutation and genetic variability, posing challenges in developing effective antiviral therapies. This study explores the prediction of binding affinities between potential antiviral drug inhibitors and the NS2B-NS3 protease of the dengue virus using machine learning models. Molecular docking simulations were conducted with AutoDock Vina to generate interaction data between viral proteins and ligands. The generated datasets were used to train several machine learning models, including Random Forest Regressor (RF Regressor), Support Vector Regression (SVR), and Extreme Gradient Boosting Regressor (XGBoost Regressor). The RF Regressor model demonstrated the highest accuracy in predicting binding affinities, measured through Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Pearson Correlation Coefficient (R). However, the XGBoost Regressor and SVR models showed better generalization in practical scenarios. This study highlights the potential of machine learning to optimize the drug discovery process and provides significant insights into antiviral drug development for dengue fever.*

Keywords: *binding affinities, dengue virus, machine learning, molecular docking, NS2B-NS3 protease.*

Intisari—*Virus dengue mengalami mutasi cepat dan variabilitas genetik, yang menimbulkan tantangan dalam pengembangan terapi antivirus yang efektif. Penelitian ini mengeksplorasi prediksi afinitas pengikatan antara inhibitor obat antiviral potensial dan protease NS2B-NS3 virus dengue menggunakan model pembelajaran mesin. Simulasi docking molekuler dilakukan dengan AutoDock Vina untuk menghasilkan data interaksi antara protein virus dan ligan. Dataset yang dihasilkan digunakan untuk melatih beberapa model pembelajaran mesin, termasuk Random Forest Regressor (RF Regressor), Support Vector Regression (SVR), dan Extreme Gradient Boosting Regressor (XGBoost Regressor). Model RF Regressor menunjukkan akurasi tertinggi dalam memprediksi afinitas pengikatan, diukur dengan Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan Pearson Correlation Coefficient (R). Namun, model XGBoost Regressor dan SVR menunjukkan generalisasi yang lebih baik dalam skenario praktis. Penelitian ini menyoroti potensi pembelajaran mesin untuk mengoptimalkan proses penemuan obat dan memberikan wawasan penting dalam pengembangan obat antivirus untuk demam dengue.*

Kata Kunci: *afinitas pengikatan, virus dengue, pembelajaran mesin, docking molekuler, protease NS2B-NS3.*

INTRODUCTION

Dengue fever remains a significant public health challenge in Indonesia. In 2023, there were 114,435 cases and 894 deaths, while in the first eight weeks of 2024, there were 15,977 cases and 124 deaths. Notably, the DENV-3 and DENV-2 serotypes dominate, contributing 53.4% and 38.6% of total cases, respectively, with a higher risk of severe complications such as dengue shock syndrome (DSS) [1]. The genetic variability of the pathogen and the limitations of antiviral treatments exacerbate the situation, emphasizing the urgent need for effective therapies [2].

The NS2B-NS3 protease plays a crucial role in dengue virus replication, functioning in the

processing of proteins necessary for the formation of infectious virus particles. The interaction between NS2B and NS3 forms a protease complex that cleaves viral polyproteins into functional proteins. Recent research indicates that inhibiting this protease can significantly reduce viral replication, making it an attractive therapeutic target. A study reported that compounds inhibiting this interaction can decrease viral replication by up to 80% in cell cultures [3]. Additionally, epidemiological data show that dengue infections can lead to severe complications, so the development of effective inhibitors against the NS2B-NS3 protease could contribute to the reduction of dengue-related morbidity and mortality. Therefore, these proteins are a primary focus in the development of new antiviral therapies.

Despite promising research on protease inhibitors, there remains a gap in understanding the specific mechanisms behind the molecular interactions between ligands and this protease. Many existing studies still rely on limited datasets that do not encompass the broad genetic variability of DENV, which can affect the effectiveness of the developed inhibitors. Consequently, further research is needed to create a comprehensive protein-ligand dataset and to apply machine learning techniques to improve binding affinity predictions and the efficacy of antiviral therapies.

