

COMPUTATIONAL QSAR-BASED MACHINE LEARNING APPROACH FOR PREDICTING ACTIVITY OF SGLT2 INHIBITORS USING THE KNIME PLATFORM

ADHA DASTU ILLAHI¹, GATOT FATWANTO HERTONO², ARRY YANUAR^{1*}

¹Laboratory of Biomedical Computation and Drug Design, Faculty of Pharmacy, Universitas Indonesia, Depok-16424, Jawa Barat, Indonesia. ²Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, Indonesia Barat, Indonesia

*Corresponding author: Arry Yanuar; *Email: arry.yanuar@ui.ac.id

Received: 07 Jun 2024, Revised and Accepted: 12 Nov 2024

ABSTRACT

Objective: This study aims to identify optimal predictive models and key molecular fragments by preparing a dataset and using machine learning techniques within the Konstanz Information Miner (KNIME) platform.

Methods: The human sodium-glucose cotransporter 2 (SGLT2) target dataset was obtained from the ChEMBL database and refined by removing salts, incomplete/incorrect data, and duplicates. The data was classified into active and inactive compounds, and fingerprints and descriptors were calculated. Christian Borgelt's Molecular Substructure Miner (MoSS) was employed to identify frequent molecular fragments. Following data partitioning, various 'classification' and 'regression' machine learning (ML) based Quantitative Structure-Activity Relationship (QSAR) models were developed and evaluated using different techniques, including sensitivity and mean Squared Error (MSE).

Results: In QSAR classification, the Support Vector Machine (SVM) model demonstrated the best performance with an accuracy of 81.66%, while in QSAR Regression, the Extreme Gradient Boosting (XGB) model exhibited the best coefficient of determination (R^2) and mean Absolute Error (MAE) values of 0.69 and 0.47 respectively. The identification of frequent Molecular Fragments highlighted common characteristics in active SGLT2 inhibitors.

Conclusion: The results of developing these QSAR models indicate that machine learning methods can be effectively used to predict SGLT2 inhibitors virtually, thereby expediting the drug discovery process.

Keywords: QSAR, SGLT2 inhibitor, Machine learning, KNIME, Artificial intelligent, In silico

© 2025 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>) DOI: <https://dx.doi.org/10.22159/ijap.2025v17i1.51726> Journal homepage: <https://innovareacademics.in/journals/index.php/ijap>

INTRODUCTION

Sodium-glucose Cotransporter 2 (SGLT2) is a protein that plays a crucial role in glucose reabsorption in the kidneys [1]. SGLT2 transports glucose from the renal tubules back into the bloodstream, thereby maintaining blood glucose levels. In individuals with type 2 diabetes mellitus (T2DM), insulin resistance leads to elevated blood glucose levels. SGLT2 inhibitors can lower blood sugar by inhibiting glucose reabsorption in the kidneys and increasing glucose excretion through urine, which helps reduce blood glucose levels and improve diabetes management [2].

The development of SGLT2 inhibitors offers several significant benefits to T2DM. Their insulin-independent mechanism of action makes SGLT2 inhibitors effective in patients with insulin resistance [3]. In addition to lowering blood glucose levels, SGLT2 inhibitors provide additional benefits such as weight loss and blood pressure reduction. Clinical studies have shown that SGLT2 inhibitors can reduce the risk of cardiovascular events and heart failure in T2DM patients [4]. Therefore, SGLT2 inhibitors are an important target in developing more effective and comprehensive diabetes therapies.

In addressing this challenge, computational approaches have become a precious tool in drug development in this field [5]. Computational pharmacology offers significant advantages in developing drugs with SGLT2 inhibitory mechanisms. One commonly used method is the quantitative Structure-Activity Relationship (QSAR). QSAR is a computational method used to predict the biological activity of a compound based on its chemical structure [6]. QSAR is an essential tool in drug discovery as it allows for the virtual screening of thousands of compounds, saving time and costs [7, 8].

