StackBRAF: A Large-Scale Stacking Ensemble Learning for BRAF Affinity Prediction

Nur Fadhilah Syahid, Natthida Weerapreeyakul, and Tarapong Srisongkram*

Cite This: *ACS Omega* 2023, 8, 20881−20891

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The B-rapidly accelerated fibrosarcoma (BRAF) is a proto-oncogene that plays a vital role in cell signaling and growth regulation. Identifying a potent BRAF inhibitor can enhance therapeutic success in high-stage cancers, particularly metastatic melanoma. In this study, we proposed a stacking ensemble learning framework for the accurate prediction of BRAF inhibitors. We obtained 3857 curated molecules with BRAF inhibitory activity expressed as a predicted half-maximal inhibitory concentration value (pIC_{50}) from the ChEMBL database. Twelve molecular fingerprints from PaDeL-Descriptor were calculated for model training. Three machine learning algorithms including extreme gradient boosting, support vector regression, and multilayer perceptron were utilized for constructing new predictive features

(PFs). The meta-ensemble random forest regression, called StackBRAF, was created based on the 36 PFs. The StackBRAF model achieves lower mean absolute error (MAE) and higher coefficient of determination (R^2 and Q^2) than the individual baseline models. The stacking ensemble learning model provides good y-randomization results, indicating a strong correlation between molecular features and pIC_{50}. An applicability domain of the model with an acceptable Tanimoto similarity score was also defined. Moreover, a large-scale high-throughput screening of 2123 FDA-approved drugs against the BRAF protein was successfully demonstrated using the StackBRAF algorithm. Thus, the StackBRAF model proved beneficial as a drug design algorithm for BRAF inhibitor drug discovery and drug development.
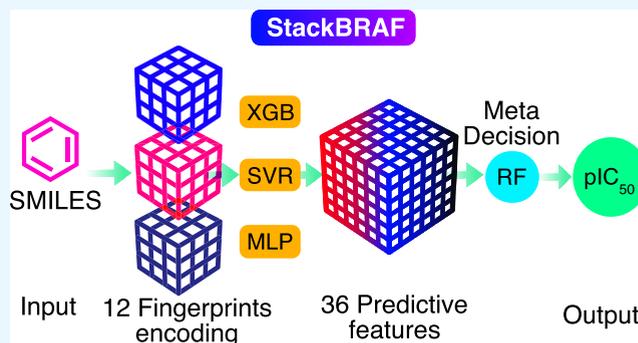
1. INTRODUCTION

The B-rapidly accelerated fibrosarcoma (BRAF) is a proto-oncogene belonging to the rat sarcoma (RAS)/rapidly accelerated fibrosarcoma (RAF) family.[1] The RAF family plays a crucial role in the mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) signaling pathway.[1] The RAF signaling module transduces signals from the cell surface to the nucleus. This module is composed of the RAS-GTPase enzyme, which is the upstream regulator of the RAF proteins.[2] The binding of growth factors, cytokines, and hormones to the membrane receptors activates RAS signaling and subsequent RAF activity. This signaling pathway leads to the activation of the dual-specificity protein kinases including MEK1 and MEK2, which in turn activate the ERK1 and ERK2 kinases.[3] Mutations in this pathway promote cell growth, proliferation, and survival in most cancers.

BRAF mutations are present in 49% of melanoma patients and increase the risk of mortality in melanoma patients by 1.7 times higher than in patients without BRAF mutations.[4] The majority (90%) of BRAF mutations occur at valine residue at position 600 that changes to glutamic acid (V600E).[5] To date, three generations of selective BRAF inhibitors such as sorafenib, dabrafenib, and TAK-632 have been approved for advanced-stage melanoma targeting BRAF V600E protein.[6]

These drugs have significantly increased patient responses; however, their acquired drug resistance limits their effectiveness as a single therapy, and most patients relapse within a year.[7] Identification of new BRAF inhibitors has thus been pursued to seek better-adapted BRAF inhibitors and more effective candidates.[6] *In silico* techniques employing powerful computing resources and computational frameworks contribute significantly to BRAF inhibitor research.[8,9]

A previous study constructed a 3D-quantitative structure–activity relationship (3D-QSAR) model for evaluating the BRAF inhibitory activity of 71 pyridopyrazinones using a partial least square (PLS) regression mode.[8] Recently, Gaussian field-based 3D-QSAR and molecular simulation studies were performed to design a potent pyrimidine–sulfonamide, also using the PLS model.[9] Two studies also built 3D-QSAR models using the PLS algorithm to predict BRAF

inhibitors.[10,11] Additionally, one study applied a deep learning (DL) generative model for designing BRAF inhibitors.[12] In total, four previous studies used 3D-QSAR and PLS models, and one study applied DL for predicting BRAF inhibitors. However, none of them implemented more sophisticated machine learning (ML) techniques.

ML and DL have accelerated the search for potent compounds with desired properties.[13,14] ML has significantly impacted kinase drug design as evidenced by the discovery of dual fibroblast growth factor receptor and epidermal growth factor receptor inhibitors.[15] Commonly used ML algorithms, such as random forest (RF), extreme gradient boosting (XGB), or multilayer perceptron (MLP), are widely employed in predicting drug bioactivity and pharmaceutical compositions.[16−18] However, one limitation of conventional ML models is a high generalization error, as predictions rely on a single model output. In contrast, stacked generalization refers to any scheme that feeds information from one set of estimators to another before the final estimator, which reduces generalization error.[19] This approach addresses the potential bias prediction from single ML models that rely on a single output. Stacking ensemble learning, on the other hand, produces the meta-decision from all ML output, which can improve the model accuracy.[19]

In this study, we developed StackBRAF—the stacking ensemble learning model to accurately predict the negative log of the inhibitory concentration at 50 percent (pIC_{50}) against BRAF protein. The experimental inhibitory concentration at 50 percent (IC_{50}) of all compounds against BRAF protein was obtained from the ChEMBL database. We used 12 molecular fingerprints from PaDeL-Descriptor software to train the model. The 10-fold cross-validation (10-CV) results of baseline models served as new predictive features (PFs) for constructing the meta-predictor using an ensemble RF regressor algorithm. The proposed ensemble learning model was evaluated for predicting ligand affinity against BRAF protein. The model's applicability was determined based on the principle that the model can only accurately predict the compounds with structures similar to those in the training dataset.[20] The goodness of fit, model robustness, and predictability were also measured. In conclusion, the StackBRAF model was applied to predict drug affinity against BRAF protein for 2123 US FDA-approved drugs as a part of a performance evaluation.

2. MATERIALS AND METHODS

2.1. Dataset. Chemical compounds and IC_{50} values against BRAF protein were collected from the ChEMBL database with assay ID: ChEMBL5145.[21] The chemical features were stored in a canonical isomeric simplified molecular input-line entry system (SMILES) format. Any missing SMILES or IC_{50} values were removed. Inorganics and mixed compounds were excluded from the dataset, as they could interfere with the molecular fingerprint calculation. Only compounds with numeric IC_{50} values were collected. The IC_{50} values were converted into molar units (M) and then transformed into pIC_{50} values by negative log transformation of the IC_{50} value. Duplicated molecules were eliminated using two criteria: if the duplicates had different pIC_{50} values less than or equal to 0.2, the pIC_{50} values were averaged, and only one entry with the average pIC_{50} value was kept. However, if duplicates had different pIC_{50} values greater than 0.2, both entries were removed from the dataset to ensure the model has reliable

prediction and reduced overfitting.[22] FDA-approved drugs were removed at the stage for use in the validating model performance in the FDA-approved drugs dataset. Molecules with a molecular weight greater than 700 (Da) and a LogP greater than 8 were also excluded because they are less likely to pass through the cell membrane.[23] The dataset for model construction and validation contained a total of 3857 compounds, consisting of 2697 compounds in the training dataset and 1157 compounds in the test dataset. The FDA-approved drugs' SMILES dataset was collected from Phase V drugs in the ChEMBL database. The FDA-approved drugs' SMILES were also transformed into canonical isomeric SMILES then the inorganic compounds, mixed compounds, and duplicate entries were eliminated. Finally, the 2123 curated FDA-approved drug datasets were used in this study. All datasets were included in the Supporting Information.