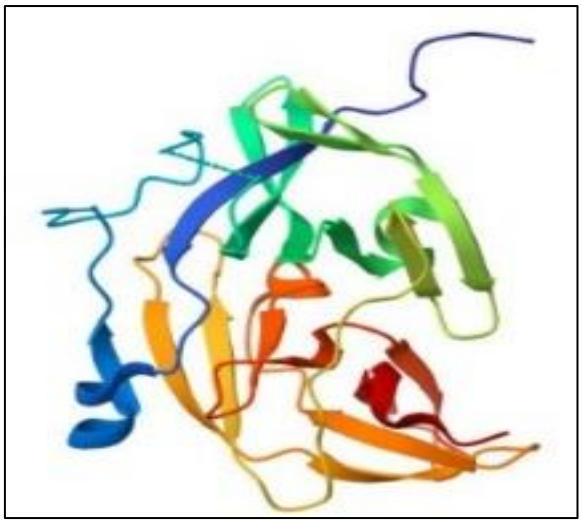
The specific objectives of this research include: (1) creating a new protein-ligand dataset that encompasses DENV genetic variability, (2) developing more accurate binding affinity prediction models through the integration of machine learning algorithms, and (3) exploring key interactions between ligands and the NS2B-NS3 protease to design more effective inhibitors.

Antiviral agents for Dengue Hemorrhagic Fever (DHF) can be evaluated using molecular docking techniques, which predict interactions between small molecules (ligands) and proteins (receptors) [4]. DHF, caused by four DENV serotypes, involves non-structural proteins (NS) that play roles in viral replication and immune evasion. In Indonesia, NS3 and NS2 are the primary targets for treatment. NS2A and NS2B assist in viral assembly and replication, while NS2B-NS3 must collaborate for DENV replication, making it a crucial therapeutic target [5]. NS3 also interacts with NS4A for viral genome cleavage and DENV RNA processing [6], [7].

This research focuses on developing a new protein-ligand dataset and binding affinity prediction models using machine learning scoring functions, aiming to reduce reliance on existing datasets. Docking simulations between antiviral agents and the NS2 and NS3 proteins of the dengue virus were conducted using AutoDock Vina due to its significance in Dengue Fever (DF) cases in Indonesia. The generated protein-ligand interaction data enhance the accuracy of the machine learning models by identifying key interaction patterns. Random Forest Regressor (RF Regressor) and Support Vector Regression (SVR) are commonly used algorithms in this study [8], [9], with RF Regressor often excelling in Pearson Correlation Coefficient (R) evaluations [10], [11], although Extreme Gradient Boosting Regressor (XGBoost Regressor) sometimes performs better [12], [13]. SVR dominates in Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) evaluations [14], [15]. This study integrates three algorithms into a new scoring function, aiming to improve accuracy and predictive resilience.

MATERIALS AND METHODS

As a vital part in the replication of Dengue virus, NS2B-NS3pro protein plays a major role in the virus polypeptide process and is a potential target for the development of antivirus. The NS2B-NS3pro protein obtained from the RSCB PDB database is labeled with the PDB code 4M9K [15] . The 4M9K protein is selected based on the dengue virus type 2 virus protein. In this study, the NS2B-NS3pro protein is chosen as the main target for molecular docking with various ligands in order to identify candidate molecules that could be effective protease inhibitors. Figure 1 depicts the structure of the NS2B-NS3pro receptor protein (PDB ID: 4M9K) used in this study.



Source: (https://www.rcsb.org/structure/4M9K. [Accessed: Jul. 05, 2024])

Figure 1. Protein Reseptor PDB ID 4M9K

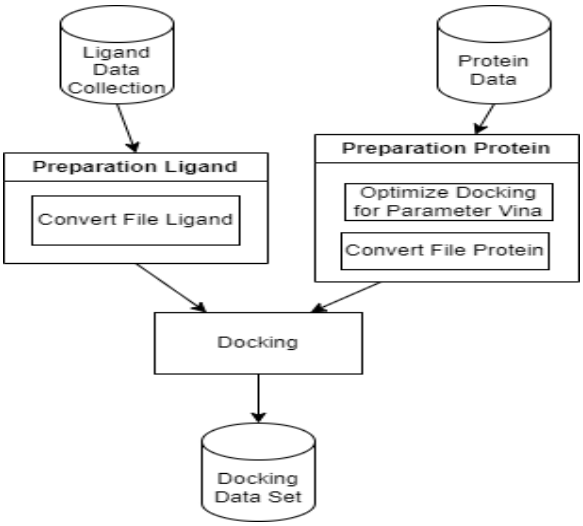
Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5994

System Design

This research utilizes AutoDock Vina version 1.1.2 to obtain a dataset containing the receptor-ligand binding affinity and the strength of receptor-ligand complex binding as shown in Figure 2 [16]. The search for binding affinity is carried out with the following flow.