The QSAR process aided by Machine Learning (ML) has significant advantages due to ML's ability to recognize complex patterns in chemical data, enabling more accurate and efficient predictions

[9–11]. One platform that can be utilized is Konstanz Information Miner (KNIME). KNIME is an open-source data analysis platform popular among computational scientists, providing an intuitive visual interface for building complex data analysis workflows, including QSAR model implementation [12]. Using KNIME, researchers can easily integrate various analytical tools and visualize the results [13].

In this study, validation was performed to develop an ML-based QSAR model to predict the activity of SGLT2 inhibitors using the KNIME platform. This research aims to produce accurate QSAR prediction models to accelerate new drug discovery and enhance efficiency in screening potential compounds by employing ML algorithms such as linear regression, decision trees, Bayesian principles, and neural networks.

MATERIALS AND METHODS

Computational methodologies

This research was conducted at the Biomedical Computation and Drug Design Laboratory, Faculty of Pharmacy, University of Indonesia. The computational aspects were facilitated by a computer system equipped with an Advanced Micro Devices (AMD) Ryzen 5900x 12-Core Processor running at GHz and 128 gigabytes (GB) of Random Access Memory (RAM), operating on Windows 10 Pro. The QSAR ML model was developed using Konstanz Information Miner (KNIME) version 4.11.

Preparation

The dataset was downloaded from the ChEMBL website, focusing on the human SGLT2 target and filtering for IC50 activity (https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL3884/) in CSV format. The data, selected from the ChEMBL database using scientific literature filters, includes 1050 compounds [14]. The data was then filtered to remove entries with missing activity values.

Molecular structures that were duplicated or used standard relations other than "=" were also removed. Additionally, salts were eliminated, and the data was classified into active and inactive compounds [15]. Compounds with IC50 values less than 21.1 nM were considered active, while those with IC50 values greater than 21.1 nM were considered inactive [16]. The activity of each molecular structure (IC50) was converted into logarithmic values in molar units [$\text{pIC}_{50} = -\log(\text{IC}_{50} \times 10^{-9})$] and then sorted from largest to smallest based on their pIC_{50} values.

Calculation of fingerprints and descriptors

The calculation of descriptors and fingerprints is conducted using three nodes in KNIME. The fingerprint options utilized are RDKit Fingerprint (Daylight-like topological fingerprint), FeatMorgan (FCFP), and RDKit Descriptor Calculation [17].

Development of ML-based QSAR model

The dataset is partitioned into 80:20 using a stratified technique based on activity values for Classification and linear sampling for the Regression Model. 80% of the data is utilized as the training set, while the remaining 20% serves as the test set or External Validation [18]. The QSAR Classification model algorithms include Multi-Layer Perceptron (MLP), Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB). Meanwhile, for Regression, RF, SVM, XGB, and Linear Regression (LR) are employed.

Evaluation of QSAR

The classification model in evaluating QSAR classification utilizes internal and external validation assessments, which include sensitivity, accuracy, F-value, and precision. The calculation of the QSAR classification evaluation is as follows:

$$\text{Sensitivity/Recall} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP}$$

$$\text{Precision (FP Rate)} = \frac{TP}{TP+FP}$$

$$\text{F-Value} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

The evaluation of QSAR regression involves assessing the training results with the pre-separated 20% data. To determine the best model performance, we analyze the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), mean signed difference (MSD), and mean absolute percentage error (MAPE) for each machine learning algorithm. The calculation of QSAR regression evaluation is as follows:

$$R^2_{\text{Pred}} = 1 - \frac{\sum(Y_{\text{Obs}} - Y_{\text{Pred}})^2}{\sum(Y_{\text{Obs}} - \bar{Y}_{\text{Training}})^2}$$

$$\text{MSE} = \frac{\sum(Y_{\text{Obs}} - Y_{\text{Pred}})^2}{n}$$

$$\text{RMSE} = \sqrt{\frac{\sum(Y_{\text{Obs}} - Y_{\text{Pred}})^2}{n}}$$

Frequent molecular fragments

Before partitioning, the dataset, which had already been classified into active and inactive compounds, was tested using

Christian Borgelt's MoSS (Molecular Substructure Miner) implementation. This testing aimed to match active compounds within the data by employing specific parameters: a minimum support threshold of 10% to ensure substructures appear in at least 10% of active compounds, a maximum sub-complement support of 5% to limit substructures appearing in more than 5% of inactive compounds, a minimum fragment size of 1, and a maximum fragment size of 100.