2.2. Molecular Fingerprints. The canonical isomeric SMILES format was used to generate molecular fingerprints using PaDeL-Descriptor software.[24] The SMILES strings have advantages in terms of their storage and easy handling compared to the 3D molecular structure or graph representation.[25] The canonical isomeric SMILE is a stereochemistry unique SMILES notation that can be used to identify a unique compound with its stereochemistry.[26] Molecular fingerprints represent the chemical substructures that are inherently present or absent in a molecule. To calculate molecular fingerprints, salts were removed, and the tautomer and nitro groups were standardized during fingerprint encoding. Herein, the 12 molecular fingerprints were used: AtomPairs2D (780 values),[27] AtomPairs2DCount (780 values),[27] CDK (1024 values),[28] CDKextended (1024 values),[28] CKDgraphonly (1024 values),[28] Estate (79 values),[29] KlekotaRoth (4860 values),[30] KlekotaRothCount (4860 values),[30] MACCS (166 values),[31] PubChem (881 values),[32] Substructure (307 values),[24] and SubstructureCount (307 values)[24] were used in this study.

2.3. Model Architecture. Stacking ensemble learning was constructed using 10-CV based on 36 PFs derived from 3 ML algorithms including XGB, support vector regression (SVR), and MLP. XGB is a tree-boosting algorithm that enhances its performance through gradient-boosting ensemble and regularization methods. The final output of the XGB algorithm consists of consecutively trained decision tree models. For all 12 molecular fingerprints, the XGB hyperparameters are set as follows: gamma = 0, reg_lambda = 1, reg_alpha = 0, max_depth = 6, n_estimators = 100, and learning_rate = 0.3. SVR is a nonparametric regression model that predicts outputs based on a linear hyperplane and an acceptable error margin. The regularization parameter (C) of SVR was randomly searched from 1 to 10 for each of the 12 molecular fingerprints using grid search fivefold CV. MLP is a feed-forward artificial neural network (ANN) that serves as the foundation for deep neural networks. The MLP is composed of three primary components: the input layer, the hidden layer, and the output layer, which are fully connected to each other. The MLP hyperparameters, including hidden_layer_sizes = (100, 100, 100), activation = ReLU, solver = Adam, alpha = 0.001, learning_rate_init = [0.001, 0.01], max_iter = 500, were randomly searched for each of the 12 molecular fingerprints by using grid search fivefold CV. The best estimators were chosen and constructed for all 36 baseline models. Subsequently, the 10-fold CV pIC_{50} values from each baseline model were used as a new PF. These new PFs can be defined by eq 1:

$$PF = \{PF1, PF2, PF3, .., PF36\} \qquad (1)$$

where PF represents the 10-CV of the baseline model for each fingerprint from each molecule. The RF algorithm was employed as an ensemble method to predict BRAF affinity. The hyperparameters search in grid search fivefold CV for the RF model included max_depth = [10, 20, 50], n_estimator = [10, 100], and max_features = [2, 3, 4, 5]. The summary of the StackBRAF algorithm was presented in Scheme 1.

Scheme 1. StackBRAF's Model Architecture
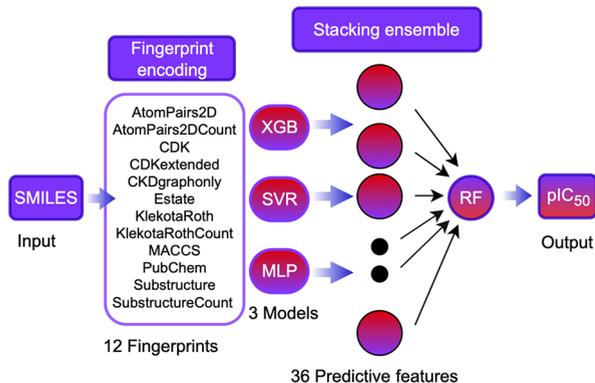
2.4. Model Evaluation. The coefficient of determination (R^2 or Q^2) and the mean absolute error (MAE) are the two essential performance metrics for measuring baseline and StackBRAF models. High prediction performance is associated with lower MAE and higher R^2 or Q^2 values. The MAE and R^2

for the training dataset (MAE_{Tr} and R^2_{Tr}), CV dataset (MAE_{CV} and Q^2_{CV}), and external test dataset (MAE_{Ext} and Q^2_{Ext}) were calculated based on the following eqs 2 and 3:

$$MAE = \frac{1}{n} \sum_1^n \left(|y_i - \hat{y}_i| \right) \qquad (2)$$

$$R^2 \text{ or } Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y)^2} \qquad (3)$$

where y_i is an experimental pIC_{50} value and the $\hat{y}_i$ is the corresponding predictive y_i value. The $\bar{y}$ represents the average of experimental y_i values. The n value is the total number of molecules in the dataset. The MAE value indicates the prediction errors of the model. Lower values of MAE (close to 0) suggest low prediction errors. The coefficient of determination (R^2 or Q^2) determines the percentage of variance from the experimental pIC_{50} that can be explained by the model. High values of R^2 and Q^2 (close to 1) indicate the high goodness of fit and high predictability, respectively. Generally, R^2 or $Q^2 > 0.6$ suggests that the prediction model is performing well, $R^2 - Q^2 \leq 0.2$ represents the goodness of fit, and the MAE less than 1 was evident that the model has a good prediction accuracy.[33]

2.5. *y*-Randomization. The correlation of the QSAR models was determined using a *y*-randomization experiment. The *x−y* pairings in the training set were randomly scrambled to produce the incorrect *x−y* pairs. These false *x−y* pairs were utilized to train the model, which was then used to predict the pIC_{50} values and calculate the R^2 and Q^2 values. The
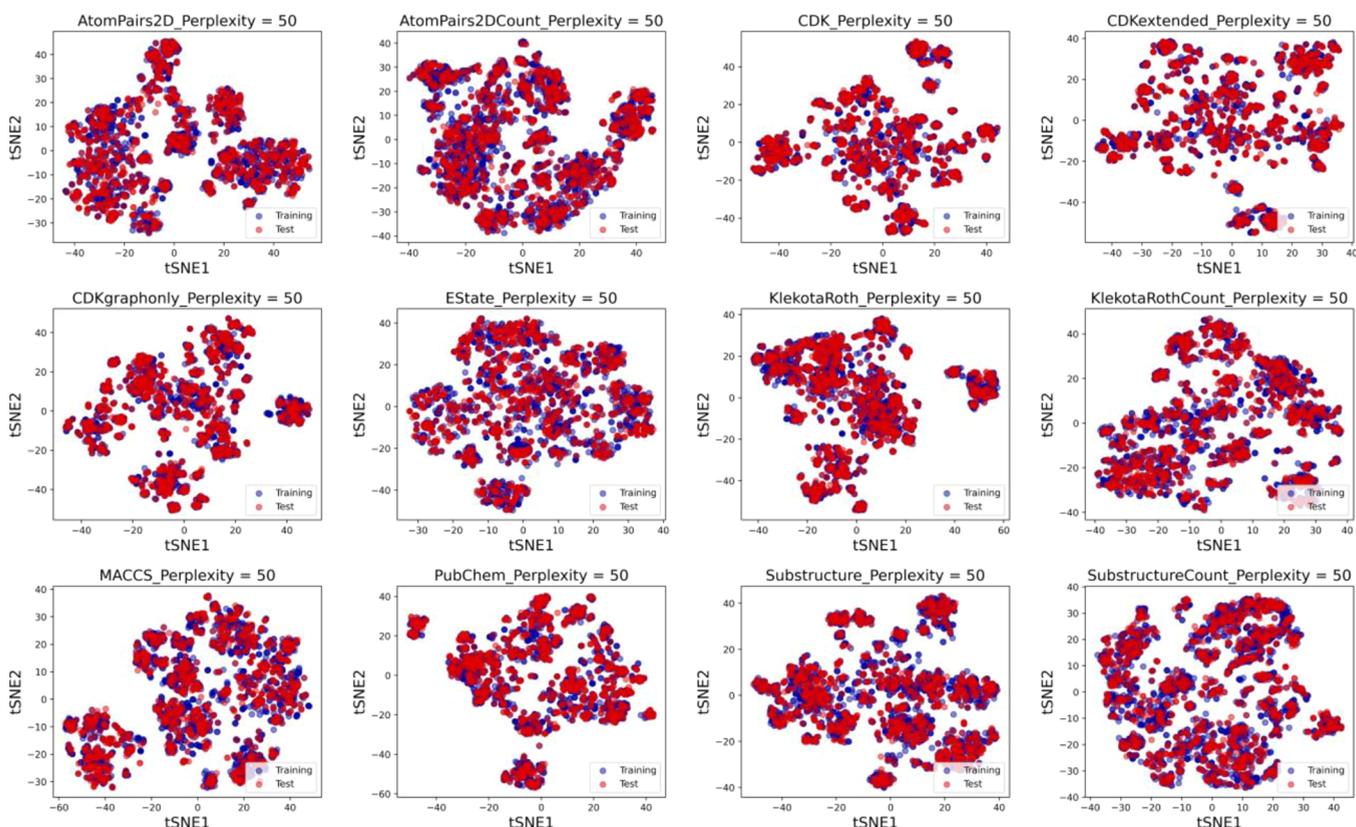
Figure 1. Chemical distribution of the BRAF dataset with 12 fingerprints displayed by the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm. Blue and red colors indicate the training and test datasets, respectively.