Preparing a set of ligand molecules obtained through the DrugBank database to interact with the NS2B-NS3pro receptor of dengue virus type 2 [17]. The docking process uses both data to simulate the interaction between the ligand and receptor. A total of 1,138 ligand configurations were meticulously curated for various viral infections like hepatitis C, malaria, cancer, and more. Additionally, the ligands were selected based on confirmed or tested features. Furthermore, the ligands are converted from PDB into PDBQT files before docking.

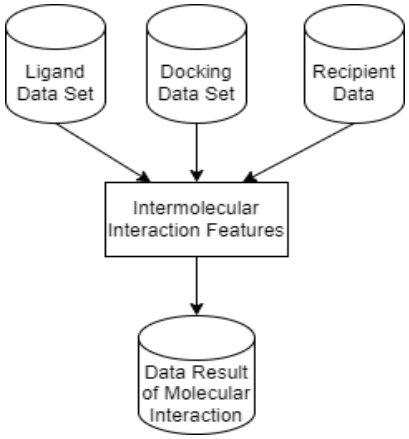


Source: (Danestiara, 2021) [16]

Figure 2. Overview of the process of creating a new data set with docking using AutoDock Vina.

Protein 4M9K loses water molecules as water can disrupt the interaction between the protein and ligand. Then, charged hydrogen atoms are added to improve accuracy in predicting binding locations. Determining the grid box on the protein is done to determine the ligand's location during the docking process, with the grid box determining the center and spatial dimensions of the docking area.

Few important parameters were set to get the best results prior to running molecular docking simulations using AutoDock Vina. The search space of ligand-receptor inter actions was defined by a grid box with the center coordinates x, y, z of 17.771, 9.012, 5.203 angstroms and dimensions x, y, z of 50,50, and 40Å. Furthermore, the protein is converted into PDBQT files before docking with Autodock Vina.

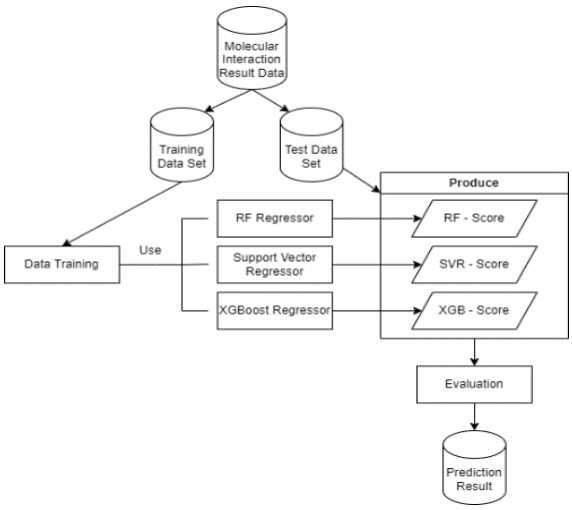


Source: (Danestiara, 2021) [16]

Figure 3. The Process Flow of Molecular Interaction Feature Using A Set Of Docking Data

Once the docking process is finished, the data is saved in a Docking Data Set, which includes information about the binding affinity or interaction energy between the ligand and receptor. Next, we preprocess this data using molecular interaction features to map interactions at the atomic level by measuring distances between atoms. Finally, the data is exported to a .csv file for further analysis with machine learning. Figure 3 presents an overview of the process flow for docking data set processing.

After obtaining the set of data on intermolecular interactions, the prediction of binding affinity values of receptor-ligand complexes is carried out using three machine learning algorithms for scoring functions, namely RF Regressor, SVR, and XGBoost Regressor.



Source: (Research Result, 2024)

Figure 4. The Process Flow Of Predicting The Affinity Value Bond Using Machine Learning Scoring Function by RF Regressor, SVR, Xgboost Regressor.

