Pharmaceutical Sciences and Drug Design
ISSN: 3062-4428
2021, Volume 1, Page No: 164-178

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Galaxy Publication

A Basic Machine Learning Approach for predicting pIC50 Inhibition Values of FLT3 Tyrosine Kinase Using Quantitative Structure–Activity Relationship Modeling

Lukas Schneider[1], Anna K. Müller[1]*, Felix Braun[1]

[1]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Heidelberg University, Heidelberg, Germany.

***E-mail** ✉ anna.mueller@outlook.com

Received: 19 August 2021; **Revised:** 25 November 2021; **Accepted:** 26 November 2021

ABSTRACT

Acute myeloid leukemia (AML) remains a major therapeutic hurdle, especially in patients with FLT3 tyrosine kinase mutations. The goal of this work was to create a reliable and accessible machine learning-driven quantitative structure–activity relationship (QSAR) model capable of forecasting the inhibitory activity (expressed as pIC50) of FLT3 inhibitors, overcoming the shortcomings of earlier models related to limited dataset scale, chemical diversity, and forecasting precision. A substantially expanded dataset—approximately 14-fold larger than those used in previous investigations (comprising 1350 molecules and 1269 descriptors)—was utilized to train a random forest regression model, selected for its outstanding performance and robustness against overfitting. Thorough internal assessment through leave-one-out and 10-fold cross-validation produced Q^2 values of 0.926 and 0.922, respectively. External testing on a separate set of 270 compounds achieved an R^2 of 0.941 with a standard error of 0.237.Critical molecular features governing inhibitory strength were pinpointed, enhancing understanding of the essential structural elements. Furthermore, an intuitive computational platform was built to allow quick estimation of pIC50 values and support ligand-based virtual screening, which successfully highlighted several potential FLT3 inhibitors. This study marks a notable progress in FLT3 inhibitor research by providing a dependable, practical, and streamlined method for initial drug discovery phases, with the potential to expedite the development of precision treatments for AML.

Keywords: AML treatment, FLT3 inhibitors, QSAR modeling, Ligand-based drug design

How to Cite This Article: Schneider L, Müller AK, Braun F. A Basic Machine Learning Approach for predicting pIC50 Inhibition Values of FLT3 Tyrosine Kinase Using Quantitative Structure–Activity Relationship Modeling. Pharm Sci Drug Des. 2021;1:164-78. https://doi.org/10.51847/qNHVgLpYw0

Introduction

Acute myeloid leukemia (AML) poses a major obstacle in cancer treatment, marked by the rapid and uncontrolled growth of abnormal clonal cells in the blood-forming system, resulting in widespread tissue invasion and advancing illness. Central to AML's underlying mechanisms is the Fms-like tyrosine kinase 3 (FLT3) receptor gene [1]. Mutations in this gene, especially internal tandem duplications (ITDs), provide leukemic cells with a strong growth edge by triggering various signaling cascades essential for disease advancement and patient prognosis. As a result, FLT3 mutations play a critical role in the aggressive behavior of AML and have been thoroughly investigated for their contributions to pathogenesis and clinical outcomes [2-4].

For many years, standard AML therapy has centered on a classic protocol involving prolonged cytarabine infusion paired with anthracyclines [5]. The success of this established regimen depends on elements like the leukemia's genetic makeup and patient age, where elderly individuals frequently show considerably reduced response rates. This has highlighted the pressing demand for innovative treatment options that can boost results across diverse patient populations [3]. Accordingly, AML management has progressed with the approval of targeted FLT3 inhibitors including midostaurin, gilteritinib, sorafenib, and quizartinib. This move toward precision medicine, incorporating newer regimens such as CPX-351 and gemtuzumab ozogamicin, demonstrates continuous attempts to tailor therapies to the disease's molecular profile [6]. The emergence of these targeted agents and the broadening

of therapeutic options represent a meaningful advance in delivering more personalized and effective AML care, raising prospects for better survival in varied patient cohorts [7-9].

Progress in comprehending and designing FLT3 inhibitors has been substantially aided by quantitative structure–activity relationship (QSAR) studies and molecular docking techniques [10-16]. Sandoval *et al*. [16] illustrated the application of QSAR through linear discriminant analysis and multilinear regression to precisely forecast the antileukemic potential of various compounds. In a similar vein, Shih and Bhujbal *et al*. [11, 13] combined molecular docking with 3D-QSAR methods to pinpoint vital structural attributes and propose new molecules with superior FLT3 inhibition. Ghosh *et al*. [15] showcased the value of computational tools, such as molecular dynamics simulations paired with 3D-QSAR, in clarifying structure–activity correlations for FLT3 inhibitors. These strategies, supported by work from Fernandes and Islam *et al*. [14, 17], have yielded essential knowledge about molecular binding interactions and affinities of candidate FLT3 inhibitors, reinforcing the importance of such computational methods in AML drug research and design.

More recently, machine learning (ML) has emerged as a powerful tool in classifying AML subtypes, demonstrating versatile uses and impressive diagnostic precision. Liu *et al*. [18] built a random forest classifier for automatically distinguishing AML-M1 and M2 subtypes from bone marrow smear images, attaining 99.8% accuracy. Abhishek *et al*. [19] employed deep learning for AML classification among other leukemias, securing 97% accuracy in binary classification and 95% in multiclass scenarios. Monaghan *et al*. [20] used ML on flow cytometry datasets, reaching 94.2% accuracy in separating acute leukemias from non-malignant cytopenias. Awada *et al*. [21] incorporated genomic information via Bayesian latent class models to uncover new AML molecular subgroups, with 97% accuracy in cross-validation. Dese *et al*. [22] applied support vector machines for effective image segmentation and leukemia identification, delivering 97.69% accuracy and cutting diagnostic time from 30 minutes to less than one minute. Talaat *et al*. [23] refined convolutional neural networks (CNNs) through hyperparameter tuning for leukemia detection, achieving 99.99% accuracy. Collectively, these efforts highlight ML's disruptive impact on AML diagnosis, augmenting conventional techniques and enabling more streamlined clinical processes.

Extending these developments, ML has transformed drug discovery, especially in detecting and forecasting kinase inhibitors, including FLT3-targeted ones. Nasimian *et al*. [24] showed how an ML-driven framework could forecast drug responses, uncovering important details about AXL reliance in AML. Janssen *et al*. [25] presented the Drug Discovery Maps (DDM) framework, utilizing algorithms like t-SNE for visualizing and anticipating kinase interactions, which facilitated the identification of highly effective FLT3 inhibitors. Zhao *et al*. [26] implemented ML techniques to categorize and examine structure–activity patterns across a large collection of FLT3 inhibitors, revealing essential structural elements linked to potent inhibition. As reviewed by Eckardt *et al*. [27], these innovations emphasize ML's increasing role in AML management, spanning diagnostics to treatment refinement. Such ML-integrated strategies hold great promise for improving the accuracy and efficiency of FLT3 inhibitor design, marking a fundamental change in AML therapeutic approaches.

Although considerable progress has been made in FLT3 inhibitor research, lingering issues remain, particularly regarding the forecasting reliability of existing QSAR models. These often display restricted accuracy and highlight the demand for greater precision alongside ease of use. A frequent drawback is their dependence on constrained molecular datasets during training, which restricts broad applicability across diverse chemical structures. Insufficient variety and uniformity in training data impair their effectiveness in addressing the full spectrum of possible FLT3 inhibitors. Moreover, the scarcity of accessible, intuitive tools that deliver fast and dependable predictions underscores the requirement for more applicable and robust solutions in drug development.