RESULTS AND DISCUSSION

QSAR workflow of SGLT2

The research workflow, illustrated in fig. 1, began with several key steps. First, Dataset preparation (W1) involved filtering compounds to remove duplicate, missing, and data that did not meet the requirements. Additionally, pIC_{50} values are calculated, and special labelling is applied to the compounds in the SGLT2 dataset. The next step was the Calculation of Fingerprints and descriptors (W2) using the RDKit node, where molecular fingerprints and descriptors are combined for further analysis. Following this, the dataset Partitioning (W3) is performed using stratified sampling for classification data and linear sampling for regression data. Subsequently, QSAR Classification (W4) is implemented using several ML models, including MLP, NB, RF, SVM, and XGB. The QSAR Regression (W5) models include RF, XGB, SVM, and LR. Finally, frequent molecular fragments (W6) are identified to find molecular fragments that frequently appear in the dataset. Each of these steps is carried out systematically to ensure the quality and accuracy of the research results [19].

The dataset preparation (W1) began with retrieving 1,049 compounds targeting SGLT2 from ChEMBL. The data curation process involved removing compounds without test values and those with non-standard relation types and eliminating duplicate data. Additionally, compounds containing salts were removed to ensure precise observations. As a result, 899 compounds remained, with 385 identified as active, as shown in table 1.

The acquired data were then processed to generate fingerprints and descriptors (W2). This study employed three types of nodes: RDKit Fingerprint, FeatMorgan, and RDKit Descriptor. The RDKit Fingerprint is a widely used topological fingerprint in the literature, specifically the Daylight fingerprint, which has a length of 2048 bits and is generated using the RDKit algorithm. FeatMorgan, or Extended-Connectivity Fingerprint (ECFP), also with a length of 2048 bits, was created using a more abstract and pharmacophoric set of initial atom identifiers through the circular fingerprint method. Additionally, the RDKit Descriptor provides up to 894 bits of information, including parameters such as Log P and TPSA [20].

Evaluation model QSAR classification

The development of predictive models involved partitioning the dataset (W3) into training and testing data. In QSAR classification (W4), testing was conducted using five ML models: MLP, NB, RF, SVM, and XGB, which were subsequently evaluated based on sensitivity, accuracy, F-score, and precision, as depicted in fig. 2. The results indicated that the best ML models in internal validation were RF, SVM, and XGB. Meanwhile, in external validation, the SVM model performed the best, with an accuracy of 81.66% (fig. 2). Additionally, the data suggested that MLP and NB classification models exhibited poor performance, with accuracy values less than 57.22%. These findings underscore the effectiveness of the best model in predicting SGLT2 inhibitors, which is crucial in developing effective target therapies.