Figure 4 overviews machine learning applied to intermolecular interaction data. The results of the collection of intermolecular interaction data are split into 80% training data and 20% test data from 1118 data. Subsequently, predictions are made on the training data set using RF Regressor, SVR, XGBoost Regressor. This process results in RF-Score, SVR-Score, and XGBoost-Score, and the predicted values of binding affinity for the receptor-ligand complex as a whole.

Features of Intermolecular Interactions

The relationship between receptor-ligand complexes and affinity is influenced by intermolecular interactions, and Machine Learning can use regression techniques to analyze these non-linear relationships. Molecular interaction involves measuring distances between atoms in receptor-ligand complexes. For example, the interaction between carbon and nitrogen is a key feature of each complex. This research focuses on heavy atoms typically found in PDB structures, which helps efficiently analyze intermolecular relationships [18]. While using various types of atoms can increase the number of features, it's important to keep their values low to avoid impacting predictions and delays in feature clustering. In this study, nine types of heavy atoms are considered in the receptor and ligand.

$$\{P(i)\}_{i=1}^{9} = \{C, N, O, F, P, S, Cl, Br, I\} \quad (1)$$
$$\{L(i)\}_{i=1}^{9} = \{C, N, O, F, P, S, Cl, Br, I\} \quad (2)$$

The molecular interaction between receptor-ligand is defined as follows:

$$x\left(A(P(j)), A(L(i))\right) \equiv \sum_{p=1}^{P_j} \sum_{l=1}^{L(i)} \theta(d_{cutoff} - d_{pl}) \quad (3)$$

All the required data is contained in the PDB files of the receptor and ligand. Where Pj and Li represent the number of receptor and ligand atoms. The role that changes the atom number and is used to replace the characteristics that have been determined as A. dpl is the Euclidean distance that calculates the distance between atoms. d_{cutoff} <12 Å, the cutoff selection is inspired by PMF which captures the solvation effect very well even though there is no claim stating that this selection is the optimal choice. Θ is the Heaviside step function that determines the receptor-ligand complexes involved.

$$\Theta\left(d_{pl}, d_{cutoff}\right)\begin{cases} d_{pl} \leq 12\ \text{Å}, 1 \\ d_{pl} > 12\ \text{Å}, 0 \end{cases} \quad (4)$$

If the distance between two molecules (dpl) is greater than 12 Å, then the intermolecular interactions in the receptor-ligand complex will have a value of zero, indicating that it is not included in the feature set and vice versa.

Furthermore, each receptor-ligand complex has 36 features characterized in vector form, namely:

$$\vec{x} = \begin{pmatrix} x_{6,6}, x_{6,7}, x_{6,8}, x_{6,9}, x_{6,15}, x_{6,16}, x_{6,17}, x_{6,35}, x_{6,53}, \\ x_{7,6}, \dots, x_{53,53} \end{pmatrix} \in \aleph^{36} \quad (5)$$

The calculation process of intermolecular interactions is carried out throughout the receptor-ligand complex that will be predicted against the receptor-ligand. After that, the results of a series of collected features will be compiled into the following dataset:

$$D = \{(\vec{x}^n, y^n)\}_{n=1}^{N} \quad (6)$$

Where D is a set of pre-processed coal data. $\vec{x}^n$ is a 36-feature vector containing complex receptor-ligand atom interactions. y^n is the actual receptor-ligand binding affinity.

Binding Affinity Prediction by Random Forest Regressor

RF Regressor is an ensemble method that combines multiple decision trees. The algorithm works by randomly splitting the dataset into subsets, building decision trees for each subset, and then averaging the results of the trees to make final predictions. RF is defined as follows:

$$RF \equiv \left(x\,; m_{try}\right) \equiv \frac{1}{P}\sum_{p=1}^{P} T_p(\vec{x}^{(n)}; m_{try})$$
$$T_p: \aleph^{36} \rightarrow R^{+} \forall p \quad (7)$$

Where the number of trees is 5000 trees, p is the tree index and Tp is the regression value of individual trees in the forest, m_{try} is some randomly selected number of features, x affinity bonds will be predicted by RF Regressor.