To address these gaps, the present study presents a novel ML-enhanced QSAR model developed using a larger and more varied compound collection, promoting improved stability and wider extrapolation. By combining cutting-edge machine learning algorithms with detailed molecular descriptors, this model seeks to overcome the shortcomings of prior QSAR efforts. It is also engineered for straightforward use, providing swift and consistent results. This framework is poised to advance the discovery of novel FLT3 inhibitors for AML therapy, establishing a benchmark for more effective and approachable tools in pharmaceutical research. In the long term, it could reshape FLT3 inhibitor creation and hasten advancements in tailored AML therapies.

Materials and Methods

Data curation

Information on compounds inhibiting FLT3, along with their reported IC50 data, was collected from the PubChem repository [28, 29] via the Requests package [30] and arranged in a structured table using Pandas [31] within Python 3. The compiled dataset was thoroughly refined by eliminating redundant records. Emphasis was placed on selecting compounds exhibiting IC50 below 10 μM to emphasize those with greater potency.

During the final preparation phase, feature standardization was applied to normalize the scales of input variables, optimizing suitability for the ANN methods. This procedure was vital for preserving dataset quality and compatibility with ANN-K and ANN-P, given their susceptibility to variations in input scaling. Standardization was achieved using the StandardScaler from scikit-learn [32], which centers features around zero mean and unit variance. The fit_transform operation was used on training samples to derive mean and standard deviation values, which were subsequently applied to the test samples via transform, preventing any leakage of test information into the model and upholding rigorous statistical standards.

Molecular descriptor calculation

An initial set of 1511 numeric molecular descriptors was generated employing PaDEL-Descriptor version 2.21 [33] alongside RDKit [34]. Subsequent refinement removed descriptors that were inapplicable to the full set of compounds or exhibited no variation, yielding a final count of 1269 descriptors. This filtering step was critical to guarantee uniform descriptor availability across all molecules, preserving only those relevant to the structural variety and appropriate for robust machine learning applications.

Benchmarking machine learning methods with external validation

The assembled dataset, including 1350 molecules and 1269 descriptors, was loaded in Python 3 with assistance from Pandas [31]. Target values were derived from experimental pIC50 measurements. For equitable distribution, the data was divided into training and testing portions in an 80:20 proportion via the train_test_split utility in sklearn [32], fixing random_state at 11 to ensure repeatability.

The evaluated machine learning approaches, executed through sklearn [32], encompassed random forest regression (RFR) [35], gradient boosting regression (GBR) [36], support vector regression (SVM) [37], kernel ridge regression (KRR) [37], Gaussian process regression (GPR) [38], and bagging regression with random forests (BRF) [39]. In addition, artificial neural network frameworks were constructed using Keras 2.13.1 (ANN-K) [40] and replicated in PyTorch 2.4.0 (ANN-P) [41]. Random state parameters were uniformly applied where relevant. Default settings were retained for the conventional machine learning algorithms, whereas the ANN underwent targeted hyperparameter tuning for best results.

ANN architecture

The neural network was structured as a sequential setup with three dense layers: an initial layer of 500 units to accommodate the extensive feature set, a hidden layer of 5 units for feature abstraction, and an output layer with one unit for pIC50 regression. ReLU activation was applied to the first and second layers, while linear activation served the output; HeNormal initialization was used for weights. Input normalization relied on StandardScaler from sklearn, with training conducted using a batch size of 10 across 100 epochs to promote effective learning while mitigating overfitting risks.

Hyperparameter selections for the ANN prioritized practicality and resource efficiency to enable equitable benchmarking against other techniques. The tuned ranges covered layer1_sizes = [100, 300, 500], layer2_sizes = [1, 3, 5, 10, 15], epochs_list = [20, 50, 100, 120, 150], and batch_sizes = [5, 10, 20, 40]. These choices balanced network depth with computational demands, supporting efficient handling of large-scale QSAR data.

Assessment of model performance and external validation

The effectiveness of the models was gauged through several statistical measures, including the coefficient of determination (R^2), mean absolute error (MAE), standard deviation (SD), and root mean squared error (RMSE), applied to both training and testing sets. These calculations were performed leveraging functions from the sklearn.metrics package, providing dependable evaluation capabilities. The test sets were particularly employed for external validation purposes, offering an in-depth perspective on predictive reliability and error profiles across the various models, with reproducibility prioritized throughout the process. This was ensured by applying fixed

random seeds (value of 11) uniformly to numpy, TensorFlow-Keras, PyTorch, and sklearn components, guaranteeing consistent and dependable performance evaluations.

Refinement of model components via feature selection

Evaluation of individual descriptors

To determine the impact of each molecular descriptor on forecasting FLT3 inhibitor potency, assessments were carried out under the same conditions (80:20 train-test division, random state = 11). Individual descriptors were tested using the random forest regressor (RFR), previously determined as the top-performing approach from benchmarking. The focus was on the test set coefficient of determination (R^2 test), which served as a key indicator of descriptor significance by directly associating it with improvements in predictive precision.

Descriptor analysis and selection procedure

Subsequently, the leading 100 descriptors were scrutinized based on their R^2 test values to explore their associations with FLT3 inhibitory potency. This examination guided a gradual incorporation strategy, beginning with the highest-ranked descriptor and sequentially including those with lower rankings. The objective was to identify the ideal combination that optimized predictive performance while controlling model complexity.

Internal validation procedures

Following the initial benchmarking and feature refinement stages, the chosen optimal model was subjected to internal validation employing leave-one-out and 10-fold cross-validation methods. The leave-one-out approach, executed through the LeaveOneOut function in sklearn.model_selection, trains the model on all but one sample, using the excluded point for validation, repeating this for every instance. Alternatively, 10-fold cross-validation, via the KFold function from the same library, partitions the data into 10 groups, training on nine and validating on the held-out one, cycling through all partitions.

In these validation steps, model reliability was measured primarily with the R^2 statistic (denoted Q^2_LOO for leave-one-out and Q^2_10-fold for the 10-fold variant). These metrics allowed direct comparisons with earlier research on similar topics, confirming the model's strength independent of the specific data split.

Ligand-based virtual screening approach

To identify novel prospective inhibitors of FLT3 tyrosine kinase, a virtual screening effort was undertaken utilizing the PubChem resource [28, 29]. The screening involved assessing structural resemblance to the 100 most potent compounds from the dataset, evaluating each reference compound separately. Specifically, every one of these top 100 actives was matched against approximately 10.2 million entries in PubChem [29, 42] via the Tanimoto similarity index [43], applying a cutoff of 90%. This procedure relied on the requests package and PubChem's PUG-REST similarity functionality. Resulting SMILES strings were then refined by excluding any established FLT3 tyrosine kinase inhibitors.

The refined SMILES collection was inputted into the predictive script for pIC50 estimation, enabling the ranking and selection of the five most favorable candidates. This streamlined prioritization supports subsequent experimental testing and expedites the identification of effective FLT3 tyrosine kinase inhibitors.