Table 1: Dataset of SGLT2 inhibitor

Partition	Active	Inactive
Training set	308	411
Test Set	77	103

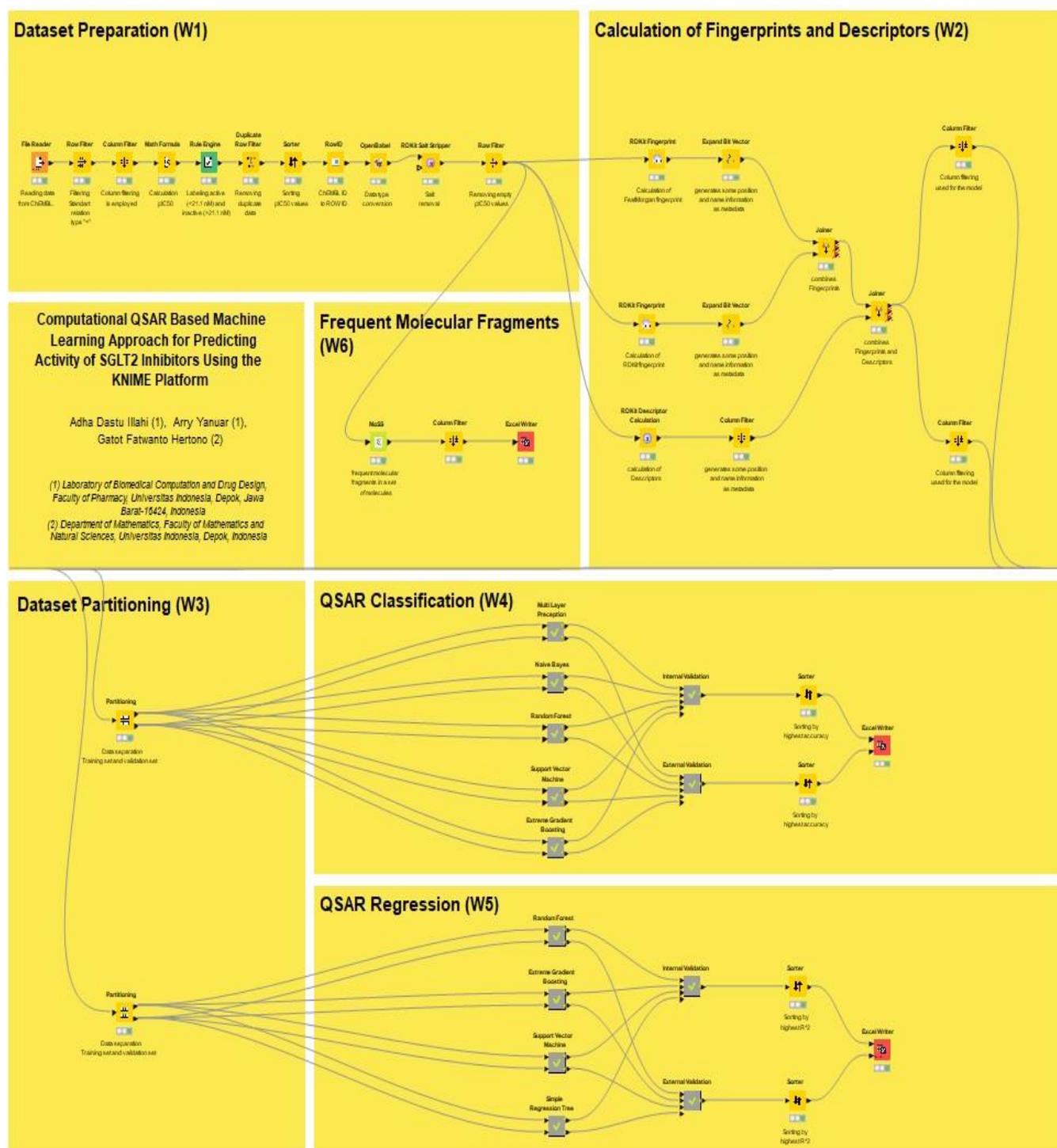


Fig. 1: QSAR SGLT2 inhibitor workflow model

Evaluation model QSAR regression

The QSAR regression model (W5) utilized four ML models: RF, XGB, SVM, and LR, as shown in table 2. The primary evaluation metrics were the coefficients of determination (R^2) and mean Absolute Error (MAE). The results revealed that the RF and XGB models exhibited superior R^2 values compared to the other models. In internal validation, the R^2 values for RF and XGB were 0.91 and 0.95, respectively, while in external validation, they were 0.67 and 0.69, respectively. These R^2 values illustrate the model's ability to explain the variability [21] (table 2).

Furthermore, the mean Absolute Error (MAE) was used to evaluate the average prediction error. The results indicated that the RF and XGB models exhibited the smallest MAE values in external validation, with values of 0.49 and 0.47, respectively. Lower MAE values indicate smaller prediction errors and better model performance. The best model in this study was determined based on external validation results, considering both R^2 and MAE values. The XGB model was selected as the best model due to its highest R^2 value and smallest MAE compared to the other models [22]. These findings suggest that the XGB model demonstrates superior predictive ability in the context of external validation.