In order to assess the effectiveness of the machine learning models (RF Regressor, SVR, and XGBoost Regressor) in a structured manner, we utilized three common statistical measures: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (R). These techniques are employed to measure the discrepancy between observed and estimated values. A good prediction model is one where the predicted value $f(\vec{x}^{(n)})$ closely matches the actual value $y^{(n)}$, namely:

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5994

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|y^{(n)} - f(\vec{x}^{(n)})\right| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N}}\sum_{n=1}^{N}(y^{(n)} - f(\vec{x}^{(n)}))^2 \quad (9)$$

$$R \equiv \frac{N\sum_{n=1}^{N}(y^{(n)}, f(\vec{x}^{(n)})) - \sum_{n=1}^{N}(y^{(n)})\sum_{n=1}^{N} f(\vec{x}^{(n)}))}{\sqrt{(N\sum_{n=1}^{N}(y^{(n)})^2 - (\sum_{n=1}^{N}(y^{(n)})^2)(N\sum_{n=1}^{N} f(\vec{x}^{(n)})^2 - (\sum_{n=1}^{N} f(\vec{x}^{(n)})^2))} \quad (10)$$

The MAE, RMSE, and R methods are widely used to evaluate the accuracy of regression models. These metrics measure how close the model predictions are to the actual values in the dataset, so they can be consistently applied to various regression algorithms, including RF, SVR, and XGBoost. The difference lies in how each model generates prediction values $f(\vec{x}^{(n)})$.

Binding Affinity Prediction by Support Vector Regressor

The SVR technique uses kernels such as Radial Basis Function (RBF) to map data into higher dimensions, predicting the binding affinity between ligands and receptors. This approach reveals non-linear relationships between molecular features from atom interactions [19]. The RBF kernel enables SVR to capture complex, non-linear patterns, enhancing prediction accuracy. SVR is defined as follows:

$$SVR = \sum_{i=1}^{N}(a_i - a_i^*)\, K(x_i, x) + b \quad (11)$$

Where N is the number of data points from the dataset used to build the regression model. $a_i - a_i^*$ are lagrange multipliers, which are weights used by SVR for each support vector. K(x_i, x) is a kernel function that measures the similarity between the support vector and new input data. b is a constant that helps adjust predictions to be more aligned with the actual data. x is an interaction feature between receptor-ligand. x_i is a support vector from the training dataset.

Binding Affinity Prediction by XGBoost Regressor

XGBoost is ideal for predicting molecular binding affinity due to its capability to handle complex features and non-linear interactions, where chemical and structural factors affect binding strength. Using gradient boosting, it iteratively builds decision trees, making it a powerful and efficient algorithm. Moreover, XGBoost includes optimizations such as parallel processing and regularized learning, boosting its effectiveness in predicting non-linear relationships in ligand-receptor binding affinity. The objective function includes a loss function to measure prediction error and regularization terms to control model complexity and prevent overfitting [20].

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \quad (12)$$

Where the number of trees is K = 5000, with k as the tree index, f_k (x_i) indicates the result of the k-th decision tree when applied to the input vector x_i. Each tree f_k functions by linking the input features x_i to the predicted value for the target variable.

RESULTS AND DISCUSSION

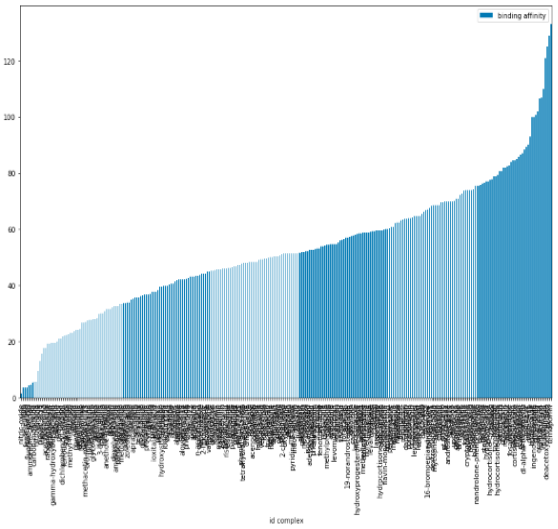
RF is deemed superior to the other algorithms as indicated by the provided Table 1, owing to its exceptional performance on the testing data. It exhibits lower MAE value of 0.55 and RMSE value of 0.78, alongside R^2 value of 0.89, in comparison to both the SVR and XGBoost. Furthermore, RF demonstrates a commendable ability to generalize from training data to testing data, revealing a minimal decline in performance and circumventing the overfitting issues encountered with the XGBoost. This algorithm also sustains an advantageous equilibrium between bias, variance and exhibits heightened resistance to overfitting as a result of its ensemble learning methodology. Consequently, the RF is identified as the preferred option within the parameters of the provided dataset.