Results and Discussion

Molecular diversity of the dataset

In order to assess the chemical variety within the dataset, clustering was conducted with RDKit [34] by generating MACCS key fingerprints [44] for every compound. The chosen clustering method was Butina [45], employing a Tanimoto similarity cutoff of 0.3, meaning that compounds grouped together exhibited a similarity score of no less than 0.7. The arrangement of compounds into clusters is depicted in **Figure 1.**

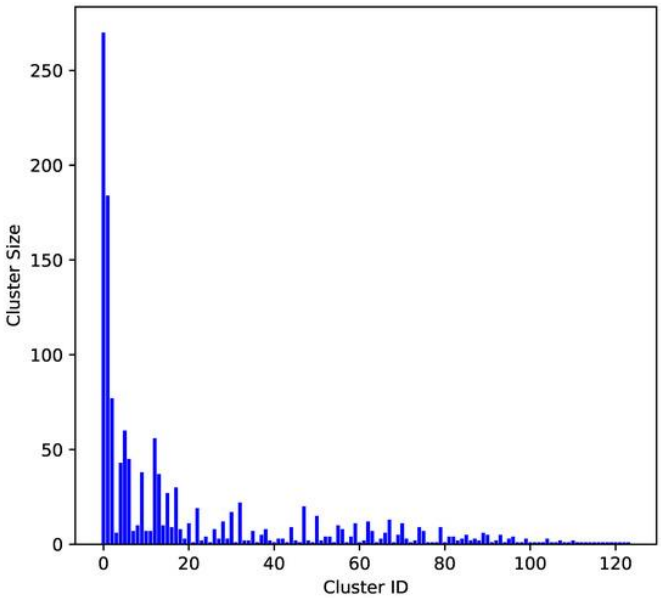


Figure 1. Analysis of molecular diversity through clustering.

Figure 1 displays the outcomes of clustering the compound dataset using MACCS key fingerprints and the Butina clustering method [44, 45]. The horizontal axis corresponds to cluster identifiers, whereas the vertical axis indicates the count of compounds within each cluster. Clusters of greater size reflect collections of compounds sharing substantial structural resemblance, pointing to areas of redundancy in the data. In contrast, smaller clusters highlight more distinctive chemical entities, signifying higher levels of diversity.

The clustering revealed an equitable mix of structural similarities and differences, demonstrating that the dataset included both closely related groups and distinctly varied compounds. The most populated cluster contained 20% of all molecules, followed by the next largest cluster with 13.6%. All other clusters individually represented less than 6% of the total compounds. In total, the dataset formed 124 separate clusters—a number substantially exceeding the molecule counts utilized in earlier investigations [10-15]. This outcome underscores the markedly greater chemical diversity incorporated in the current study relative to previous efforts, thereby providing an expanded structural landscape for examination and model construction. Such enhanced diversity is essential for building reliable and broadly applicable machine learning frameworks capable of accurately forecasting FLT3 inhibitor activity in the context of AML therapy.

Comparison of machine learning approaches

The present investigation assessed the capabilities of several machine learning algorithms trained on the same dataset to estimate pIC50 values for 1350 FLT3 tyrosine kinase inhibitors, employing 1269 molecular descriptors. The algorithms under comparison encompassed random forest regression (RFR), gradient boosting regression (GBR), kernel ridge regression (KRR), Gaussian process regression (GPR), bagging regression with random forest (BRF), as well as two artificial neural network configurations developed with Keras (ANN-K) and PyTorch (ANN-P).

Table 1 provides a detailed side-by-side evaluation of these machine learning techniques in forecasting pIC50 values for FLT3 tyrosine kinase inhibitors, reporting key performance indicators such as R^2, MAE, SD, and RMSE for both training and testing sets.

Table 1. Performance comparison of machine learning models for predicting pIC_{50} values of FLT3 tyrosine kinase inhibitor compounds.

Metric and ML	ANN-P	ANN-K	BRF	GPR	KRR	GBR	RFR
R^2 training	0.983	0.988	0.967	0.641	0.546	0.973	0.988
MAE training	0.082	0.070	0.136	0.469	0.489	0.126	0.082
SD training	0.121	0.101	0.172	0.526	0.638	0.154	0.102
RMSE training	0.123	0.103	0.172	0.568	0.638	0.154	0.102
R^2 test	0.895	0.907	0.931	−0.228	0.592	0.939	0.936

MAE test	0.248	0.235	0.207	0.876	0.484	0.195	0.197
SD test	0.313	0.296	0.255	0.932	0.619	0.237	0.246
RMSE test	0.315	0.297	0.256	1.076	0.620	0.239	0.246

Overview of model performance

Performance on training data

The effectiveness of the various machine learning models during training (evaluated on 1080 compounds) was assessed using key metrics including R^2, MAE, SD, and RMSE **(Table 1).** The random forest regressor (RFR) and the Keras-based artificial neural network (ANN-K) achieved the highest R^2 values of 0.988, demonstrating outstanding fit to the data. These models also displayed minimal errors, with MAE, SD, and RMSE values of 0.082, 0.102, and 0.102 for RFR, and 0.070, 0.101, and 0.103 for ANN-K, respectively, reflecting their strong capacity to explain and reproduce the variance in the training set.

The gradient boosting regressor (GBR) performed well overall, attaining an R^2 of 0.973, but showed marginally higher error metrics (MAE of 0.126, SD and RMSE both at 0.154) compared with RFR and ANN-K **(Table 1)**. This suggests solid predictive ability, though with slightly less consistency in closeness to observed values.

In comparison, kernel ridge regression (KRR) and Gaussian process regression (GPR) yielded considerably lower R^2 scores of 0.546 and 0.641, respectively **(Table 1)**. Their elevated error values (MAE, SD, and RMSE of 0.489, 0.638, and 0.638 for KRR; 0.469, 0.526, and 0.568 for GPR) indicate reduced accuracy and greater scatter in predictions.

Taken together, RFR and ANN-K emerged as the most reliable options for tasks demanding high precision, while GBR offers a reasonable alternative when minor trade-offs in accuracy are tolerable. KRR and GPR, however, may benefit from additional optimization or may be less suitable without further adjustments. These findings emphasize the need to choose models aligned with required performance standards and application contexts.

Performance on testing data

Evaluation of the models on an independent external set (270 compounds not used in training) revealed notable differences in generalization ability, using the same metrics **(Table 1)**. Both RFR and GBR delivered excellent results, with R^2 values approaching 0.94, accompanied by low MAEs (0.197 for RFR; 0.195 for GBR) and RMSEs (0.246 for RFR; 0.239 for GBR). These outcomes confirm the strength of tree-based ensemble techniques in maintaining accuracy on unseen data when applied to QSAR prediction of pIC50 for FLT3 inhibitors.

The GPR, however, performed poorly on the test set, producing a negative R^2 of −0.228—worse than a baseline mean predictor—along with substantially elevated MAE and RMSE. This points to severe overfitting during training or unsuitable assumptions for the dataset, rendering it impractical for real-world use.

Kernel ridge regression achieved a moderate R^2 of 0.592 **(Table 1)**, outperforming GPR but lagging behind RFR and GBR, indicating partial capture of data patterns without comparable reliability. The neural network models (ANN-K and ANN-P) experienced sharp drops in performance from training to testing, with R^2 values of 0.907 and 0.895, respectively, and increased errors, highlighting classic overfitting issues. Although ANNs excel at complex nonlinear relationships, they are vulnerable to overfitting in noisy or highly correlated feature sets, necessitating robust regularization. By contrast, RFR mitigates such risks effectively through ensemble averaging and random feature subsetting, contributing to its superior generalization.

These test results reinforce the value of prioritizing models that balance training fit with strong extrapolation to novel compounds. RFR and GBR proved most dependable for routine deployment, whereas GPR, KRR, and the ANN approaches may demand extra precautions or modifications to achieve acceptable robustness.

Description of the selected model

This subsection examines the optimization process via feature selection for the chosen random forest regressor (RFR), which was selected for forecasting FLT3 tyrosine kinase inhibitor potency. As depicted in **Figure 2,** the top five molecular descriptors alone contributed substantially, yielding a test R^2 of 0.893 and underscoring their dominant influence on predictive power and model explainability (detailed further in the "Model Interpretation" section). Adding descriptors progressively up to the ninth raised the test R^2 to 0.930, with clear gains that subsequently tapered off. After including 41 descriptors, the test R^2 stabilized at 0.941, indicating that incorporating additional features provided negligible further improvement in prediction accuracy.