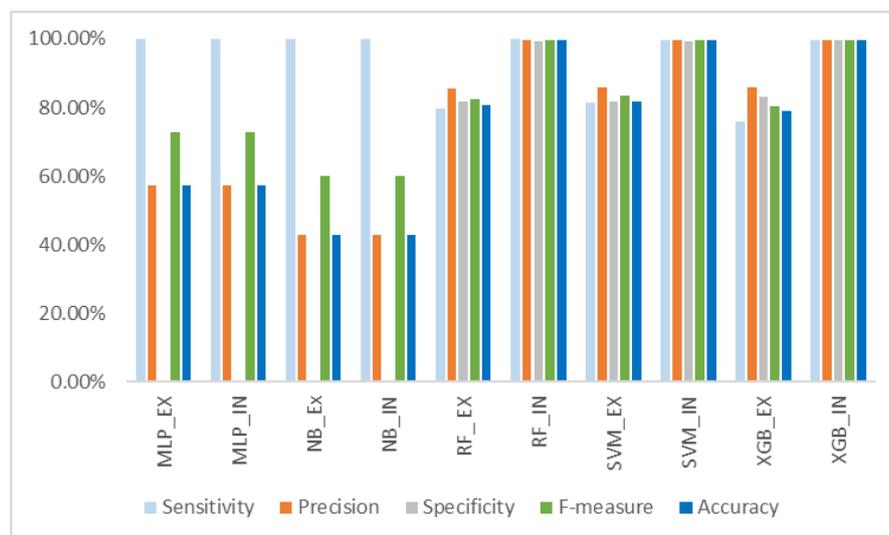


Fig. 2: Results of QSAR classification validation

Table 2: Evaluation of QSAR regression model

ML		R ²	MAE	MSE	RMSE	MSD	MAPE
RF	External	0.67	0.49	0.45	0.67	0.03	0.07
	Internal	0.95	0.19	0.07	0.27	0.01	0.03
LR	External	0.50	0.58	0.68	0.82	-0.03	0.08
	Internal	0.99	0.00	0.01	0.08	0.00	0.00
SVM	External	0.51	0.58	0.67	0.82	0.08	0.09
	Internal	0.81	0.31	0.25	0.50	0.06	0.05
XGB	External	0.69	0.47	0.43	0.66	-0.03	0.07
	Internal	0.91	0.25	0.11	0.34	-0.01	0.03

Mean absolute error (MAE), Mean squared error (MSE), Root mean squared error (RMSE), Mean signed difference (MSD) and mean absolute percentage error (MAPE)

Frequent molecular fragments

The results of the MoSS analysis indicate the frequency of fragments in the dataset. Testing was conducted by comparing active and inactive compounds based on their frequencies. One hundred thirteen fragments were obtained, with several structures displaying

the highest values in table 3. These fragments appear with higher frequency in SGLT2 inhibitors compared to non-inhibitors. The structure of SGLT2 inhibitor drugs has four main parts: 1) glucose ring, 2) central benzene ring, 3) methylene bridge, and 4) distal benzene ring, as shown in fig. 3. Some approved drugs include Dapagliflozin (IC₅₀ 1.1 nM) and Canagliflozin (2.2 nM) [23, 24].