Table 1. Baseline Model Performances of 36 Machine Learning Models

	XGBoost		MLP		SVR	
	R^2_{Tr}[a]	MAE_{Tr}[b]	R^2_{Tr}[a]	MAE_{Tr}[b]	R^2_{Tr}[a]	MAE_{Tr}[b]
AtomPairs2D	0.89	0.31	0.84	0.40	0.75	0.42
AtomPair2DCount	0.97	0.16	0.91	0.31	0.63	0.59
CDK	0.99	0.08	0.97	0.20	0.98	0.13
CDKextended	0.99	0.08	0.99	0.11	0.98	0.13
CDKgraphonly	0.96	0.17	0.96	0.18	0.93	0.21
EState	0.77	0.45	0.75	0.51	0.67	0.52
KlekotaRoth	0.93	0.28	0.99	0.10	0.97	0.14
KlekotaRothCount	0.93	0.27	0.99	0.11	0.88	0.29
MACCS	0.94	0.23	0.95	0.21	0.87	0.28
PubChem	0.95	0.22	0.93	0.25	0.88	0.29
Substructure	0.82	0.41	0.84	0.39	0.75	0.44
SubstructureCount	0.92	0.27	0.89	0.32	0.51	0.71

[a]R^2_{Tr}: determination coefficient of training dataset. [b]MAE_{Tr}: mean absolute error of training dataset.

Table 2. 10-Fold Cross-Validation Performances of 36 Machine Learning Models

	XGBoost		MLP		SVR	
	Q^2_{CV}[a]	MAE_{CV}[b]	Q^2_{CV}[a]	MAE_{CV}[b]	Q^2_{CV}[a]	MAE_{CV}[b]
AtomPairs2D	0.62	0.61	0.51	0.70	0.57	0.64
AtomPair2DCount	0.68	0.55	0.60	0.65	0.57	0.66
CDK	0.72	0.51	0.69	0.55	0.77	0.47
CDKextended	0.71	0.52	0.67	0.57	0.77	0.47
CDKgraphonly	0.69	0.54	0.63	0.60	0.71	0.53
EState	0.55	0.67	0.47	0.72	0.53	0.68
KlekotaRoth	0.73	0.52	0.64	0.60	0.76	0.49
KlekotaRothCount	0.73	0.52	0.64	0.59	0.72	0.53
MACCS	0.65	0.58	0.54	0.68	0.66	0.57
PubChem	0.72	0.52	0.66	0.58	0.73	0.51
Substructure	0.61	0.63	0.52	0.69	0.60	0.62
SubstructureCount	0.67	0.57	0.60	0.65	0.48	0.75

[a]Q^2_{CV}: determination coefficient of the 10-fold cross-validation model. [b]MAE_{CV}: mean absolute error of the 10-fold cross-validation model.

differences in the R^2 and Q^2 values between the original $x–y$ pairs and the false $x–y$ pairs indicate the model's robustness. The model is considered robust if the original $x–y$ pairings outperformed their y-randomized models. If the y-randomized models produced similar performance compared to the original $x–y$ pairings, the $x–y$ correlation of the model was defined as not trustworthy. In this study, the robustness of the model was validated by 100 y-randomization models.

2.6. Feature Importance. The SHapley Additive exPlanations (SHAP) feature importance value was used to identify the predictive fingerprints that affect the predicted pIC_{50} output of the StackBRAF model. SHAP algorithm works through game theory, where each feature is an agent that affects the model's output. The SHAP algorithm determines both the impact of each feature and the direction (positive and negative) of the model's output. The SHAP values explain the contribution of a given molecule in the dataset (local explainer). All SHAP values were then averaged to calculate the mean SHAP value, which can explain the global impact of each feature on the model's output (global explainer).[34] Positive SHAP contributions have an impact on the positive pIC_{50} value, whereas negative SHAP contributions have an impact on the negative pIC_{50} value.

2.7. Applicability Domain. The applicability region is specified as the field of application of the QSAR model, which is a set of chemicals for which the model should produce accurate predictions.[20] This principle is important because the

model can only offer reliable predictions for compounds that are identical to those utilized in the model development. The applicability domain was assessed using two methods: (1) a plot between normalized residuals of model prediction and the experimental pIC_{50} value, and (2) a Tanimoto similarity score to the non-outlier training and test datasets. The standardized residuals are calculated by using eq 4:

$$\text{Standardized residual} = \frac{y_i - \widehat{y_i}}{\text{Standard deviation } (y_i - \widehat{y_i})} \quad (4)$$

where y_i is an experimental pIC_{50} value and the $\hat{y}_i$ is the corresponding predictive y_i value. An absolute standard residual higher than 3 in both the training and test dataset was considered an outlier of the model. Once the outliers were identified, the Tanimoto similarity score was calculated using the extended circular fingerprint at radius 3 with 2048 bits from the RDKit software. This score was used to assess the similarity of a new predictive compound to the non-outliers training and test datasets and was calculated based on eq 5 for compound A and compound B ($T_{(A,B)}$):

$$T_{(A,B)} = \frac{(A \cap B)}{A + B - (A \cap B)} \quad (5)$$

where $(A \cap B)$ is the number of common presence features in both fingerprints A and B, while A and B are the number of presence features in fingerprints A and B, respectively. If the

Table 3. External Test Performances of 36 Machine Learning Models

	XGBoost		MLP		SVR	
	Q^2_{Ext}[a]	MAE_{Ext}[b]	Q^2_{Ext}[a]	MAE_{Ext}[b]	Q^2_{Ext}[a]	MAE_{Ext}[b]
AtomPairs2D	0.68	0.55	0.53	0.67	0.65	0.58
AtomPair2DCount	0.75	0.49	0.67	0.59	0.60	0.64
CDK	0.77	0.46	0.74	0.51	0.81	0.43
CDKextended	0.77	0.47	0.72	0.53	0.81	0.44
CDKgraphonly	0.72	0.51	0.68	0.56	0.77	0.49
EState	0.58	0.65	0.44	0.77	0.58	0.64
KlekotaRoth	0.76	0.49	0.69	0.57	0.80	0.45
KlekotaRothCount	0.77	0.49	0.70	0.56	0.75	0.50
MACCS	0.72	0.52	0.59	0.65	0.73	0.52
PubChem	0.77	0.47	0.71	0.53	0.76	0.48
Substructure	0.68	0.58	0.57	0.66	0.65	0.59
SubstructureCount	0.72	0.53	0.65	0.59	0.51	0.73