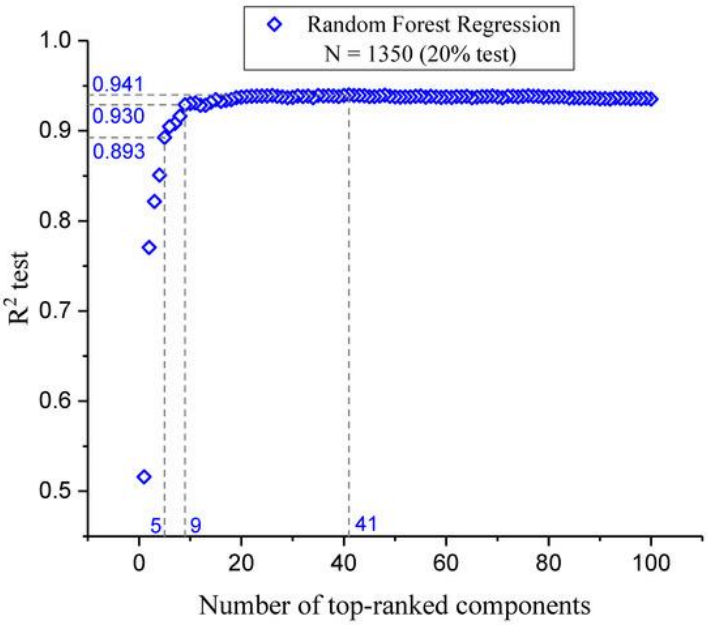


Figure 2. Changes in test R^2 scores with increasing numbers of descriptors, ordered by decreasing importance.

A key element of this evaluation involves examining the relationships between the selected descriptors, as shown in the correlation matrix presented in **Figure 3.** This matrix illustrates the pairwise Pearson correlation coefficients among the 41 chosen descriptors, as well as their individual correlations with the target inhibitory potency (pIC50). The color scale spans from −1 (deep blue), representing strong negative correlations, to +1 (deep red), denoting strong positive correlations, while values near zero appear in white. Highly correlated descriptor pairs were defined using an absolute correlation threshold of 0.90 and were eliminated prior to final model building. Consequently, all correlations displayed in **Figure 3** are below |0.90| in magnitude. The detection and exclusion of such strongly intercorrelated features is vital because they can introduce redundant information, potentially compromising model stability and interpretability. By ensuring low intercorrelation among the retained descriptors, each contributes distinct and independent information, thereby enhancing the overall reliability and predictive strength of the random forest regressor model.

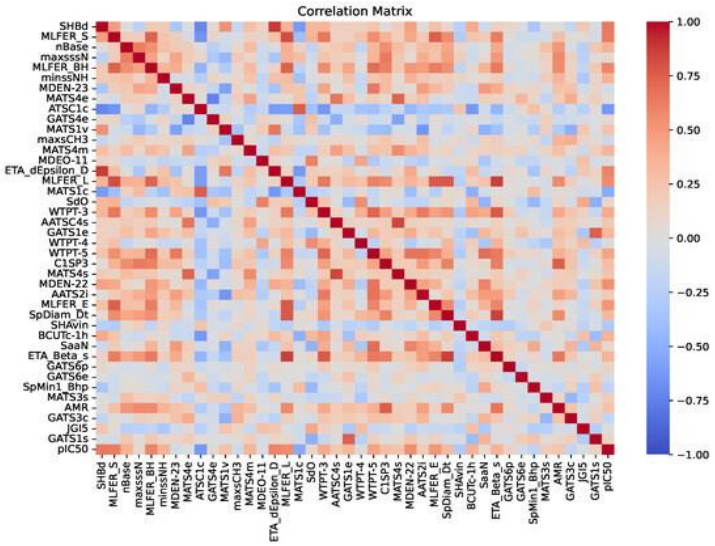
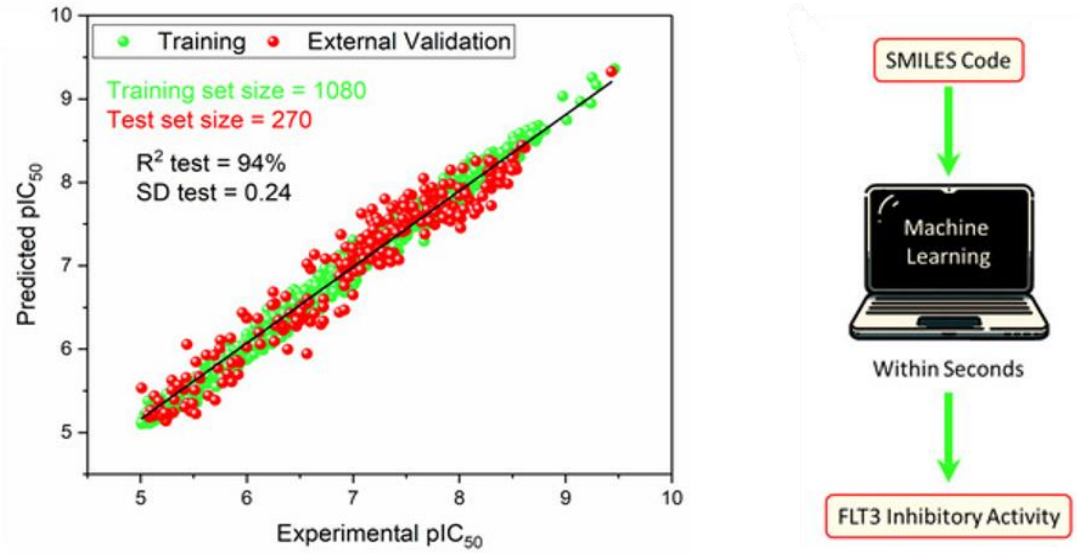


Figure 3. Heatmap depicting correlations among the 41 selected descriptors and the target variable (pIC50). The visualization presents pairwise Pearson correlation coefficients between descriptors and their associations with inhibitory potency (pIC50). The color gradient extends from −1 (deep blue) for strong negative correlations to +1 (deep red) for strong positive correlations, with values close to zero shown in white.

These findings highlight the strength of the random forest regressor (RFR) in modeling intricate nonlinear associations between a limited set of descriptors and pIC50 values, thereby achieving an optimal trade-off between model complexity and forecasting precision. This reinforces the advantages of ensemble approaches such as RFR for managing high-dimensional datasets [46] and stresses the critical role of careful feature selection in constructing effective and dependable predictive tools for drug discovery.

The performance metrics summarized in **Table 2** and visualized in **Figure 4a** illustrate the robust predictive power of the optimized RFR model, built using 41 descriptors, for FLT3 tyrosine kinase inhibitors. The model delivered an impressive R^2 of 0.989 on the training set and 0.941 on the test set, demonstrating superior accuracy and stability in relating molecular descriptors to pIC50 values across 270 compounds. Supporting error measures, including MAE, SD, and RMSE, confirmed high precision in both training and testing phases. The QLOO2 value of 0.926 and Q10fold2 value of 0.922 indicate notable predictive reliability through cross-validation, emphasizing consistency in the model.



a) b)

Figure 4. (a) Scatter diagram demonstrating the effectiveness of the random forest regressor (RFR) model. The model was developed using 1080 compounds (represented by green dots) and validated externally on 270 compounds (shown as red dots). The graph depicts the relationship between predicted and observed pIC50 values for inhibitors of FLT3 tyrosine kinase. (b) Screenshot of the user-friendly application created for estimating pIC50 values of potential FLT3 tyrosine kinase inhibitors.