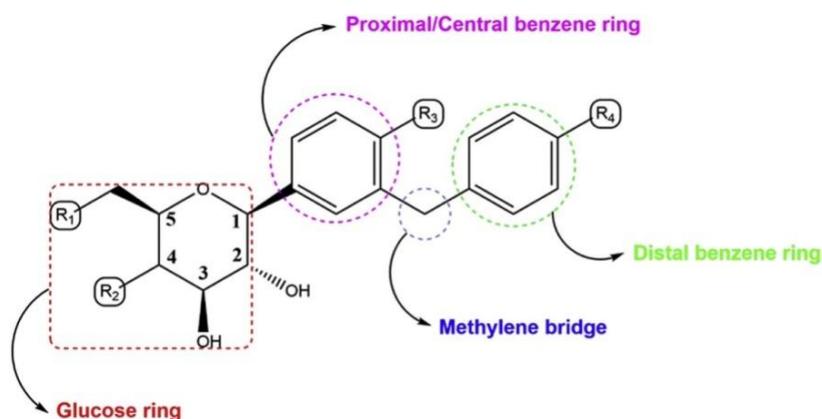
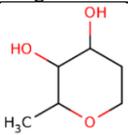
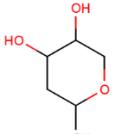
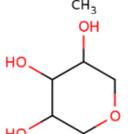
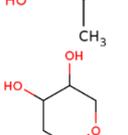
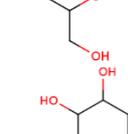
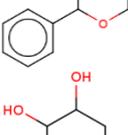
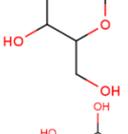
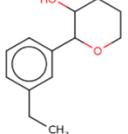
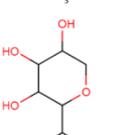
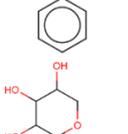


Fig. 3: Structure of SGLT2 inhibitor [23]

Dapagliflozin is the first drug for SGLT2 inhibition, featuring a glucose moiety at position 6, which binds to a central benzene ring substituted with chlorine at the para position (R₃). This central benzene ring is linked to a methylene bridge and a distal benzene ring with an ethoxy group (OCH₂CH₃) at R₄. Clinical trials have

shown that dapagliflozin can reduce HbA_{1c} levels by an average of 1.2% [25]. The MoSS analysis conducted in KNIME shows similarities between approved drugs and the MoSS findings, suggesting potential structural motifs associated with SGLT2 inhibition efficacy.

Table 3: Frequent molecular fragments

No	Fragment	Atom count	Bond count	Support in focus (abs)	Support in complement (abs)
1		336	0	0.87	0
2		336	0	0.87	0
3		318	0	0.83	0
4		284	0	0.74	0
5		266	0	0.69	0
6		266	0	0.69	0
7		265	0	0.69	0
8		248	0	0.64	0
9		247	0	0.64	0
10		236	0	0.61	0

CONCLUSION

The study used five machine learning models for QSAR Classification and four models for QSAR Regression to predict the inhibitory potential of SGLT2. The SVM model showed the highest accuracy of 81.66% in QSAR Classification. In QSAR regression, XGB showed the best R2 and MAE values, with 0.69 and 0.47, respectively, in external

validation. Using frequently occurring Molecular Fragments helped identify common characteristics in active compounds as SGLT2 inhibitors. The results suggest that machine learning methods can effectively predict SGLT2 inhibitors, which could speed up the drug discovery process. These findings could pave the way for further advancements in this field and contribute to more efficient drug discovery efforts.

FUNDING

This research was funded by The Directorate of Research and Development, Universitas Indonesia, under Hibah PUTI (Grant No. NKB-602/UN2.RST/HKP.05.00/2024).

AUTHORS CONTRIBUTIONS

Adha Dastu Illahi: Writing – original draft, Data Curation, Validation, Software, Methodology, Investigation, Conceptualization. Gatot Fatwanto Hertono: Writing – review and editing, Validation, Supervision, Methodology, Visualization. Arry Yanuar: Writing – review and editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

The manuscript was written with contributions from all authors, and all authors have approved the final version.

CONFLICT OF INTERESTS

The authors declare no conflict of interest

REFERENCES

- Guo W, Li H, Li Y, Kong W. Renal intrinsic cells remodeling in diabetic kidney disease and the regulatory effects of SGLT2 inhibitors. *Biomed Pharmacother.* 2023 Sep;165:115025. doi: [10.1016/j.biopha.2023.115025](https://doi.org/10.1016/j.biopha.2023.115025), PMID 37385209.
- Vallon V, Verma S. Effects of SGLT2 inhibitors on kidney and cardiovascular function. *Annu Rev Physiol.* 2021 Feb 10;83(1):503-28. doi: [10.1146/annurev-physiol-031620-095920](https://doi.org/10.1146/annurev-physiol-031620-095920), PMID 33197224.
- Alsereidi FR, Khashim Z, Marzook H, Gupta A, Al Rawi AM, Ramadan MM. Targeting inflammatory signaling pathways with SGLT2 inhibitors: insights into cardiovascular health and cardiac cell improvement. *Curr Probl Cardiol.* 2024 May;49(5):102524. doi: [10.1016/j.cpcardiol.2024.102524](https://doi.org/10.1016/j.cpcardiol.2024.102524), PMID 38492622.
- El Khayari A, Hakam SM, Malka G, Rochette L, El Fatimy R. New insights into the cardio-renal benefits of SGLT2 inhibitors and the coordinated role of miR-30 family. *Genes Dis.* 2024;11(6):101174. doi: [10.1016/j.gendis.2023.101174](https://doi.org/10.1016/j.gendis.2023.101174), PMID 39224109.
- Gandhi A, Masand V, Zaki ME, Al Hussain SA, Ben Ghorbal AB, Chapolikar A. QSAR analysis of sodium glucose co-transporter 2 (SGLT2) inhibitors for anti-hyperglycaemic lead development. *SAR QSAR Environ Res.* 2021 Sep 2;32(9):731-44. doi: [10.1080/1062936X.2021.1971295](https://doi.org/10.1080/1062936X.2021.1971295), PMID 34494464.
- Shah M, Patel M, Shah M, Patel M, Prajapati M. Computational transformation in drug discovery: a comprehensive study on molecular docking and quantitative structure-activity relationship (QSAR). *Intell Pharm.* 2024 Mar;2(5):589-95. doi: [10.1016/j.ipha.2024.03.001](https://doi.org/10.1016/j.ipha.2024.03.001).
- Hasan MR, Alsaiani AA, Fakhurji BZ, Molla MH, Asseri AH, Sumon MA. Application of mathematical modeling and computational tools in the modern drug design and development process. *Molecules.* 2022 Jun 29;27(13):4169. doi: [10.3390/molecules27134169](https://doi.org/10.3390/molecules27134169), PMID 35807415.
- Makhijani S. Revitalizing therapeutics: drug repurposing as a cost-effective strategy for drug development. *Int J App Pharm.* 2024 May 7;16(3):56-61. doi: [10.22159/ijap.2024v16i3.49581](https://doi.org/10.22159/ijap.2024v16i3.49581).
- Singh B, Crasto M, Ravi K, Singh S. Pharmaceutical advances: integrating artificial intelligence in QSAR combinatorial and green chemistry practices. *Intell Pharm.* 2024 May;2(5):598-608. doi: [10.1016/j.ipha.2024.05.005](https://doi.org/10.1016/j.ipha.2024.05.005).
- Pillai N, Dasgupta A, Sudsakorn S, Fretland J, Mavroudis PD. Machine learning guided early drug discovery of small molecules. *Drug Discov Today.* 2022 Aug;27(8):2209-15. doi: [10.1016/j.drudis.2022.03.017](https://doi.org/10.1016/j.drudis.2022.03.017), PMID 35364270.
- Ankith M, Surya Teja SP, Damodharan N. Artificial neural networks: functioning and applications in pharmaceutical industry. *Int J App Pharm.* 2018 Sep 8;10(5):28. doi: [10.22159/ijap.2018v10i5.28300](https://doi.org/10.22159/ijap.2018v10i5.28300).
- Berthold MR, Cebron N, Dill F, Gabriel TR, Kotter T, Meinel T. Knime the konstanz information miner. *SIGKDD Explor Newsl.* 2009 Nov 16;11(1):26-31. doi: [10.1145/1656274.1656280](https://doi.org/10.1145/1656274.1656280).
- Hermansyah O, Bustamam A, Yanuar A. Virtual screening of dipeptidyl peptidase-4 inhibitors using quantitative structure-activity relationship based artificial intelligence and molecular docking of hit compounds. *Comput Biol Chem.* 2021 Dec;95:107597. doi: [10.1016/j.compbiolchem.2021.107597](https://doi.org/10.1016/j.compbiolchem.2021.107597), PMID 34800858.
- Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D. The ChEMBL database in 2017. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D945-54. doi: [10.1093/nar/gkw1074](https://doi.org/10.1093/nar/gkw1074), PMID 27899562.
- Kausar S, Falcao AO. An automated framework for QSAR model building. *J Cheminform.* 2018 Dec 16;10(1):1. doi: [10.1186/s13321-017-0256-5](https://doi.org/10.1186/s13321-017-0256-5), PMID 29340790.
- Moinul M, Amin SA, Kumar P, Patil UK, Gajbhiye A, Jha T. Exploring sodium-glucose cotransporter (SGLT2) inhibitors with machine learning approach: a novel hope in anti-diabetes drug discovery. *J Mol Graph Model.* 2022 Mar;111:108106. doi: [10.1016/j.jmgl.2021.108106](https://doi.org/10.1016/j.jmgl.2021.108106), PMID 34923429.
- Beisken S, Meinel T, Wiswedel B, DE Figueiredo LF, Berthold M, Steinbeck C. Knime CDK: workflow driven cheminformatics. *BMC Bioinformatics.* 2013 Dec 22;14(1):257. doi: [10.1186/1471-2105-14-257](https://doi.org/10.1186/1471-2105-14-257), PMID 24103053.
- Myint KZ, Wang L, Tong Q, Xie XQ. Molecular fingerprint-based artificial neural networks QSAR for ligand biological activity predictions. *Mol Pharm.* 2012 Oct 1;9(10):2912-23. doi: [10.1021/mp300237z](https://doi.org/10.1021/mp300237z), PMID 22937990.
- Carracedo Reboredo P, Linares Blanco J, Rodriguez Fernandez N, Cedron F, Novoa FJ, Carballal A. A review on machine learning approaches and trends in drug discovery. *Comput Struct Biotechnol J.* 2021;19:4538-58. doi: [10.1016/j.csbj.2021.08.011](https://doi.org/10.1016/j.csbj.2021.08.011), PMID 34471498.
- Yang J, Cai Y, Zhao K, Xie H, Chen X. Concepts and applications of chemical fingerprint for hit and lead screening. *Drug Discov Today.* 2022 Nov;27(11):103356. doi: [10.1016/j.drudis.2022.103356](https://doi.org/10.1016/j.drudis.2022.103356), PMID 36113834.
- Veerasingh R, Rajak H, Jain A, Sivadasan S. Validation of QSAR models-strategies and importance. *Int J Drug Des Discov.* 2011;2(3):511-9.
- Roy K, Editor. *Advances in QSAR modeling.* Berlin: Springer International Publishing: Vol. 24. Challenges and advances in computational chemistry and physics; 2017.
- Bhattacharya S, Rathore A, Parwani D, Mallick C, Asati V, Agarwal S. An exhaustive perspective on structural insights of SGLT2 inhibitors: a novel class of antidiabetic agent. *Eur J Med Chem.* 2020 Oct;204:112523. doi: [10.1016/j.ejmech.2020.112523](https://doi.org/10.1016/j.ejmech.2020.112523), PMID 32717480.
- Ramani J, Shah H, Vyas VK, Sharma M. A review on the medicinal chemistry of sodium-glucose co-transporter 2 inhibitors (SGLT2-I): update from 2010 to present. *Eur J Med Chem Rep.* 2022 Dec;6:100074. doi: [10.1016/j.ejmcr.2022.100074](https://doi.org/10.1016/j.ejmcr.2022.100074).
- Hussain M, Atif M, Babar M, Akhtar L. Comparison of efficacy and safety profile of empagliflozin versus dapagliflozin as add on therapy in type 2 diabetic patients. *J Ayub Med Coll Abbottabad.* 2021;33(4):593-7. PMID 35124914.