Table 1. Comparison of the Evaluations From Training Data And Testing Data

No	Machine Learning Algorithm	Training Data			Testing Data		
		MAE	RMSE	R	MAE	RMSE	R
1	Random Forest Regressor	0,27	0,28	0,98	0,55	0,78	0,89
2	Support Vector Regressor	0,57	0,78	0,9	0,56	0,8	0,89
3	XGBoost Regressor	0,09	0,26	0,99	0,64	0,89	87

Source: (Research Result, 2024)

Figure 5 displays a graph with the values of receptor-ligand binding affinity on the y-axis using RF. The affinities shown have undergone a learning process and are arranged from smallest to largest. Each complex id represents one value of receptor-ligand binding affinity.

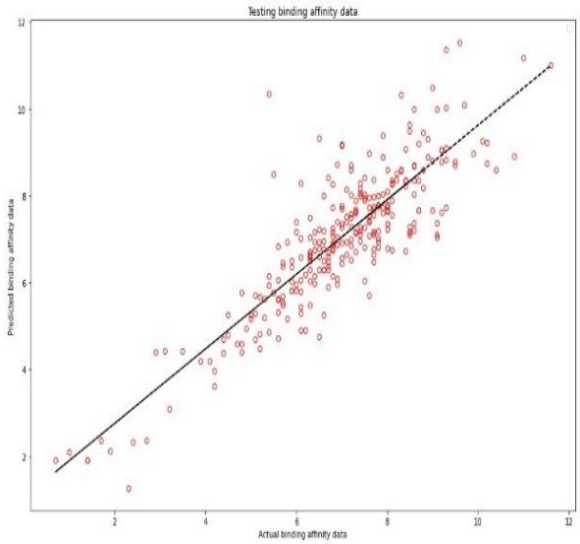


Source: (Research Result, 2024)

Figure 5. ID Complex and Receptor-Ligand Binding Affinities Generated By RF Score.

Every algorithm have different predictions on binding affinities. Although not significantly different, each model has its own unique way of evaluating data, resulting in slightly varied affinity binding values. Table 2 displays a comparison of the predicted values from the three models and the rankings produced based on the affinities. This research will provide a deeper understanding of the most suitable model in predicting binding compatibility.

Figure 6 presents the correlation between predicted binding affinity and actual binding affinity of receptor-ligand complexes. The proximity of points to the diagonal line, which represents the ideal prediction, reflects the model's accuracy in predicting binding affinity. A smaller RMSE and MAE indicate that most points are close to the line, signifying higher accuracy of the model's predictions. Additionally, the *R* provides insights into the linear correlation between predicted and actual values, with a higher *R* value indicating a stronger positive correlation. In terms of the graph, when points cluster closely along the diagonal, it indicates a high R value and confirms a strong alignment between predicted and actual binding affinity. Therefore, this visual representation successfully supports the model performance as evaluated through RMSE, MAE, and *R* metrics.



Source: (Research Result, 2024)

Figure 6. The Distribution and Correlation of Binding Affinity Before and After Predicted.

Table 2. Comparison of Binding Affinity Predictions and Rankings for Antiviral Compounds using XGBoost, RF, and SVR Models

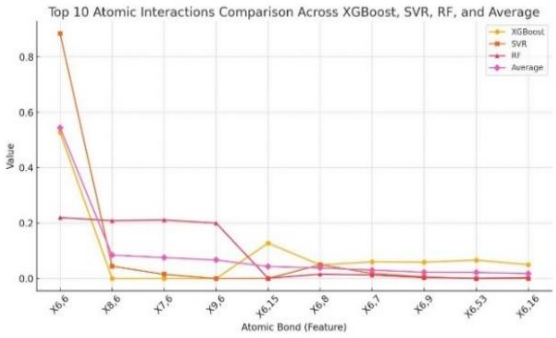
Antiviral	XGBoost		RF		SVR		Average Value of Binding Affinity	Ranking from Average Value
	Binding Affinity	Ranking	Binding Affinity	Ranking	Binding Affinity	Ranking		
Fidaxomicin	10.999245	4	10.8428	1	10.277158	1	10.706401	1
Temoporfin	11.53072	1	10.1777	4	9.662838	5	10.457086	2
Deacetoxyvinzolidine	11.17846	3	10.288517	3	9.838864	4	10.43528033	3
Gantacurium	10.34101	6	10.5264	2	10.23925	2	10.36888667	4
Jtk-853	11.358109	2	9.410833	14	8.991529	15	9.920157	5
Caspofungin	9.188359	21	9.420833	13	9.874012	3	9.494401333	6