[a]Q^2_{Ext}: determination coefficient of the external model. [b]MAE_{Ext}: mean absolute error of the external model.
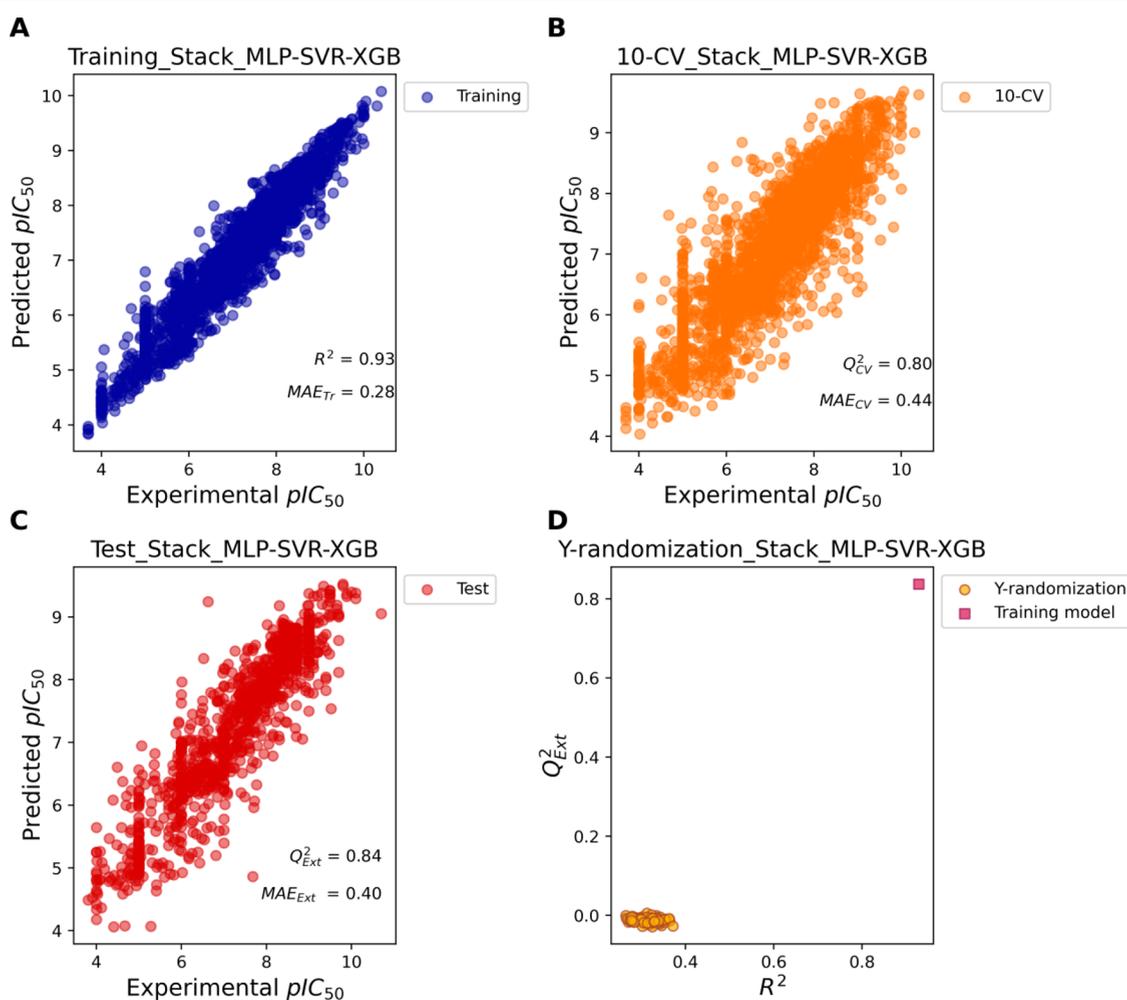
Figure 2. StackBRAF model performance. (A) training dataset performance, (B) 10-fold cross-validation (CV) performance, (C) external test dataset performance, and (D) y-randomization results between the original model (red) and the y-randomization models (yellow color, $n = 100$).

maximum Tanimoto similarity score of a given compound is higher than moderate similarity (0.5), when compared to the non-outliers of training and test datasets, the predictive compound was defined as in the applicability domain.

3. RESULTS

3.1. Chemical Distribution of the BRAF Dataset.
Figure 1 illustrates the chemical distribution of BRAF training and test datasets by the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm. The t-SNE visualization is based on the distribution of the compounds from 12 fingerprints. The blue and red circles represent the training

Table 4. BRAF Inhibitor QSAR Performance Comparison

models	training dataset	test dataset	R^2	Q^2_{CV}	$R^2_{y\text{-random}}$	Q^2_{Ext}	references
3D-QSAR	71	19	0.907	0.66	not reported	0.976	8
Gaussian-3D-QSAR	36	15	0.96	0.78	0.68	0.63	9
3D-QSAR	27	189	not reported	0.70	not reported	0.56	10
3D-QSAR	243	60	0.71	0.68	0.002	0.82	11
StackBRAF	2697	1157	0.93	0.80	0.32	0.84	our study
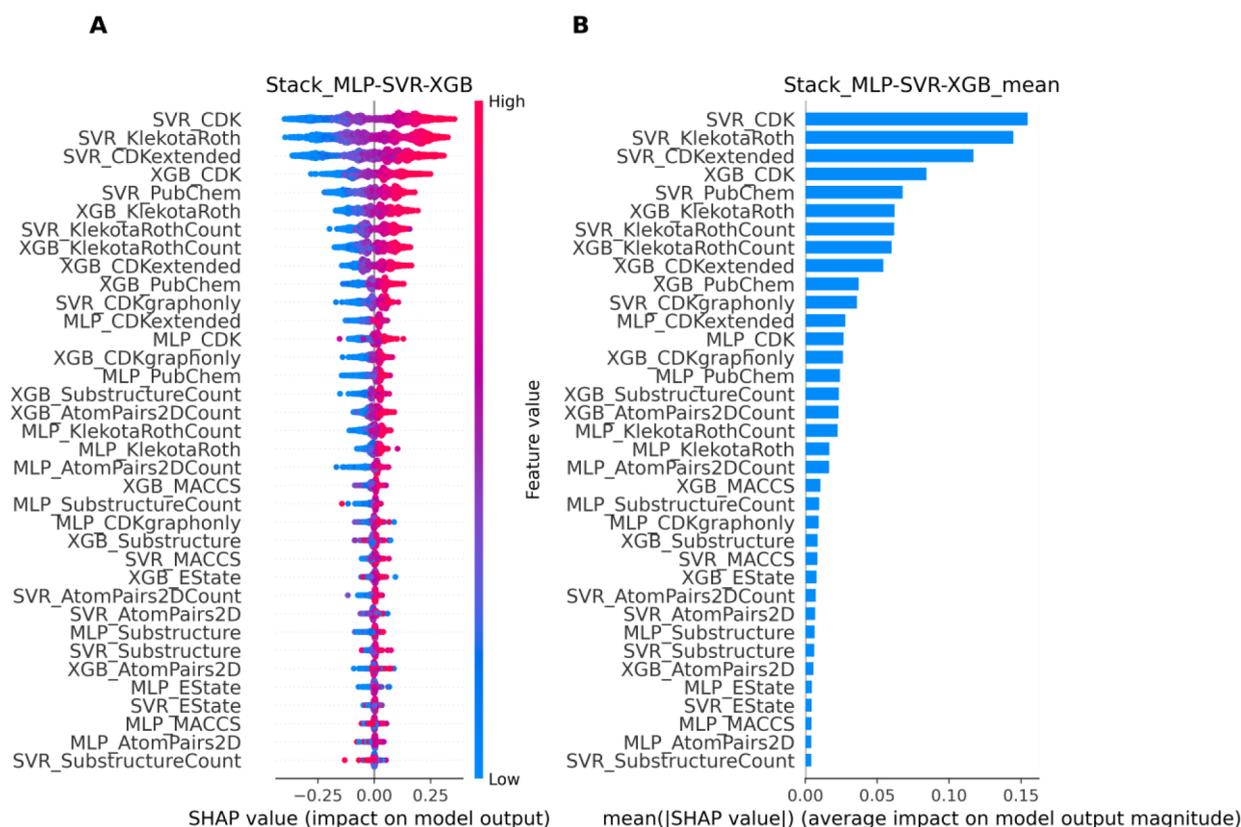
Figure 3. Predictive feature importance of the StackBRAF model. (A) Summary plot and (B) average SHAP value per predictive feature. The red and blue colors indicate the high and low predictive feature values in the summary plot, respectively.

and test datasets, respectively. The results reveal that the chemical distribution of all 12 fingerprints exhibits the similarity distribution between training and test datasets. No distinct separation between these datasets was observed in the distribution plots (Figure 1). This result indicates that the 12 molecular fingerprints of the test dataset share a similar chemical distribution with those of the training dataset. Therefore, the test dataset can be used to determine the model's predictability, and all 12 molecular fingerprints can be utilized for model construction.

3.2. Machine Learning Model Performance. Tables 1−3 present the machine-learning performance of 36 models, resulting from the combination of 12 molecular fingerprints with 3 algorithms. Table 1 shows that the XGB_CDK, XGB_CDKextended, MLP_CDKextended, MLP_KlekotaRoth, and MLP_KlekotaRothCount had the highest R^2_{Tr} at 0.99, and the lowest MAE$_{Tr}$ ranged from 0.08 to 0.11 compared to other models. However, the highest Q^2_{CV} and the lowest MAE$_{CV}$ of 10-CV were found in the SVR_CDK, and the SVR_CDKextended (Q^2_{CV} = 0.77, 0.77, 0.76, and the MAE$_{CV}$ ranged from 0.47 to 0.49, respectively) as displayed in Table 2. In contrast, XGB_CDK, XGB_CDK, MLP_CDKex-

tended, MLP_KlekotaRoth, and MLP_KlekotaRothCount extended yielded the Q^2_{CV} only at 0.64−0.73, respectively. Those results signify that the CDK and CDK extended of SVR have better goodness of fit than the XGB_CDK, XGB_CDKextended, MLP_CDKextended, MLP_ KlekotaRoth, and MLP_ KlekotaRothCount models.

Table 3 shows that the Q^2_{Ext} of SVR_CDK, SVR_CDKextended, and SVR_KlekotaRoth produced the top three highest Q^2_{Ext} at 0.81, 0.81, and 0.80, respectively. Besides, the MAE$_{Ext}$ of the SVR_CDK, SVR_CDKextended, and SVR_KlekotaRoth also exhibited the top three lowest MAE$_{Ext}$ values similar to the 10-fold CV model performances. As a result, the top three baseline models for predicting BRAF affinity are SVR_CDK, SVR_CDKextended, and SVR_KlekotaRoth.

The goodness of fit of the baseline models was also assessed using R^2_{Tr}−Q^2_{CV} and R^2_{Tr}−Q^2_{Ext} metrics. We found that the R^2_{Tr}−Q^2_{CV} and R^2_{Tr}−Q^2_{Ext} metrics of the SVR_CDK, SVR_CDK extended, and SVR_KlekotaRoth were 0.21 and 0.16−0.17, respectively. An R^2_{Tr}−Q^2_{CV} metric higher than 0.2 indicates that the top three baseline models were slightly overfitted models.[16] Furthermore, the highest Q^2_{Ext} of the baseline model was 0.81, which does not meet the desired
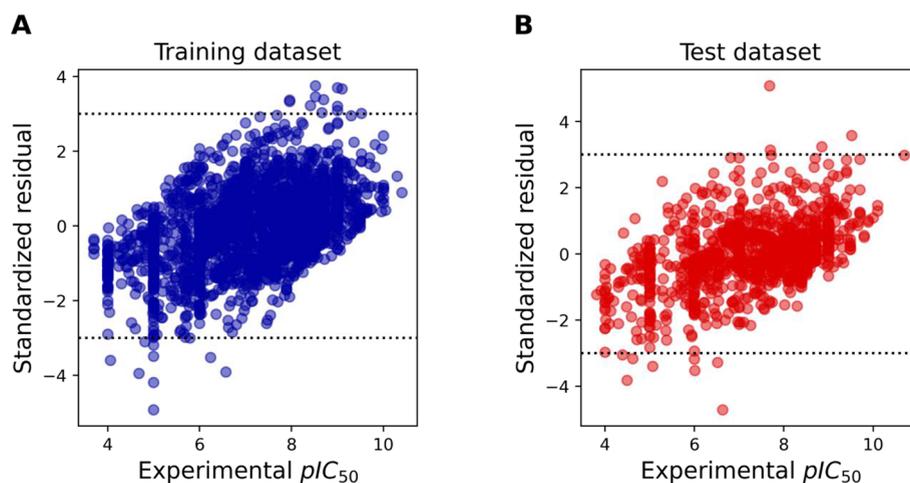
Figure 4. Standard residual plot and experimental pIC$_{50}$ analysis of the StackBRAF model. (A) Standard residual plot of the training dataset, and (B) standard residual plot of the test dataset. Blue and red dots indicate training and test datasets, respectively. The dashed line indicates an absolute standardized residual at 3.

predictability performance. Therefore, to reduce overfitting and improve the predictability of the BRAF model, a stacking ensemble learning model was constructed.

3.3. Stacking Ensemble Model. Figure 2 illustrates the performance of the stacking ensemble model for predicting BRAF affinity (StackBRAF). Figure 2A−C depicts the StackBRAF model with the R^2_{Tr}, Q^2_{CV}, and Q^2_{Ext} at 0.93, 0.80, and 0.84, while the MAE_{Tr}, MAE_{CV}, and MAE_{Ext} of the StackBRAF model were 0.28, 0.44, and 0.40, respectively. We observed that the StackBRAF provided the highest Q^2_{CV} and Q^2_{Ext} values, while yielding the lowest MAE_{CV} and MAE_{Ext} values compared to the baseline models. Additionally, the differences between $R^2_{Tr}-Q^2_{CV}$ and $R^2_{Tr}-Q^2_{Ext}$ metrics were 0.13 and 0.09, which were significantly lower than 0.2 and lower than the top three high-performance baseline models, respectively. Comparing Figure 2 with Tables 2 and 3, the results indicate that the StackBRAF model exhibits the highest goodness of fit and predictability compared to baseline models.

We examined the false $x−y$ correlation using y-randomization experiments (Figure 2D). The results reveal that the y-randomization models (yellow circles, n = 100) were in the opposite quadrant compared to the original model (red circles). The average of $R^2_{y-random}$ and $Q^2_{y-random}$ was 0.32 ± 0.02 and −0.01 ± 0.01, respectively. The significant difference between the R^2_{Tr} and Q^2_{Ext} of the original model and the $R^2_{y-random}$ and $Q^2_{y-random}$ of the y-randomization models proves that the correlation between the x variables and y variables of the StackBRAF model is reliable. This result confirms the robustness of the model.

Previous studies have constructed the validated QSAR models for designing new drugs against the BRAF protein (Table 4).[8−11] Most of them utilized 3D descriptors calculated from comparative molecular force field and comparative molecular similarity indices analysis models. They also applied the PLS as a regression model with steric, electrostatic, hydrophobic, and hydrogen bonding between the drug and target as molecular descriptors.[8−11] The training datasets of previous studies included 27−243 compounds, while the test datasets comprised 15−189 compounds. In comparison, our StackBRAF has 11−100 times larger training and 6−77 times larger test datasets than those of previous studies. The StackBRAF demonstrated the highest goodness of fit in the

CV test (Q^2_{CV}) compared to the four previous studies.[8−11] StackBRAF also exhibits better predictability (Q^2_{Ext}) than the three previous studies of the 3D-QSAR model.[9−11] Overall, our StackBRAF model offers improved predictability, robustness, and applicability across a wide range of chemical molecules.

3.4. Model Interpretation. Figure 3 highlights the importance of PFs in the StackBRAF model using the SHAP algorithm. In Figure 3A, the high and low PF values influencing the StackBRAF's output are represented by red and blue colors, respectively. The mean SHAP value demonstrates the impact each PF has on the model output value. We found that the top five important features include SVR_CDK (mean |SHAP| = 0.15), SVR_KlekotaRoth (mean |SHAP| = 0.14), SVR_CDKextended (mean |SHAP| = 0.12), the XGB_CDK (mean |SHAP| = 0.08), and SVR_PubChem (mean |SHAP| = 0.07) (Figure 3B). The top three most important features of the StackBRAF align with the top three baseline models. Notably, the top five PFs have a combined average impact of up to 0.56 on the pIC$_{50}$ value of the model's output.

Figure 3B also reveals the low-importance PFs of the StackBRAF model. The MLP_Estate, SVR_EState, MLP_MACCS, MLP_AtomPairs2D, and SVR_SubstructureCount all have the mean |SHAP| value of 0.004, which is less than 0.01. This indicates that the pIC$_{50}$ is rarely influenced by these features. Thus, the stacking ensemble model is primarily influenced by the accurate predictive fingerprints and less influence by the low-accuracy ones. Notably, computing the SHAP algorithm for each SVR_CDK, SVR_KlekotaRoth, and SVR_CDKextended fingerprints was hindered by system shutdown and kernel issues with the SVR algorithm.

3.5. Applicability Domain. The ML model cannot predict all chemical substances, only the chemicals utilized in the training dataset will produce very accurate predictions. The application scope of the model is defined as the evaluation of the chemical having a standardized residual lower than ±3. The standardized residual is calculated as the ratio between the difference of the observed and predicted value of a given molecule (residual) and the standard deviation of the residuals of all molecules in the dataset. The compound with an absolute standardized residual higher than ±3 was considered an
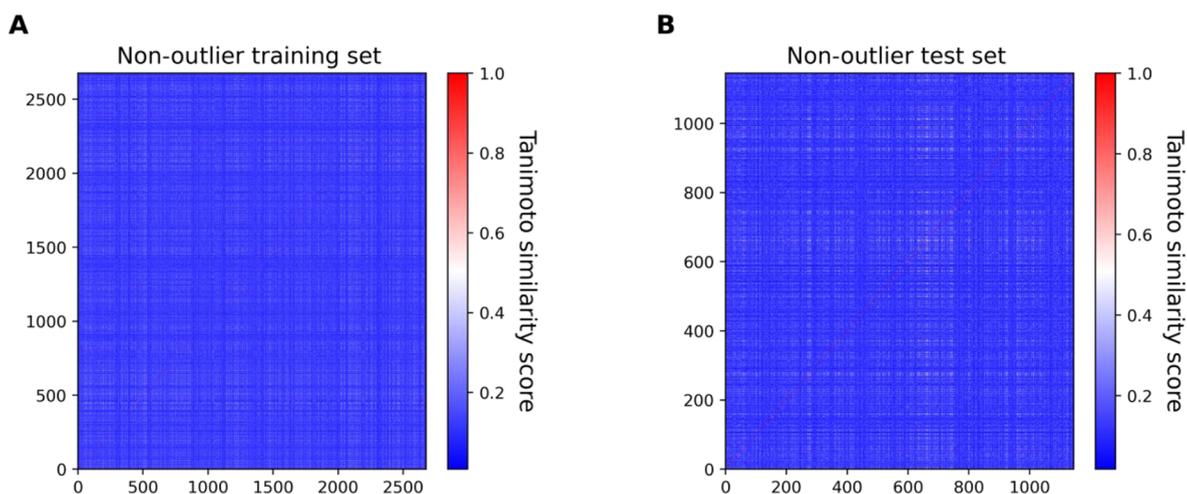
Figure 5. Tanimoto similarity score of the compound in non-outlier of (A) training and (B) test datasets. The red (1.0), white (0.5), and blue (0.0) colors illustrate high similarity, median similarity, and low similarity among the compounds, respectively.

Chart 1. Top Five Predictive pIC$_{50}$ of FDA-Approved Drugs[a]

Name	Predictive pIC$_{50}$				Average actual pIC$_{50}$	Similarity		Structure
	SVR_CDK	SVR_CDK Extended	SVR_Klekota Roth	StackB RAF		training	test	
Encorafenib	8.37	8.09	8.14	8.19	9.00	0.82	0.85	
Vemurafenib	8.20	8.13	8.22	8.00	7.46	0.83	0.82	
Regorafenib	6.80	6.75	7.84	7.02	7.28	0.62	0.65	
Ponatinib	6.84	6.77	7.01	6.70	Not reported	0.52	0.58	
Sorafenib	6.47	6.09	7.01	6.59	7.23	0.69	0.66	
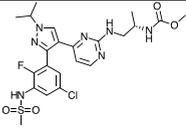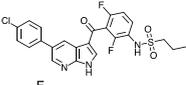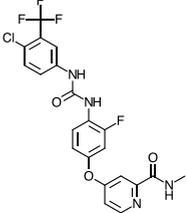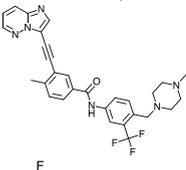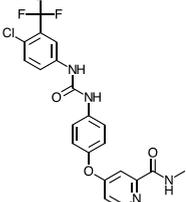
[a]NR: Not reported in the ChEMBL database.

outlier. The result shows that 21 molecules from the training dataset and 12 molecules from the test dataset contained higher ±3 standardized residual values (Figure 4A,B). This means the outliers of this model contained 33 molecules from both the training and test dataset (Supporting Information). Noted that the standardized residuals of low experimental

pIC$_{50}$ values (pIC$_{50}$ < 6) tend to exceed the −3 standardized residuals, while the high experimental pIC$_{50}$ values (pIC$_{50}$ > 6) tend to increase higher than 3 standardized residuals in both training and test dataset (Figure 4A,B). This result implies that the StackBRAF model could overestimate low and high experimental pIC$_{50}$ if the molecules are similar to the outliers.

Figure 5 depicts the Tanimoto similarity of the non-outlier compounds in the training and test dataset. This experiment defined the similarity degree of the new compound compared to the non-outliers training and test datasets. High similarity scores (closer to 1) indicate high chemical structure similarity, while low similarity scores (closer to 0) indicate low similarity between two compounds. The red (1.0), green (0.5), and blue (0.0) colors represent low, median, and high chemical diversity between the pairs of compounds, respectively. We found that the average similarity scores of the compounds in the non-outlier training and test datasets were both 0.13 ± 0.09, suggesting high molecular diversity and strong model generalization. Moreover, a median similarity score of 0.5 was set as the minimum acceptable level of similarity before categorizing a compound into the applicability domain. Compounds with similar structures to the non-outlier training or test dataset will obtain a confident prediction of the pIC_{50} value from the StackBRAF model.

3.6. Prediction of BRAF Inhibitor from 2123 FDA-Approved Drugs. The performance of the StackBRAF algorithm was examined using 2123 curated FDA-approved drugs for predicting BRAF inhibitors. Drugs with a Tanimoto similarity score below 0.5 when compared to the non-outliers training and test datasets were excluded. Chart 1 illustrates the top five predictive pIC_{50} values from the StackBRAF model and the top three baseline models alongside the average actual pIC_{50} value found in the ChEMBL database. The StackBRAF model identified encorafenib, vemurafenib, regorafenib, ponatinib, and sorafenib as the top five FDA-approved drugs with the highest predictive pIC_{50} values, which were 8.19, 8.00, 7.02, 6.70, and 6.59, respectively. In contrast, their actual pIC_{50} values were 9.00, 7.46, 7.28, not reported (NR), and 7.23, respectively (Chart 1). When evaluating the prediction error using the MAE metric, StackBRAF had the lowest MAE at 0.56, surpassing SVR_CDK, SVR_CDKExtended, and SVR_KlekotaRoth, which had values of 0.65, 0.81, and 0.60, respectively. This result indicates that the StackBRAF had the lowest MAE compared to the top three baseline models.

Encorafenib and vemurafenib are FDA-approved BRAF V600E inhibitors for melanoma patients.[35,36] Regorafenib is an FDA-approved multitarget kinase inhibitor that targets BRAF and RAF-1 proteins.[37] The subsequent two compounds, ponatinib and sorafenib, are FDA-approved drugs for chronic myeloid leukemia with BCR mutation and advanced renal cell carcinoma with BRAF mutation, respectively. These top five predictions of FDA-approved drugs, including four FDA-approved BRAF inhibitors, suggest that the StackBRAF model can understand the molecular fingerprints of the drug targeting BRAF protein. On top of that, the running time for predicting a single molecule was less than 8 s. Thus, this result suggests that the StackBRAF algorithm can be used in BRAF inhibitor ligand-based drug design.

4. DISCUSSION

This study provides valuable proof of concept, demonstrating that the stacking ensemble ML model can outperform the baseline models and increase accuracy for BRAF inhibitory activity prediction. The advantages of stacking ensemble learning include the fact that the stacked model does not depend on only one ML model, which could lead to bias and prediction error. Instead, the stacked model utilizes the abilities of several well-performing algorithms to make an optimal decision that surpasses any of the individual algorithms

used to build the ensemble model. In addition, the stacked model exhibits improved diversity due to the different baseline algorithms employed in building PFs, such as tree-based models, support vector models, or ANN models. Nevertheless, the primary limitation of stacked models is that they can require high computational time when training with large datasets.[38] Despite that the stacked models have been found to achieve high accuracies and have been applied in numerous drug design and healthcare prediction problems.[25,39−41]

The baseline algorithms that we used offer distinct benefits and diversity. The advantages of the XGB model include its fast-running time and high-performance model prediction, derived from the gradient-boosting tree-based method.[42] The XGB has been widely used in QSAR modeling due to its superior performance compared to other ML models.[43] The SVR algorithm provides excellent generalization capabilities and high prediction accuracy, which are based on its non-linear kernel, and epsilon tube.[44] The epsilon tube aims to find the most optimal flattest tube containing the training dataset while balancing model complexity and prediction error.[44] In our study, the performances of the SVR models ranked among the top three baseline models and contributed to the top three PFs in terms of SHAP importance. The benefits of an MLP regressor include its ability to solve non-linear problems and perform well with large datasets. In our study, we used 100 nodes with 3 hidden layers to construct the MLP models. However, we noticed that the MLP models exhibited lower importance compared to the SVR and XGB models as identified by the SHAP experiment (Figure 3). This might be due to the nature of neural networks, which require model optimization and an extensive dataset for efficient model training.[45]

To summarize, the StackBRAF model was developed with experiment pIC_{50} values against BRAF protein as the target endpoint. The applicability domain of the model was well-defined for the non-outlier training and test datasets. The goodness of fit, the model's robustness, and the model's predictability were evaluated with both CV and external validation. The StackBRAF outperforms the established acceptance criteria for a high-performance QSAR model, which include R^2 and $Q^2 > 0.6$, $R^2 - Q^2 \leq 0.2$, and MAE < 1.[33] Additionally, the StackBRAF also exhibits superior validated performance (Q^2 values) higher than the previously reported models as well as the baseline models.

5. CONCLUSIONS

This paper demonstrates the process of building, validating, and utilizing a stacking ensemble ML model called StackBRAF for predicting BRAF inhibitors. The first layer of the StackBRAF model consists of outputs from 36 PFs, constructed by combining 12 molecular fingerprints with XGB, MLP, and SVR models. The final layer of StackBRAF employed a RF regressor that takes the 36 PFs as input. The StackBRAF model offers a fast-running time, high accuracy, and low prediction error compared to both the baseline model and other previously published. We also applied the Tanimoto similarity score for screening the applicability domain of the StackBRAF model to increase the confident prediction of the model. Consequently, this model can be used in BRAF inhibitors drug design or FDA-drug repurposing strategies against BRAF protein. To expedite the usability of the StackBRAF model, we have hosted our model at https://

github.com/taraponglab/stackbraf for further use in research and development.

ASSOCIATED CONTENT

Data Availability Statement

All data used in this manuscript are available in the Supporting Information. The StackBRAF source code and software can be downloaded at https://github.com/taraponglab/stackbraf. Any questions or feedback regarding the StackBRAF software can directly inquire via email to the corresponding author (tarasri@kku.ac.th).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c01641.

> Raw dataset of BRAF inhibitors obtained from ChEMBL5145, training dataset, test dataset, FDA-approved drugs dataset, outlier compound dataset, and outlier structures (PDF)

AUTHOR INFORMATION

Corresponding Author

Tarapong Srisongkram − *Division of Pharmaceutical Chemistry, Faculty of Pharmaceutical Sciences, Khon Kaen University, Khon Kaen 40002, Thailand; Human High Performance and Health Promotion Research Institute, Khon Kaen University, Khon Kaen 40002, Thailand;* ● orcid.org/0000-0001-8512-5379; Email: tarasri@kku.ac.th

Authors

Nur Fadhilah Syahid − *Graduate School in the Program of Pharmaceutical Chemistry and Natural Products, Faculty of Pharmaceutical Sciences, Khon Kaen University, Khon Kaen 40002, Thailand*

Natthida Weerapreeyakul − *Division of Pharmaceutical Chemistry, Faculty of Pharmaceutical Sciences, Khon Kaen University, Khon Kaen 40002, Thailand; Human High Performance and Health Promotion Research Institute, Khon Kaen University, Khon Kaen 40002, Thailand*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c01641

Author Contributions

Conceptualization, N.F.S. and T.S.; software, N.F.S. and T.S.; validation, T.S.; formal analysis, N.F.S. and T.S.; investigation, N.F.S. and T.S.; resources, T.S.; data curation, T.S.; writing—original draft, N.F.S., N.W. and T.S.; writing—review and editing, N.F.S., N.W., and T.S.; visualization, T.S.; supervision, N.W. and T.S.; project administration, T.S.; funding acquisition, T.S. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ABBREVIATIONS

ANN   artificial neural network
CV   cross-validation
PF   predictive features
FDA   Food and Drug Administration
IC_{50} inhibitor concentration at 50 percent
MAE   mean absolute error
MAPK   mitogen-activated protein kinase
MLP   multilayer perceptron
PLS   partial least square
QSAR   quantitative structure−activity relationship
RAF   rapidly accelerated fibrosarcoma
RAS   rat sarcoma
RF   random forest
SHAP   SHapley Additive exPlanations
SMILES   simplified molecular input line entry system
SVR   support vector regression
t-SNE   t-Distributed Stochastic Neighbor Embedding

REFERENCES

(1) Leicht, D. T.; Balan, V.; Kaplun, A.; Singh-Gupta, V.; Kaplun, L.; Dobson, M.; Tzivion, G. Raf Kinases: Function, Regulation and Role in Human Cancer. *Biochim. Biophys. Acta, Mol. Cell Res.* **2007**, *1773*, 1196−1212.

(2) Wellbrock, C.; Karasarides, M.; Marais, R. The RAF Proteins Take Centre Stage. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 875−885.

(3) Catalanotti, F.; Reyes, G.; Jesenberger, V.; Galabova-Kovacs, G.; de Matos Simoes, R.; Carugo, O.; Baccarini, M. A Mek1−Mek2 Heterodimer Determines the Strength and Duration of the Erk Signal. *Nat. Struct. Mol. Biol.* **2009**, *16*, 294−303.

(4) Ardekani, G. S.; Jafarnejad, S. M.; Tan, L.; Saeedi, A.; Li, G. The Prognostic Value of BRAF Mutation in Colorectal Cancer and Melanoma: A Systematic Review and Meta-Analysis. *PLoS One* **2012**, *7*, No. e47054.

(5) Lavoie, H.; Therrien, M. Regulation of RAF Protein Kinases in ERK Signalling. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 281−298.

(6) Agianian, B.; Gavathiotis, E. Current Insights of BRAF Inhibitors in Cancer. *J. Med. Chem.* **2018**, *61*, 5775−5793.

(7) Rizos, H.; Menzies, A. M.; Pupo, G. M.; Carlino, M. S.; Fung, C.; Hyman, J.; Haydu, L. E.; Mijatov, B.; Becker, T. M.; Boyd, S. C.; Howle, J.; Saw, R.; Thompson, J. F.; Kefford, R. F.; Scolyer, R. A.; Long, G. V. BRAF Inhibitor Resistance Mechanisms in Metastatic Melanoma: Spectrum and Clinical Impact. *Clin. Cancer Res.* **2014**, *20*, 1965−1977.

(8) Ai, Y.; Wang, S.-T.; Tang, C.; Sun, P.-H.; Song, F.-J. 3D-QSAR and Docking Studies on Pyridopyrazinones as BRAF Inhibitors. *Med. Chem. Res.* **2011**, *20*, 1298−1317.

(9) Singh, A. K; Novak, J.; Kumar, A.; Singh, H.; Thareja, S.; Pathak, P.; Grishina, M.; Verma, A.; Yadav, J. P.; Khalilullah, H.; Pathania, V.; Nandanwar, H.; Jaremko, M.; Emwas, A.-H.; Kumar, P. Gaussian Field-Based 3D-QSAR and Molecular Simulation Studies to Design Potent Pyrimidine−Sulfonamide Hybrids as Selective BRAF^{V600E} Inhibitors. *RSC Adv.* **2022**, *12*, 30181−30200.

(10) Shih, K.-C.; Lin, C.-Y.; Zhou, J.; Chi, H.-C.; Chen, T.-S.; Wang, C.-C.; Tseng, H.-W.; Tang, C.-Y. Development of Novel 3D-QSAR Combination Approach for Screening and Optimizing B-Raf Inhibitors in Silico. *J. Chem. Inf. Model.* **2011**, *51*, 398−407.

(11) Wesley, L.; Veerapaneni, S.; Desai, R.; McGee, F.; Joglekar, N.; Rao, S.; Kamal, Z. 3D-QSAR and SVM Prediction of BRAF-V600E and HIV Integrase Inhibitors: A Comparative Study and Characterization of Performance with a New Expected Prediction Performance Metric. *Am. J. Biochem. Biotechnol.* **2016**, *12*, 253−262.

(12) Yang, L.; Yang, G.; Bing, Z.; Tian, Y.; Niu, Y.; Huang, L.; Yang, L. Transformer-Based Generative Model Accelerating the Development of Novel BRAF Inhibitors. *ACS Omega* **2021**, *6*, 33864−33873.

(13) Lo, Y.-C.; Rensi, S. E.; Torng, W.; Altman, R. B. Machine Learning in Chemoinformatics and Drug Discovery. *Drug Discovery Today* **2018**, *23*, 1538−1546.

(14) Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R. K. Artificial Intelligence in Drug Discovery and Development. *Drug Discovery Today* **2021**, *26*, 80−93.

(15) Chen, X.; Xie, W.; Yang, Y.; Hua, Y.; Xing, G.; Liang, L.; Deng, C.; Wang, Y.; Fan, Y.; Liu, H.; Lu, T.; Chen, Y.; Zhang, Y. Discovery of Dual FGFR4 and EGFR Inhibitors by Machine Learning and Biological Evaluation. *J. Chem. Inf. Model.* **2020**, *60*, 4640−4652.

(16) Suvannang, N.; Preeyanon, L.; Malik, A. A.; Schaduangrat, N.; Shoombuatong, W.; Worachartcheewan, A.; Tantimongcolwat, T.; Nantasenamat, C. Probing the Origin of Estrogen Receptor Alpha Inhibition *via* Large-Scale QSAR Study. *RSC Adv.* **2018**, *8*, 11344−11356.

(17) Srisongkram, T.; Weerapreeyakul, N. Drug Repurposing against KRAS Mutant G12C: A Machine Learning, Molecular Docking, and Molecular Dynamics Study. *Int. J. Mol. Sci.* **2023**, *24*, 669.

(18) Wang, S.; Di, J.; Wang, D.; Dai, X.; Hua, Y.; Gao, X.; Zheng, A.; Gao, J. State-of-the-Art Review of Artificial Neural Networks to Predict, Characterize and Optimize Pharmaceutical Formulation. *Pharmaceutics* **2022**, *14*, 183.

(19) Wolpert, D. H. Stacked Generalization. *Neural Networks* **1992**, *5*, 241−259.

(20) OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] model.* https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2 (accessed Aug 25, 2022).

(21) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.

(22) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J. Chem. Inf. Model.* **2016**, *56*, 1243−1252.

(23) Matsson, P.; Kihlberg, J. How Big Is Too Big for Cell Permeability? *J. Med. Chem.* **2017**, *60*, 1662−1664.

(24) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466−1474.

(25) Schaduangrat, N.; Anuwongcharoen, N.; Moni, M. A.; Lio', P.; Charoenkwan, P.; Shoombuatong, W. StackPR Is a New Computational Approach for Large-Scale Identification of Progesterone Receptor Antagonists Using the Stacking Strategy. *Sci. Rep.* **2022**, *12*, 16435.

(26) Hähnke, V. D.; Kim, S.; Bolton, E. E. PubChem Chemical Structure Standardization. *J Cheminform.* **2018**, *10*, 36.

(27) Awale, M.; Reymond, J.-L. Atom Pair 2D-Fingerprints Perceive 3D-Molecular Shape and Pharmacophores for Very Fast Virtual Screening of ZINC and GDB-17. *J. Chem. Inf. Model.* **2014**, *54*, 1892−1907.

(28) Willighagen, E. L.; Mayfield, J. W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliazkova, N.; Kuhn, S.; Pluskal, T.; Rojas-Chertó, M.; Spjuth, O.; Torrance, G.; Evelo, C. T.; Guha, R.; Steinbeck, C. The Chemistry Development Kit (CDK) v2.0: Atom Typing, Depiction, Molecular Formulas, and Substructure Searching. *Aust. J. Chem.* **2017**, *9*, 33.

(29) 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.

(30) Klekota, J.; Roth, F. P. Chemical Substructures That Enrich for Biological Activity. *Bioinformatics* **2008**, *24*, 2518−2525.

(31) 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*, 1273−1280.

(32) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202−D1213.

(33) Nakarin, F.; Boonpalit, K.; Kinchagawat, J.; Wachiraphan, P.; Rungrotmongkol, T.; Nutanong, S. Assisting Multitargeted Ligand Affinity Prediction of Receptor Tyrosine Kinases Associated Non-small Cell Lung Cancer Treatment with Multitasking Principal Neighborhood Aggregation. *Molecules* **2022**, *27*, 1226.

(34) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56−67.

(35) Davis, J.; Wayman, M. Encorafenib and Binimetinib Combination Therapy in Metastatic Melanoma. *J. Adv. Pract. Oncol.* **2022**, *13*, 450−455.

(36) Kim, G.; McKee, A. E.; Ning, Y.-M.; Hazarika, M.; Theoret, M.; Johnson, J. R.; Xu, Q. C.; Tang, S.; Sridhara, R.; Jiang, X.; He, K.; Roscoe, D.; McGuinn, W. D.; Helms, W. S.; Russell, A. M.; Miksinski, S. P.; Zirkelbach, J. F.; Earp, J.; Liu, Q.; Ibrahim, A.; Justice, R.; Pazdur, R. FDA Approval Summary: Vemurafenib for Treatment of Unresectable or Metastatic Melanoma with the BRAFV600E Mutation. *Clin. Cancer Res.* **2014**, *20*, 4994−5000.

(37) Strumberg, D.; Schultheis, B. Regorafenib for Cancer. *Expert Opin. Invest. Drugs* **2012**, *21*, 879−889.

(38) Mienye, I. D.; Sun, Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access* **2022**, *10*, 99129−99149.

(39) Ahmad, S.; Charoenkwan, P.; Quinn, J. M. W.; Moni, M. A.; Hasan, M. M.; Lio', P.; Shoombuatong, W. SCORPION Is a Stacking-Based Ensemble Learning Framework for Accurate Prediction of Phage Virion Proteins. *Sci. Rep.* **2022**, *12*, 4106.

(40) Grenet, I.; Merlo, K.; Comet, J.-P.; Tertiaux, R.; Rouquié, D.; Dayan, F. Stacked Generalization with Applicability Domain Outperforms Simple QSAR on *in Vitro* Toxicological Data. *J. Chem. Inf. Model.* **2019**, *59*, 1486−1496.

(41) Hwangbo, L.; Kang, Y. J.; Kwon, H.; Lee, J. I.; Cho, H.-J.; Ko, J.-K.; Sung, S. M.; Lee, T. H. Stacking Ensemble Learning Model to Predict 6-Month Mortality in Ischemic Stroke Patients. *Sci. Rep.* **2022**, *12*, 17389.

(42) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016; pp 785−794. DOI: 10.1145/2939672.2939785.

(43) Sheridan, R. P.; Wang, W. M.; Liaw, A.; Ma, J.; Gifford, E. M. Extreme Gradient Boosting as a Method for Quantitative Structure−Activity Relationships. *J. Chem. Inf. Model.* **2016**, *56*, 2353−2360.

(44) Awad, M.; Khanna, R. Support Vector Regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Awad, M., Khanna, R., Eds.; Apress: Berkeley, CA, 2015; pp 67−80. DOI: 10.1007/978-1-4302-5990-9_4.

(45) Alom, M. Z.; Taha, T. M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M. S.; Hasan, M.; Van Essen, B. C.; Awwal, A. A. S.; Asari, V. K. A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics* **2019**, *8*, 292.