Table 2. Evaluation of Random Forest Models for predicting pIC50 Values of FLT3 Tyrosine Kinase Inhibitor Compounds Using 41 Components.

Metric	Test Set	Training Set
R^2	0.941	0.989
Size	270	1080
MAE	0.193	0.081
Standard Deviation (SD)	0.237	0.101
RMSE	0.238	0.101
Q10-fold²	0.922	
QLOO²	0.926	

Comparison with previous QSAR models

The random forest regressor (RFR) model, constructed using 41 selected descriptors, was benchmarked against earlier QSAR investigations focused on FLT3 tyrosine kinase inhibitors, with results summarized in **Table 3**. This model substantially outperformed prior efforts in forecasting pIC50 values for novel compounds, delivering a test-set R^2 of 0.941 and a standard deviation of 0.237. In comparison, previous works reported maximum R^2 values of

0.891 and minimum standard deviations of 0.3. The superior results achieved here demonstrate not only greater predictive precision but also the benefits of training and validating on a larger and more chemically diverse collection of 270 compounds, which strengthens the model's capacity to reliably estimate FLT3 tyrosine kinase inhibitory activity across a wider structural space. Additionally, the leave-one-out cross-validation Q^2 of 0.926—considerably higher than the 0.802 or lower values obtained in prior studies—indicates reduced sensitivity to individual data points or specific descriptors, highlighting enhanced robustness relative to earlier QSAR approaches.

Table 3. Performance comparison of QSAR models developed for FLT3 inhibitors.

Author (Year)	Kar (2012)[a]	Shih (2012)[a]	Abutayeh (2019)[a]	Bhujbal (2020)[a]	Fernandes (2020)[a]	Ghosh (2021)[a]	This Work
Train set size	51	25	76	45	28	30	1080
Dataset size	67	72	93	63	40	40	1350
Test set size	16	47	17	18	12	10	270
R^2 test	0.891	0.76	0.57	0.707	0.80	0.698	0.941
R^2 training	0.956	0.98	0.86	0.956	0.80	0.983	0.989
SD test	0.435	0.66	-	>0.895	0.31	0.452	0.237
Q^2_{LOO}	0.747	0.58	0.65	0.57	0.60	0.802	0.926

[a] Data obtained from [10-15].

These results highlight the strength of a purely ligand-based strategy when underpinned by an extensive and chemically varied dataset, establishing this approach as a highly practical and trustworthy instrument for drug design.

Interpretation of the Model

Model interpretability was attained through a detailed conceptual examination of the five descriptors that exerted the greatest influence on predictive performance. These top-ranked descriptors, presented in **Table 4**, are SHBdb, MLFER_S, nBase, MaxsssN, and MLFER_BH, each recognized for their critical contributions to the model's accuracy.

Table 4. Identification and description of the five most influential descriptors contributing to the model, ranked by importance.

Priority	Name	Descriptor	Description
5°	Overall solute hydrogen bond basicity	MLFER_BH [47, 48]	Total hydrogen bond basicity of the molecule, obtained by summing the contributions from all potential hydrogen bond acceptor sites.
4°	Maximum atom-type E-state: >N–	MaxsssN [49, 50]	Highest electrotopological state value among nitrogen atoms bearing three single bonds.
3°	Number of basic groups	nBase	Count of basic functional groups in the molecule, primarily nitrogen-containing groups capable of accepting protons.
2°	Molecular linear free energy relation (S)	MLFER_S [47, 51]	Overall polarizability/dipolarity descriptor derived from the cumulative contributions of solvatophilic groups, based on established empirical solvent interaction parameters.
1°	Sum of E-states for (strong) hydrogen bond donors	SHBdb [49, 51]	The sum of intrinsic electrotopological state values for all strong hydrogen bond donor atoms, accounting for their electronic and topological environment.

SHBdb

The association between SHBdb values and pIC50 levels, as shown in **Figure 5a**, illustrates the subtle balance essential for effective FLT3 tyrosine kinase inhibitor design. The SHBdb descriptor captures the strength and distribution of strong hydrogen bond donors, which are vital for forming stable contacts within the FLT3 active site. These donors enable critical hydrogen bonding with key residues, including Cys694 and Cys695 in the hinge region [13]. Optimal inhibitory potency is observed when SHBdb falls within the narrow window of 1–1.5. Values outside this interval result in reduced activity, since either too few or too many strong hydrogen bond donors can impair binding affinity or specificity. This pattern is consistent with observations reported by Kar *et al*. [10], who

stressed the importance of fine-tuning hydrogen bond donor properties to prevent loss of selectivity or excessive interaction. Thus, SHBdb serves as a key guide for optimizing hydrogen bond donor features, thereby improving both the potency and target selectivity of candidate inhibitors.

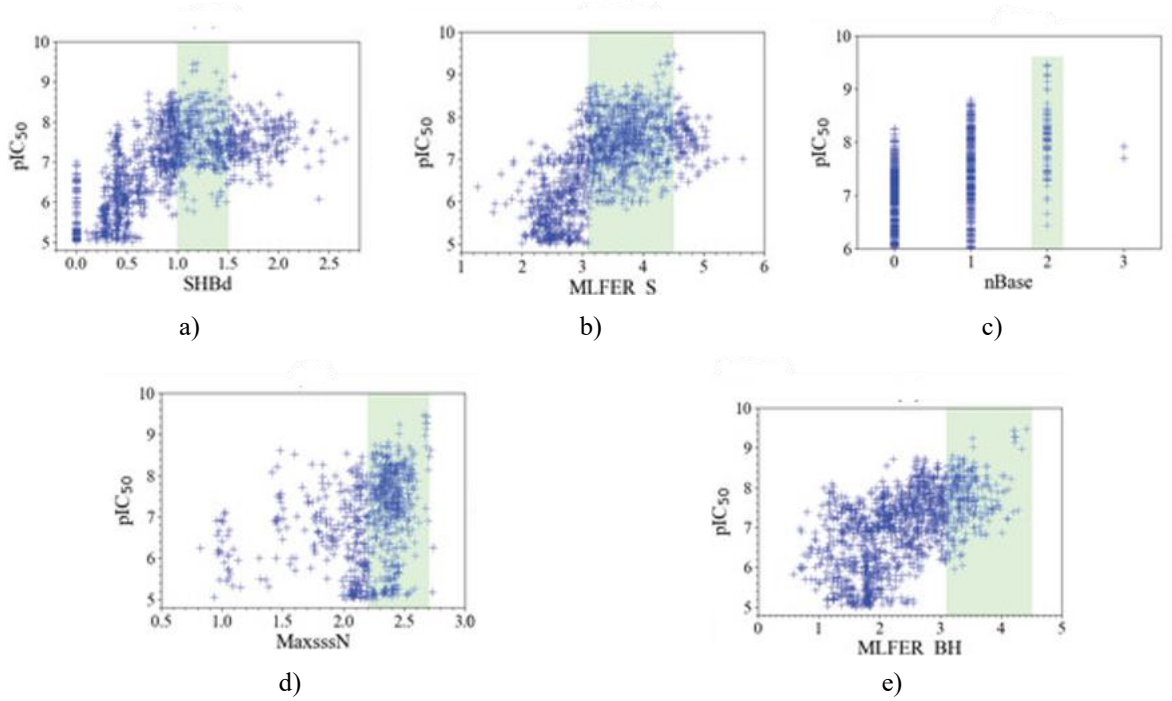


Figure 5. Scatter plots (a–e) depicting the correlations between the five most critical molecular descriptors (SHBdb, MLFER_S, nBase, MaxsssN, and MLFER_BH) and FLT3 inhibitor potency (expressed as pIC50) for compounds in both training and test datasets. The green highlighted areas in each panel mark the descriptor value ranges linked to elevated potency.

MLFER_S

Examination of **Figure 5b** indicates that the ideal range for MLFER_S in potent FLT3 tyrosine kinase inhibitors lies between 3.1 and 4.5. This descriptor measures overall polarizability and dipolarity, reflecting the molecule's capacity for hydrophobic and solvophobic interactions that are essential for fitting into the hydrophobic regions of the FLT3 binding pocket, involving residues such as Phe830 and Tyr693 [13]. Compounds falling within this range displayed maximum inhibitory potency, whereas those outside it showed reduced efficacy. These observations agree with the results reported by Shih *et al*. [11], who established that balanced hydrophobic contributions improve ligand binding affinity and target specificity. Consequently, MLFER_S emphasizes the need to carefully calibrate hydrophobicity and solubility to achieve optimal performance in FLT3 inhibitor design.

nBase, MaxsssN, and MLFER_BH

The patterns observed for nBase, MaxsssN, and MLFER_BH in **Figures 5c–5e**, respectively, together illustrate the complex interplay between structural features and FLT3 inhibitory potency.

Highest activity was achieved when compounds possessed exactly two basic groups **(Figure 5c).** Such basic functionalities promote electrostatic interactions and hydrogen bonding with key FLT3 residues, including Asp698 and Lys644 [12]. This finding corroborates the work of Kar *et al*. [10], which underscored the value of basic nitrogen atoms in strengthening ligand–receptor complexes.

The MaxsssN descriptor captures the electrotopological state of tertiary nitrogen atoms (those with three single bonds), commonly found in amine or amide moieties, which play a vital role in hydrogen bonding and electrostatic contacts. Enhanced potency was evident in compounds with MaxsssN values above 1.5, becoming particularly pronounced beyond 2.2, consistent with prior reports [11]. These nitrogen features substantially support effective binding and selectivity in the FLT3 active site.

MLFER_BH, in turn, provides a comprehensive measure of the molecule's total hydrogen bond acceptor strength. The strongest inhibitory effects were seen in compounds with MLFER_BH exceeding 3.1, where acceptor

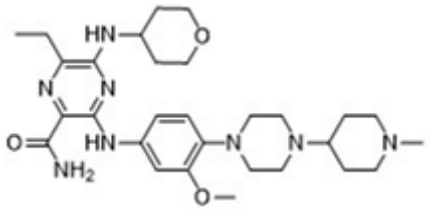
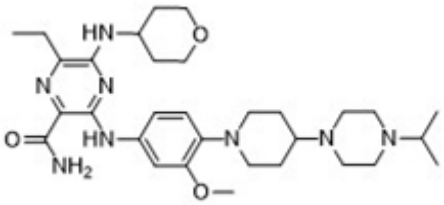
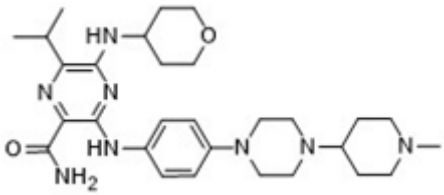
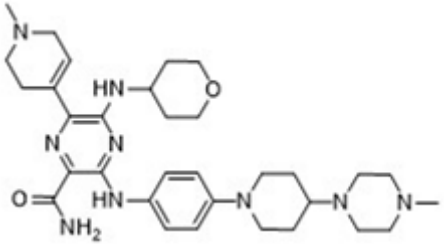
groups—such as carbonyl oxygens or heterocyclic nitrogens—establish robust hydrogen bonds with residues like Gly697 and Cys695 [15]. This broader descriptor extends beyond nitrogen-specific contributions to include all potential acceptor sites, affirming the central importance of hydrogen bond acceptance in the inhibition mechanism.

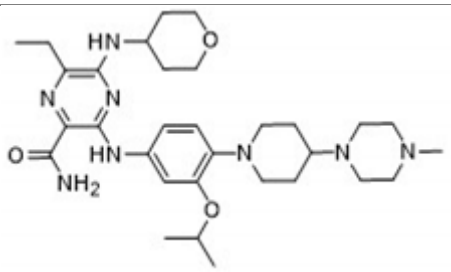
Together, nBase, MaxsssN, and MLFER_BH encapsulate key physicochemical properties—electrostatics, hydrogen bonding capacity, and acceptor distribution—that govern FLT3 inhibition. These insights, backed by multiple independent studies [10-12], validate the significance of these descriptors for guiding the rational optimization of next-generation therapeutic agents.

Discovery of novel FLT3 inhibitors via ligand-based virtual screening

By employing ligand-based virtual screening (LBVS) with the developed cheminformatics model, several highly promising candidates with strong predicted inhibitory activity against FLT3 tyrosine kinase were uncovered. The five most promising compounds are listed in **Table 5.** These selections feature structural similarities to gilteritinib, an advanced-generation FLT3 inhibitor [52]. This strategy demonstrates the power of LBVS in rapidly pinpointing molecules with favorable biological profiles solely from ligand information, bypassing the need for target structure or physical assays. The identified pyrazinecarboxamide derivatives exhibit predicted pIC50 values approaching that of gilteritinib (9.39) [53], illustrating the effectiveness of this computational pipeline in accelerating the identification of new FLT3-targeted agents for AML harboring FLT3 mutations. Overall, these results deepen our knowledge of FLT3 inhibitor structure–activity relationships and offer valuable candidates for subsequent experimental evaluation and validation.

Table 5. Top five candidates for FLT3 inhibitors identified by ligand-based virtual screening.

IUPAC Name	pIC_{50}	Structure
6-Ethyl-3-[3-methoxy-4-[4-(1-methylpiperidin-4-yl)piperazin-1-yl]anilino]-5-(oxan-4-ylamino)pyrazine-2-carboxamide	9.34	
6-Ethyl-3-[3-methoxy-4-[4-(4-propan-2-ylpiperazin-1-yl) piperidin-1-yl]anilino]-5-(oxan-4-ylamino)pyrazine-2-carboxamide	9.34	
3-[4-[4-(1-Methylpiperidin-4-yl)piperazin-1-yl]anilino]-5-(oxan-4-ylamino)-6-propan-2-ylpyrazine-2-carboxamide	9.29	
6-(1-Methyl-3,6-dihydro-2H-pyridin-4-yl)-3-[4-[4-(4-methylpiperazin-1-yl)piperidin-1-yl]anilino]-5-(oxan-4-ylamino)pyrazine-2-carboxamide	9.27	

6-Ethyl-3-[4-[4-(4-methylpiperazin-1-yl)piperidin-1-yl]-3-propan-2-yloxyanilino]-5-(oxan-4-ylamino)pyrazine-2-carboxamide	9.27	

The association between the molecules listed in **Table 5** and the descriptors SHBd, MLFER_S, nBase, MaxsssN, and MLFER_BH underscores the link between molecular substructures and inhibitory potency. In compounds such as 6-Ethyl-3-[3-methoxy-4-[4-(1-methylpiperidin-4-yl)piperazin-1-yl]anilino]-5-(oxan-4-ylamino)pyrazine-2-carboxamide, elevated SHBd levels arise from numerous hydrogen bond donor moieties, thereby boosting their inhibition effectiveness. For structures like 3-[4-[4-(1-Methylpiperidin-4-yl)piperazin-1-yl]anilino]-5-(oxan-4-ylamino)-6-propan-2-ylpyrazine-2-carboxamide, increased MLFER_S scores indicate the inclusion of polarophilic functional groups that enhance aqueous solubility and binding interactions with the target protein. The prevalence of basic functionalities (nBase), including amine groups and piperidine moieties, is common across these selected compounds. As an example, the compound 6-Ethyl-3-[4-[4-(4-methylpiperazin-1-yl)piperidin-1-yl]-3-propan-2-yloxyanilino]-5-(oxan-4-ylamino)pyrazine-2-carboxamide contains multiple nitrogen centers that increase its basic character. Compounds exemplified by 6-(1-Methyl-3,6-dihydro-2H-pyridin-4-yl)-3-[4-[4-(4-methylpiperazin-1-yl)piperidin-1-yl]anilino]-5-(oxan-4-ylamino)pyrazine-2-carboxamide show raised MaxsssN figures owing to tertiary nitrogens in piperazine units. Lastly, MLFER_BH represents the cumulative hydrogen bond basicity across all possible acceptor positions. Pyrazine-2-carboxamide-based derivatives feature abundant hydrogen bond acceptor locations, which amplify their total hydrogen bond basicity and improve affinity toward the FLT3 tyrosine kinase.

Script-like tool description

To improve accessibility of the predictive model, a freely available script-oriented tool was developed for automatic computation of pIC50 and IC50 values from any input compound's SMILES notation **(Figure 4b).** This resource is available through the link: https://github.com/Jacksonalcazar/Prediction-of-FLT3-Inhibitory-Activity (created on 6 July 2024). The tool is engineered for ease of use and speed, allowing straightforward processing of multiple SMILES entries in an automated fashion, with outputs generated in mere seconds.

Conclusion

This research effectively illustrated the utility and strength of a combined QSAR-machine learning framework for estimating pIC50 values of FLT3 tyrosine kinase inhibitors, drawing on detailed ligand structural properties. The achievement stemmed from utilizing a broad and varied compound collection, which encompassed essential elements influencing their biological effects. Through thorough dataset assembly, in-depth descriptor examination, and careful comparison of multiple machine learning techniques, the resulting model exhibited outstanding predictive power alongside relative straightforwardness.

In particular, the random forest regressor proved superior, substantiated by stringent external and internal validation protocols. It offers a straightforward yet dependable resource for spotting prospective FLT3 inhibitors, supported by Q^2_LOO of 0.926 and Q^2_10-fold of 0.922 over a diverse dataset. Furthermore, it achieved an R^2 of 0.941 and SD of 0.237 when forecasting pIC50 for 270 external FLT3 inhibitor compounds.

The feature refinement and selection efforts also underscored the vital role of certain molecular descriptors in determining inhibitor effectiveness, yielding important structural insights that can guide the targeted development of novel FLT3 inhibitors and refine drug discovery by emphasizing beneficial traits.

Additionally, the creation of an accessible script-based prediction tool marks a noteworthy advancement in cheminformatics resources, providing investigators with an effective and practical way to assess potential FLT3 inhibition of candidate molecules, inclusive of ligand-based screening applications. The tool's capacity for large-scale use was evidenced by its handling of up to 10.2 million compounds, highlighting its appropriateness for

extensive virtual screening. It also independently manages descriptor generation and predictions, with support for RDKit and Open Babel inputs, promoting efficient integration without additional software dependencies.
Overall, this work delivers a straightforward predictive framework for pIC50 estimation in FLT3 tyrosine kinase inhibitors, establishing an advanced standard in merging machine learning with QSAR for therapeutic development. The method provides improved accuracy and ease of use, aiding swift detection of promising AML treatments through FLT3 targeting. The tool's robustness, speed, and interoperability further establish it as an essential asset in cheminformatics and initial drug design phases.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Birg, F.; Courcoul, M.; Rosnet, O.; Bardin, F.; Pébusque, M.J.; Marchetto, S.; Tabilio, A.; Mannoni, P.; Birnbaum, D. Expression of the FMS/KIT-like gene FLT3 in human acute leukemias of the myeloid and lymphoid lineages. Blood 1992, 80, 2584–93.
2. Small, D. FLT3 mutations: Biology and treatment. Hematol. Am. Soc. Hematol. Educ. Program 2006, 2006, 178–84.
3. Barley, K.; Navada, S.C. Acute myeloid leukemia. Oncology 2019, 373, 308–18.
4. Kazi, J.U.; Rönnstrand, L. FMS-like tyrosine kinase 3/FLT3: From basic science to clinical implications. Physiol. Rev. 2019, 99, 1433–66.
5. Kantarjian, H.M.; Short, N.J.; Fathi, A.T.; Marcucci, G.; Ravandi, F.; Tallman, M.; Wang, E.S.; Wei, A.H. Acute Myeloid Leukemia: Historical Perspective and Progress in Research and Therapy Over 5 Decades. Clin. Lymphoma Myeloma Leuk. 2021, 21, 580–97.
6. Wei, A.H.; Tiong, I.S. Midostaurin, enasidenib, CPX-351, gemtuzumab ozogamicin, and venetoclax bring new hope to AML. Blood 2017, 130, 2469–74.
7. Daver, N.; Wei, A.H.; Pollyea, D.A.; Fathi, A.T.; Vyas, P.; DiNardo, C.D. New directions for emerging therapies in acute myeloid leukemia: The next chapter. Blood Cancer J. 2020, 10, 1–12.
8. Kantarjian, H.; Kadia, T.; DiNardo, C.; Daver, N.; Borthakur, G.; Jabbour, E.; Garcia-Manero, G.; Konopleva, M.; Ravandi, F. Acute myeloid leukemia: Current progress and future directions. Blood Cancer J. 2021, 11, 1–25.
9. Jaramillo, S.; Schlenk, R.F. Update on current treatments for adult acute myeloid leukemia: To treat acute myeloid leukemia intensively or non-intensively? That is the question. Haematologica 2023, 108, 342–52.
10. Kumar Kar, R.; Suryadevara, P.; Roushan, R.; Chandra Sahoo, G.; Ranjan Dikhit, M.; Das, P. Quantifying the Structural Requirements for Designing Newer FLT3 Inhibitors. Med. Chem. 2012, 8, 913–27.
11. Shih, K.C.; Lin, C.Y.; Chi, H.C.; Hwang, C.S.; Chen, T.S.; Tang, C.Y.; Hsiao, N.W. Design of novel FLT-3 inhibitors based on dual-layer 3D-QSAR model and fragment-based compounds in silico. J. Chem. Inf. Model. 2012, 52, 146–55.
12. Abutayeh, R.F.; Taha, M.O. Discovery of novel Flt3 inhibitory chemotypes through extensive ligand-based and new structure-based pharmacophore modelling methods. J. Mol. Graph. Model. 2019, 88, 128–51.
13. Bhujbal, S.P.; Keretsu, S.; Cho, S.J. Design of New Therapeutic Agents Targeting FLT3 Receptor Tyrosine Kinase Using Molecular Docking and 3D-QSAR Approach. Lett. Drug Des. Discov. 2019, 17, 585–96.
14. Fernandes, Í.A.; Resende, D.B.; Ramalho, T.C.; Kuca, K.; Da Cunha, E.F.F. Theoretical studies aimed at finding FLT3 inhibitors and a promising compound and molecular pattern with dual aurora B/FLT3 activity. Molecules 2020, 25, 1726.
15. Ghosh, S.; Keretsu, S.; Cho, S.J. Molecular modeling studies of n-phenylpyrimidine-4-amine derivatives for inhibiting FMS-like tyrosine kinase-3. Int. J. Mol. Sci. 2021, 22, 12511.

16. Sandoval, C.; Torrens, F.; Godoy, K.; Reyes, C.; Farías, J. Application of Quantitative Structure-Activity Relationships in the Prediction of New Compounds with Anti-Leukemic Activity. Int. J. Mol. Sci. 2023, 24, 12258.
17. Islam, M.R.; Osman, O.I.; Hassan, W.M. Identifying novel therapeutic inhibitors to target FMS-like tyrosine kinase-3 (FLT3) against acute myeloid leukemia: A molecular docking, molecular dynamics, and DFT study. J. Biomol. Struct. Dyn. 2023.
18. Liu, K.; Hu, J. Classification of acute myeloid leukemia M1 and M2 subtypes using machine learning. Comput. Biol. Med. 2022, 147, 105741.
19. Abhishek, A.; Jha, R.K.; Sinha, R.; Jha, K. Automated classification of acute leukemia on a heterogeneous dataset using machine learning and deep learning techniques. Biomed. Signal Process. Control 2022, 72, 103341.
20. Monaghan, S.A.; Li, J.L.; Liu, Y.C.; Ko, M.Y.; Boyiadzis, M.; Chang, T.Y.; Wang, Y.F.; Lee, C.C.; Swerdlow, S.H.; Ko, B.S. A machine learning approach to the classification of acute leukemias and distinction from nonneoplastic cytopenias using flow cytometry data. Am. J. Clin. Pathol. 2022, 157, 546–53.
21. Awada, H.; Durmaz, A.; Gurnari, C.; Kishtagari, A.; Meggendorfer, M.; Kerr, C.M.; Kuzmanovic, T.; Durrani, J.; Shreve, J.; Nagata, Y.; et al. Machine learning integrates genomic signatures for subclassification beyond primary and secondary acute myeloid leukemia. Blood J. Am. Soc. Hematol. 2021, 138, 1885–95.
22. Dese, K.; Raj, H.; Ayana, G.; Yemane, T.; Adissu, W.; Krishnamoorthy, J.; Kwa, T. Accurate machine-learning-based classification of leukemia from blood smear images. Clin. Lymphoma Myeloma Leuk. 2021, 21, e903–e914.
23. Talaat, F.M.; Gamel, S.A. Machine learning in detection and classification of leukemia using C-NMC_Leukemia. Multimed. Tools Appl. 2024, 83, 8063–8076.
24. Nasimian, A.; Al Ashiri, L.; Ahmed, M.; Duan, H.; Zhang, X.; Rönnstrand, L.; Kazi, J.U. A Receptor Tyrosine Kinase Inhibitor Sensitivity Prediction Model Identifies AXL Dependency in Leukemia. Int. J. Mol. Sci. 2023, 24, 3830.
25. Janssen, A.P.; Grimm, S.H.; Wijdeven, R.H.; Lenselink, E.B.; Neefjes, J.; Van Boeckel, C.A.; Van Westen, G.J.; Van Der Stelt, M. Drug Discovery Maps, a Machine Learning Model That Visualizes and Predicts Kinome-Inhibitor Interaction Landscapes. J. Chem. Inf. Model. 2019, 59, 1221–29.
26. Zhao, Y.; Tian, Y.; Pang, X.; Li, G.; Shi, S.; Yan, A. Classification of FLT3 inhibitors and SAR analysis by machine learning methods. Mol. Divers. 2023, 1, 1–17.
27. Eckardt, J.N.; Bornhäuser, M.; Wendt, K.; Middeke, J.M. Application of machine learning in the management of acute myeloid leukemia: Current practice and future prospects. Blood Adv. 2020, 4, 6077–85.
28. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2023 update. Nucleic Acids Res. 2023, 51, D1373–D1380.
29. PubChem Database. Available online: https://pubchem.ncbi.nlm.nih.gov (accessed on 28 November 2023).
30. Kenneth Reitz. Requests: HTTP for Humans™ — Requests 2.26.0 Documentation. 2021. Available online: https://docs.python-requests.org/en/latest/ (accessed on 7 February 2024).
31. McKinney, W.; Team, P.D. Pandas—Powerful Python Data Analysis Toolkit. 2015. Available online: https://pandas.pydata.org (accessed on 7 February 2024).
32. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–30.
33. Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 2011, 32, 1466–74.
34. Landrum, G. RDKit: Open-Source Cheminformatics 2022_9_5 (Q3 2022). Available online: https://zenodo.org/records/7671152 (accessed on 23 February 2023).
35. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32.
36. Friedman, J.H. Greedy function approximation: A gradient boosting machine. Ann. Stat. 2001, 29, 1189–32.
37. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. IEEE Intell. Syst. Their Appl. 1998, 13, 18–28.

38. Williams, C.K.; Rasmussen, C.E. Gaussian Processes for Machine Learning; MIT Press: Cambridge, MA, USA, 2006; Volume 2, p. 3.
39. Altman, N.; Krzywinski, M. Ensemble methods: Bagging and random forests. Nat. Pubchemds 2017, 14, 933–5.
40. Chollet, F. Keras, 2015. In: Github Repos. Available online: https://github.com/fchollet/keras (accessed on 15 September 2023).
41. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 2019, 32, 8026–37.
42. PubChem Substructure Fingerprint. Available online: https://pubchem.ncbi.nlm.nih.gov/docs/data-specification (accessed on 10 December 2023).
43. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J. Cheminform. 2015, 7, 1–13.
44. Durant, J.L.; Leland, B.A.; Henry, D.R.; Nourse, J.G. Reoptimization of MDL keys for use in drug discovery. J. Chem. Inf. Comput. Sci. 2002, 42, 1273–80.
45. Butina, D. Unsupervised data base clustering based on daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. J. Chem. Inf. Comput. Sci. 1999, 39, 747–50.
46. Marino, S.; Zhao, Y.; Zhou, N.; Zhou, Y.; Toga, A.W.; Zhao, L.; Jian, Y.; Yang, Y.; Chen, Y.; Wu, Q.; et al. Compressive Big Data Analytics: An ensemble meta-algorithm for high-dimensional multisource datasets. PLoS ONE 2020, 15, e0228520.
47. Platts, J.A.; Butina, D.; Abraham, M.H.; Hersey, A. Estimation of molecular linear free energy relation descriptors using a group contribution approach. J. Chem. Inf. Comput. Sci. 1999, 39, 835–45.
48. Ibrahim, Z.Y.; Uzairu, A.; Shallangwa, G.; Abechi, S. QSAR and molecular docking based design of some indolyl-3-ethanone-α-thioethers derivatives as Plasmodium falciparum dihydroorotate dehydrogenase (PfDHODH) inhibitors. SN Appl. Sci. 2020, 2, 1–12.
49. Hall, L.H.; Kier, L.B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. J. Chem. Inf. Comput. Sci. 1995, 35, 1039–1045.
50. Lin, C.; Xiaoxiao, Z. Optimizing Drug Screening with Machine Learning. In Proceedings of the 2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2022, Chengdu, China, 16–18 December 2022.
51. Euldji, I.; Si-Moussa, C.; Hamadache, M.; Benkortbi, O. QSPR Modelling of the Solubility of Drug and Drug-like Compounds in Supercritical Carbon Dioxide. Mol. Inform. 2022, 41, 2200026.
52. Lee, L.Y.; Hernandez, D.; Rajkhowa, T.; Smith, S.C.; Raman, J.R.; Nguyen, B.; Small, D.; Levis, M. Preclinical studies of gilteritinib, a next-generation FLT3 inhibitor. Blood 2017, 129, 257–60.
53. Shimada, I.; Kurosawa, K.; Matsuya, T.; Iikubo, K.; Kondoh, Y.; Kamikawa, A.; Tomiyama, H.; Iwai, Y. Patent US8969336. 2015. Available online: https://patents.google.com/patent/US8969336B2 (accessed on 25 April 2024).