Source: (Research Result, 2024)

Understanding how atoms connect is crucial for grasping how ligands and receptors bind together. XGBoost and RF use built-in feature importance to evaluate the impact of atom interactions, while SVR relies on permutation importance since it lacks a built-in method. This approach, which incorporates insights from all three algorithms, helps us better understand the atomic interactions that influence the binding strength of molecules.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.5994



Source: (Research Result, 2024)

Figure 7. Comparison of Feature Importance for the Top 10 Atomic Interactions in Xgboost, SVR, RF, and Average Models.

Figure 7 shows the comparison of the top 10 atomic interactions based on the feature importance of XGBoost and RF, as well as the permutation importance for SVR. The graph lines indicate that atoms interacting with X6,6 (C-C), X8,6 (O-C), and X7,6 (N-C) contribute the most to binding affinity. The peak value of the X6,6 interaction is found in each algorithm, with the C-C interaction having a value of around 0.83 for SVR, 0.51 for XGBoost, and 0.35 for RF, indicating the strength of the carbon-carbon interaction in the process. X8,6 (O-C) contributes significantly with a value of 0.18 in SVR and 0.1 in XGBoost, followed by X7,6 (N-C) with a value of 0.2 in RF and XGBoost.

When calculating feature importance and permutation importance, the X6,6 (C-C) interaction consistently ranks highest in both XGBoost, RF, and SVR. This proves that the carbon-carbon (C-C) interaction is dominant and most influential in molecular binding affinity. The collective interactions of C-C, O-C, and N-C atoms affect the strength of bonds in the target molecule, with the C-C interaction being dominant among them. To further validate these findings, laboratory tests should be conducted to assess the actual binding affinities and interactions of synthesized compounds based on these features.

CONCLUSIONS

The research identified X66 (carbon-carbon interaction) as the most dominant feature in predicting binding affinity, consistently ranking highest across all models. This highlights the significance of carbon-carbon interactions in ligand binding to the NS2B-NS3pro protein, offering insights for the future design of antiviral compounds targeting the protease. The findings shed light on the intermolecular interactions that govern binding strength, providing valuable clues for developing more effective inhibitors targeting the NS2B-NS3 protein.

Future research should focus on laboratory testing to validate binding affinity predictions from machine learning models. This includes synthesizing antiviral compounds based on dominant carbon-carbon interactions targeting the NS2B-NS3 protein and evaluating them using fluorescence binding assays and protease inhibition tests. Kinetic studies and surface plasmon resonance (SPR) methods should also be employed to assess stability and binding strength. These tests could enhance compound effectiveness and contribute to developing new antiviral therapies for dengue fever.

REFERENCE

[1] Ministry of Health of the Republic of Indonesia, “Infografis P2P,” *Pusat Promosi Kesehatan dan Pemberdayaan Masyarakat, Kementerian Kesehatan Republik Indonesia*. Accessed: Mar. 02, 2024. [Online]. Available: https://p2pm.kemkes.go.id/pages/publikasi/infografis

[2] Gauri, L. A., Bhatnagar, K., Sameja, P., Nehara, H. R., Saini, K., Dutt, R., Bhatnagar, B., Khokhar, M. S., Kumar, V., and Garg, A., "Hematological indices as predictors of mortality in dengue shock syndrome: A retrospective study," *Journal of Acute Disease*, vol. 11, no. 4, pp. 150-155, 2022.

[3] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, and E. Moreno, "AMDock: a versatile graphical tool for assisting Molecular Docking with Autodock Vina and Autodock4," *Biology Direct*, vol. 15, no. 1, pp. 1-12, 2020.

[4] J.-P. Quek et al., “Dynamic Interactions of Post Cleaved NS2B Cofactor and NS3 Protease Identified by Integrative Structural Approaches,” *Viruses*, vol. 14, no. 7, p. 1440, Jun. 2022.

[5] F. G. Saqallah, M. A. Abbas, and H. A. Wahab, "Recent advances in natural products as potential inhibitors of dengue virus with a special emphasis on NS2b/NS3 protease," *Phytochemistry*, vol. 202, p. 113362, 2022.

[6] A. Batool, N. Kanwal, M. Akram, S. Rauf, M. T. Navid, F. Masood, and L. Ali, "Molecular characterization and expression analysis of NS3 and NS4 genes of dengue virus serotype-2 from Pakistani isolates,"

Research Square, doi: 10.21203/rs.3.rs-3732242/v1, 2023.

[7] D. A. Pratiwi, R. Puspitasari, and A. Budiman, "Prediksi Ikatan Obat pada Protein NS2 Dengue dengan Machine Learning dan Docking Molekuler," *Jurnal Ilmiah Farmasi*, vol. 10, no. 1, pp. 1-10, 2020.

[8] R. Puspitasari, A. Budiman, and D. A. Pratiwi, "Deteksi Dengue dengan Analisis Antibodi NS2 dan Machine Learning," *Jurnal Biomedik*, vol. 13, no. 2, pp. 119-126, 2021.

[9] A. Budiman, D. A. Pratiwi, and R. Puspitasari, "Identifikasi Faktor Risiko Dengue dengan Analisis Protein NS2 dan Machine Learning," *Jurnal Kedokteran Tropis*, vol. 24, no. 1, pp. 45-52, 2022.

[10] J. Li and X. Wang, "Random forest-based prediction of binding affinity of dengue virus NS3 helicase inhibitors," *Journal of Chemical Information and Modeling*, vol. 60, no. 11, pp. 5360-5368, 2020.

[11] H. Feng, F. Wang, N. Li, Q. Xu, G. Zheng, X. Sun, M. Hu, G. Xing, and G. Zhang, "A random forest model for peptide classification based on virtual docking data," *Int. J. Mol. Sci.*, vol. 24, no. 14, p. 11409, 2023.

[12] J.-C. Huang, Y.-C. Tsai, P.-Y. Wu, Y.-H. Lien, C.-Y. Chien, C.-F. Kuo, J.-F. Hung, S.-C. Chen, and C.-H. Kuo, "Predictive modeling of blood pressure during hemodialysis: A comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105536, 2020.

[13] Y. Zhang and X. Chen, "Machine Learning-based prediction of binding affinity of dengue virus NS1 protein inhibitors," *Research Square*, doi: 10.21203/rs.3.rs-1117089/v1, accessed Mar. 3, 2024.

[14] B. Bargam, A. Boudhar, C. Kinnard, H. Bouamri, K. Nifa, and A. Chehbouni, "Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco," *Discover Applied Sciences*, vol. 6, no. 6, pp. 1-17, 2024.

[15] "RCSB PDB: Homepage," *RSCB PDB*. Accessed: Jul. 05, 2024. [Online]. Available: https://www.rcsb.org/

[16] V. R. Danestiara and R. Hertadi, "Prediksi Afinitas Pengikatan Kandidat Obat yang Berpotensi Menjadi Inhibitor Mpro Sars-Cov-2 Menggunakan Pendekatan Random Forest Regression," *Institut Teknologi Bandung, Bandung*, 2021.

[17] "DrugBank Online | Database for Drug and Drug Target Info," *DrugBank*. Accessed: Jul. 05, 2024. [Online]. Available: https://go.drugbank.com/

[18] V. R. Danestiara, M. Marwondo, T. A. Aziz, and I. Ismayadi, "Molecular Docking-Based Digital Screening for Dengue Virus NS2B-NS3pro Inhibitors," *EasyChair Preprint* 14490, pp. 1-6, Aug. 16, 2024.

[19] D. A. Pisner and D. M. Schnyer, "Support vector machine," in Machine learning, pp. 101-121, Academic Press, 2020.

[20] M. Amjad, I. Ahmad, M. Ahmad, P. Wróblewski, P. Kamiński, and U. Amjad, "Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation," *Appl. Sci.*, vol. 12, no. 4, p. 2126, 2022.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri