ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Machine Learning-Based Approach to Developing Potent EGFR Inhibitors for Breast Cancer—Design, Synthesis, and In Vitro Evaluation

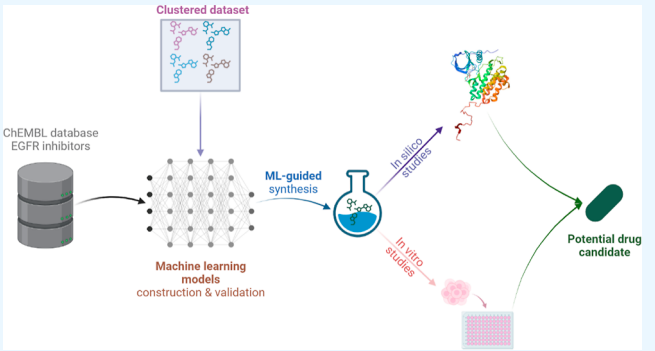
Hossam Nada, Anam Rana Gul, Ahmed Elkamhawy, Sungdo Kim, Minkyoung Kim, Yongseok Choi, Tae Jung Park,* and Kyeong Lee*

Cite This: *ACS Omega* 2023, 8, 31784−31800

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The epidermal growth factor receptor (EGFR) is vital for regulating cellular functions, including cell division, migration, survival, apoptosis, angiogenesis, and cancer. EGFR overexpression is an ideal target for anticancer drug development as it is absent from normal tissues, marking it as tumor-specific. Unfortunately, the development of medication resistance limits the therapeutic efficacy of the currently approved EGFR inhibitors, indicating the need for further development. Herein, a machine learning-based application that predicts the bioactivity of novel EGFR inhibitors is presented. Clustering of the EGFR small-molecule inhibitor (∼9000 compounds) library showed that *N*-substituted quinazolin-4-amine-based compounds made up the largest cluster of EGFR inhibitors (∼2500 compounds). Taking advantage of this finding, rational drug design was used to design a novel series of 4-anilinoquinazoline-based EGFR inhibitors, which were first tested by the developed artificial intelligence application, and only the compounds which were predicted to be active were then chosen to be synthesized. This led to the synthesis of 18 novel compounds, which were subsequently evaluated for cytotoxicity and EGFR inhibitory activity. Among the tested compounds, compound **9** demonstrated the most potent antiproliferative activity, with 2.50 and 1.96 μM activity over MCF-7 and MDA-MB-231 cancer cell lines, respectively. Moreover, compound **9** displayed an EGFR inhibitory activity of 2.53 nM and promising apoptotic results, marking it a potential candidate for breast cancer therapy.



1. INTRODUCTION

Cancer is a complex disease that involves many biological pathways, and its incidence continues to rise internationally.[1,2] One of the hallmarks of cancer is the uncontrolled proliferation of cells.[3] Among the various molecular factors that contribute to the progression of cancer, the epidermal growth factor receptor (EGFR) has emerged as a crucial factor in regulating cellular functions such as cell division, migration, survival, apoptosis, angiogenesis, inflammation, and cancer.[4−6] Overexpression of EGFR has been related to human epithelial malignancies such as breast, colon, head and neck, prostate, lung, and pancreas cancers, making it an appealing therapeutic target.[7−9]

Among these conditions, overexpression of EGFR has been heavily linked to increased tumor cell proliferation and poor survival rates in breast cancer patients.[10,11] EGFR plays a crucial role in regulating the development and balance of epithelial tissues.[12] When activated, EGFR phosphorylates and activates various proteins, initiating diverse signaling pathways that contribute to multiple cellular processes such as cell proliferation, apoptosis, angiogenesis, migration, adhesion, and invasion.[13] In cancer, the inappropriate activation of EGFR predominantly arises from amplification and point mutations at the genomic locus.[14,15] Additionally, transcriptional upregulation or excessive production of EGFR ligands through autocrine or paracrine mechanisms can also contribute to its aberrant activation.[12,16,17] Furthermore, EGFR is increasingly recognized as a biomarker of resistance in tumors as its amplification or acquisition of secondary mutations has been observed under the influence of therapeutic drugs.[18]

EGFR expression has also been connected to breast cancer differentiation loss.[19] As a result, EGFR may be a useful early prognostic marker in breast cancer patients, although it does not necessarily indicate long-term survival. These facts

Received: April 24, 2023
Accepted: August 11, 2023
Published: August 23, 2023

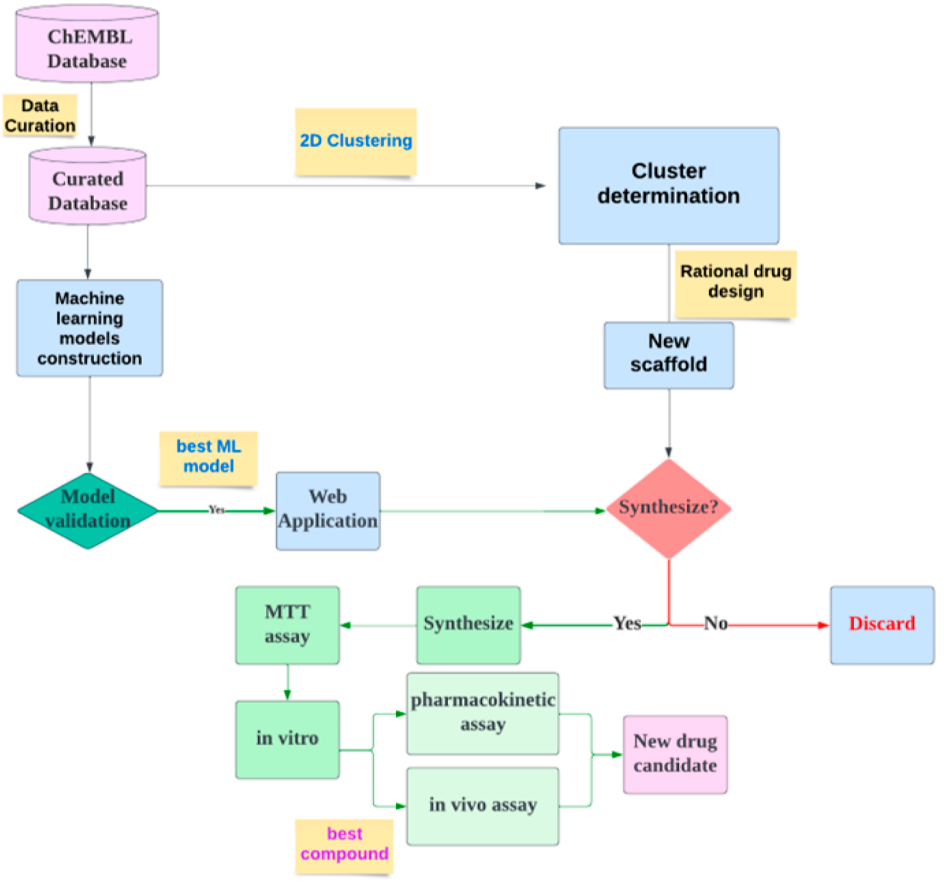


Figure 1. Design strategy of the study.

highlight the importance of EGFR as a therapeutic target for breast cancer.

The approval of gefitinib and erlotinib, first-generation reversible EGFR inhibitors, for the treatment of cancer indicates that EGFR-targeting drugs have evolved as a suitable strategy for cancer treatment.[20,21] The therapeutic efficacy of EGFR inhibitors is however limited by the emergence of drug resistance, emphasizing the need for more potent inhibitors.[22]

Unfortunately, the development of novel medicine is typically time-consuming and expensive and requires a significant number of resources.[23] The drug development process can be significantly speeded up by using cheminformatics.[24,25] Cheminformatics is a multidisciplinary field that uses computational and information technologies to find solutions to a wide range of problems in chemistry.[26] It has achieved exponential progress in the era of machine learning (ML) and artificial intelligence. In drug discovery, cheminformatics has long been applied to aid in the search for and optimization of new molecules.[24] One of the chemoinformatic solutions to speeding up the drug discovery process is the application of quantitative structure−activity relationship (QSAR) methodologies.[27,28] QSAR models generate statistically significant relationships between the chemical structure and biological activity, providing a faster and more efficient way of predicting new drug candidates. With the increased availability of chemical libraries of inhibitors, building QSAR models via ML has become more appealing.

In this study, a systematic cheminformatic analysis was performed on data sets gathered from the ChEMBL database to generate machine learning models for EGFR inhibitors. The best machine learning model was subsequently employed to construct a web application that facilitates the use of the chosen model's predictive capabilities. Clustering of the database based on their shared chemical structure led to the identification of the chemical structure with the greatest potential for creating novel EGFR inhibitors (quinazolines). Next, rational drug design was used to create a new series of EGFR inhibitors based on the identified cluster. Only the derivatives with promising predicted bioactivity, based on the results of testing each derivative in the web application, were synthesized. The synthesized compounds were tested for their anticancer activity, and the most promising derivatives were evaluated for their EGFR inhibitory activity. Finally, the most promising compound was subjected to apoptosis and cell cycle analysis. Molecular docking analysis was employed to elucidate the possible binding interactions of the compounds with EGFR. The design strategy of the study is illustrated in Figure 1.

2. RESULTS AND DISCUSSION

2.1. Machine Learning Models of EGFR and Bioactivity Prediction Application. Chemoinformatic techniques such as 2D-QSAR and 3D-pharmacophore ML models have been extensively used to complement experimental studies in identifying chemical characteristics of inhibitors and predicting their activity against various targets.[29,30] Compared to conventional in vitro methods, which can be time-consuming and labor-intensive, chemoinformatic techniques provide a quick and effective way to predict the activity of designed hybrid molecules.[31] Furthermore, as larger data sets

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

Table 1. Developed Machine Learning Algorithms

algorithm	type	description of hyperparameter setting	R^2 train	R^2 test
Random Forest	ensemble learning	*n*_estimators = 100, random_state = 42	0.959	0.717
Linear Regression	linear model		0.693	0.568
Ridge Regression	linear model	alpha = 1.0	0.693	0.573
Lasso Regression	linear model	alpha = 1.0	0.057	0.056
Elastic Net	linear model	alpha = 1.0, l1_ratio = 0.5	0.058	0.057
K-NN Regression	instance-based	*n*_neighbors = 5	0.494	0.212
SVM Regression	support vector	kernel = "linear", *C* = 1.0	0.614	0.489
MLP Regression	neural network	hidden_layer_sizes = (100), activation = "relu"	0.922	0.596
XGBoost	boosting	*n*_estimators = 100, learning_rate = 0.1, max_depth = 3	0.898	0.704
LightGBM	boosting	*n*_estimators = 100, learning_rate = 0.1, max_depth = 3	0.797	0.681
CatBoost	boosting	*n*_estimators = 100, learning_rate = 0.1, max_depth = 3	0.619	0.569

and chemical libraries become more widely available, ML has become an increasingly attractive option for harnessing these techniques. Herein, an application was built to predict the bioactivity of EGFR small-molecule inhibitors based on machine learning models. To achieve this goal, the following steps were followed: (i) a data set of EGFR inhibitors with a defined end point was compiled and organized; (ii) machine learning models were constructed and evaluated; and (iii) the best model was used to develop a web application using Python.

2.1.1. EGFR Inhibitors' Data Set Construction and Organization. A dataset of EGFR inhibitors was compiled from the ChEMBL database from an original database of 13914 compounds. The data set was curated by removing 4175 compounds with no reported IC_{50} values, resulting in a final data set of 9019 compounds (Supporting Information). The compounds were then classified into active, inactive, and intermediate inhibitors based on their IC_{50} values. Active compounds possessed IC_{50} values less than or equal to 1 μM, while inactive compounds had IC_{50} values greater than or equal to 10 μM. Next, IC_{50} values were converted to pIC_{50}.

2.1.2. Machine Learning Model Development, Evaluation, and Application Generation. The prepared database was used for construction of 11 ML models after the generation of fingerprint descriptors using RDKit AllChem software. The RDKit AllChem software was chosen due to it being a freely available tool with the ability to generate a broad range of molecular descriptors, including structural, topological, electronic, and thermodynamic properties of molecules.[32,33] The molecular descriptors used in this study included molecular weight, number of rotatable bonds, topological polar surface area (TPSA), and Morgan fingerprints with radii of 2 and 1024 bits. *K*-fold cross-validation was used as the primary validation method with *k* set to 5. The data were randomly split into 5 parts, with each part used once as a validation set and the other 4 parts used as the training set. This process was repeated 5 times, with each part used once as the validation set. The *R*-squared score was used as the evaluation metric to compare the performance of the models. The performance of the models was evaluated by calculating the mean *R*-squared scores for both the training and the validation sets. The performance results of each model as well as their hyperparameter setting are displayed in Table 1.

The results of the study showed that Random Forest had the highest mean training *R*-squared score of 0.959 and the highest mean validation *R*-squared score of 0.717 (Figure 2). The *R*-squared scores were calculated based on the average test score across all folds in the cross-validation process. This result

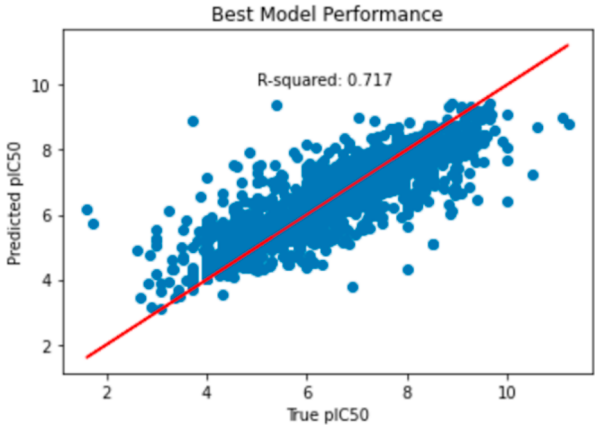


Figure 2. *R*-squared score of the Random Forest model based on the average test score across all folds in the cross-validation process.

indicated that the Random Forest ML model was deemed the most suitable model for EGFR activity prediction.

The second-best model was XGBoost with a mean training R-squared score of 0.898 and a mean validation *R*-squared score of 0.704. Multi-layer perceptron (MLP) Regression had the third highest mean training *R*-squared score of 0.922 and a mean validation *R*-squared score of 0.596. Linear Regression and Ridge Regression had similar mean training *R*-squared scores of around 0.693, but their mean validation *R*-squared scores were relatively low, ranging from 0.568 to 0.573. K nearest neighbor (K-NN) Regression had a mean training *R*-squared score of 0.494 and a relatively low mean validation *R*-squared score of 0.212. Lasso Regression had the lowest mean training *R*-squared score of 0.057 and the lowest mean validation *R*-squared score of 0.056. Support vector machine (SVM) Regression had a mean training *R*-squared score of 0.614 and a mean validation *R*-squared score of 0.489. LightGBM had a mean training *R*-squared score of 0.797 and a mean validation *R*-squared score of 0.681. CatBoost had a mean training *R*-squared score of 0.619 and a mean validation *R*-squared score of 0.569.

A python-based web application (Supporting Information) that can be run using Streamlit was then developed with the intention of being a user-friendly tool for researchers in drug discovery to predict the EGFR inhibitory activity of new drug candidates based on their chemical structures. The user interface of the web application is illustrated in Figure 3.

2.2. Rational Design and Activity Prediction of Novel EGFR Inhibitors. Using the Schrodinger Maestro cluster module, the 2D structures of the EGFR inhibitors in the

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

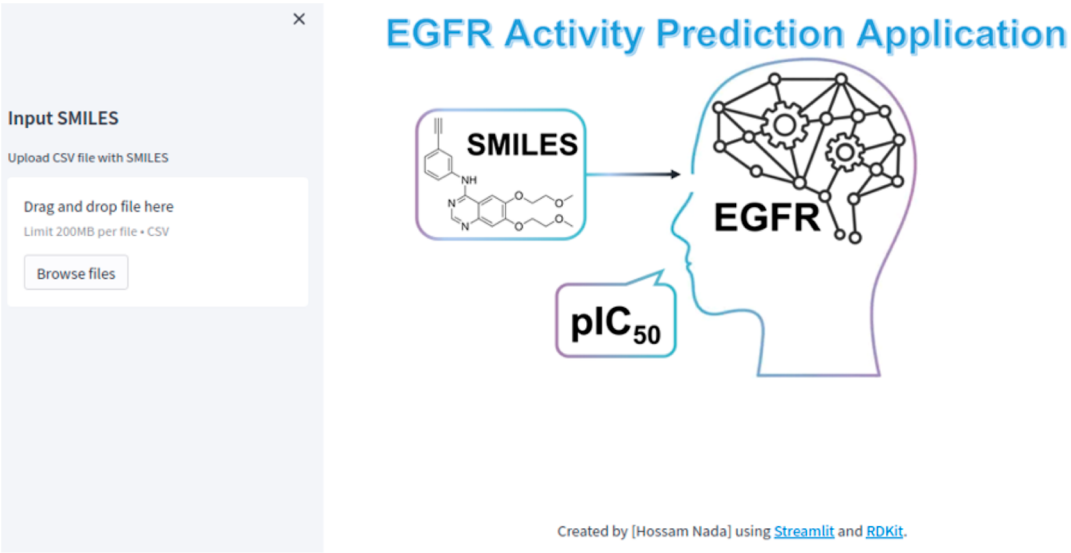


Figure 3. User interface of the EGFR activity prediction application. The figure shows the main components of the user interface, including the input form (smiles) and the prediction results (pIC_{50})

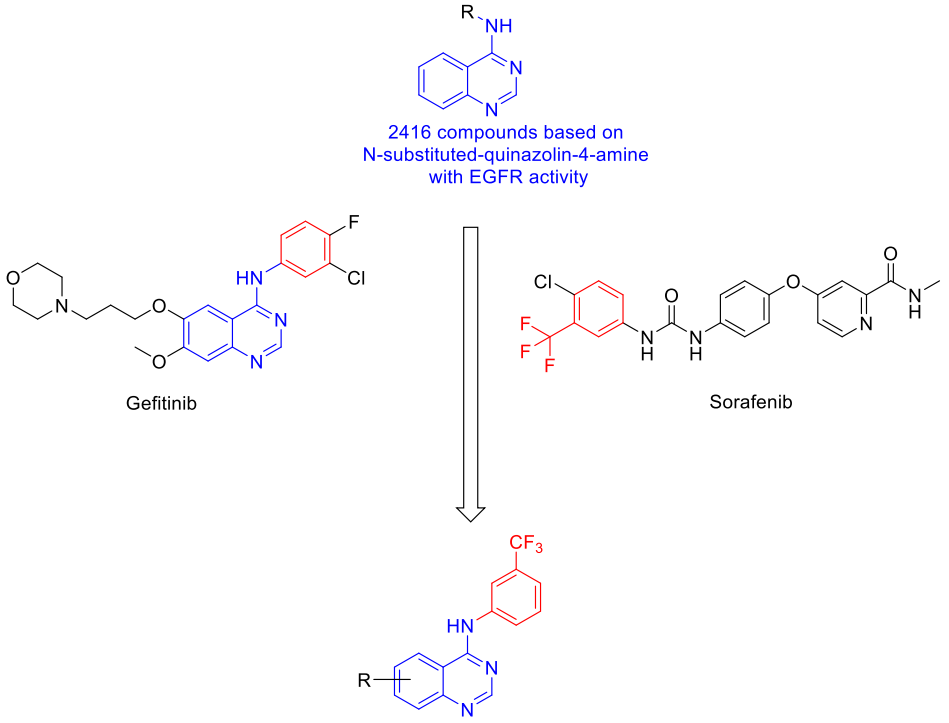


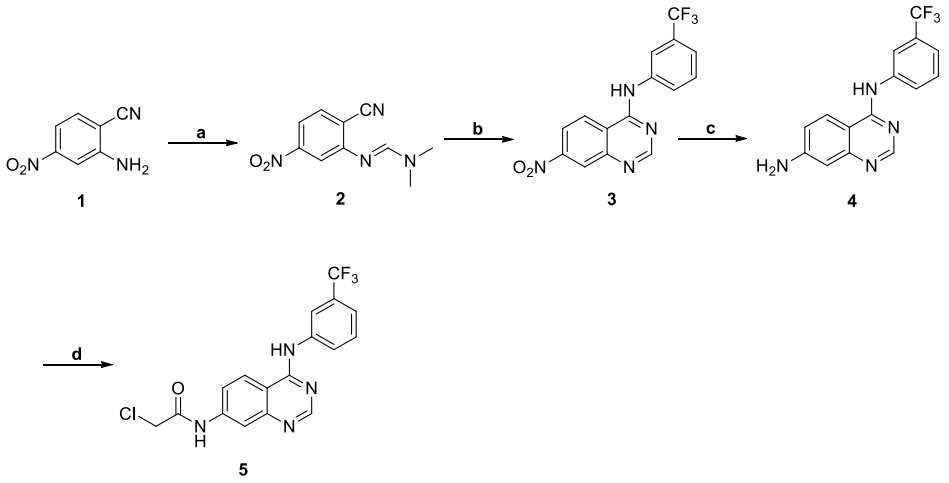
Figure 4. Design strategy employed in the design of the substituted *N*-(3-(trifluoromethyl)phenyl)quinazolin-4-amine scaffold.

curated database were clustered. This process led to the identification of 2416 compounds (Supporting Information) that were *N*-substituted quinazolin-4-amine-based, making them a potential scaffold for further development. The top-performing ML model (Random forest) was evaluated on *N*-substituted quinazolin-4-amine-based compounds, yielding an *R*-squared score of 0.86. Taking advantage of this discovery, a hybridization strategy was employed, combining the *N*-substituted-quinazolin-4-amine scaffold with various FDA-approved kinase inhibitors, resulting in the creation of a new scaffold. This hybridization involved the combination of the quinazolin-4-amine with a 3-(trifluoromethyl)benzene moiety found in gefitinib, sorafenib. The resulting structure (Figure 4) was then substituted at the 5 and 6 positions using amine and c−c linkers to yield novel structures. To streamline the process, the aforementioned EGFR bioactivity prediction application was utilized to predict the activity of the theorized compounds against EGFR. Only the compounds predicted to be active ($pIC_{50} \geq 6$) were synthesized, enabling the synthesis of 18 novel structures. The chemicals used in the synthesis of these compounds were already present in the laboratory, making this an efficient and cost-effective method for the creation of novel EGFR inhibitors.

2.3. Permutation Test. A permutation test was conducted to assess the ML model's performance on a subset of quinazolin-4-amine derivatives and determine if there is a

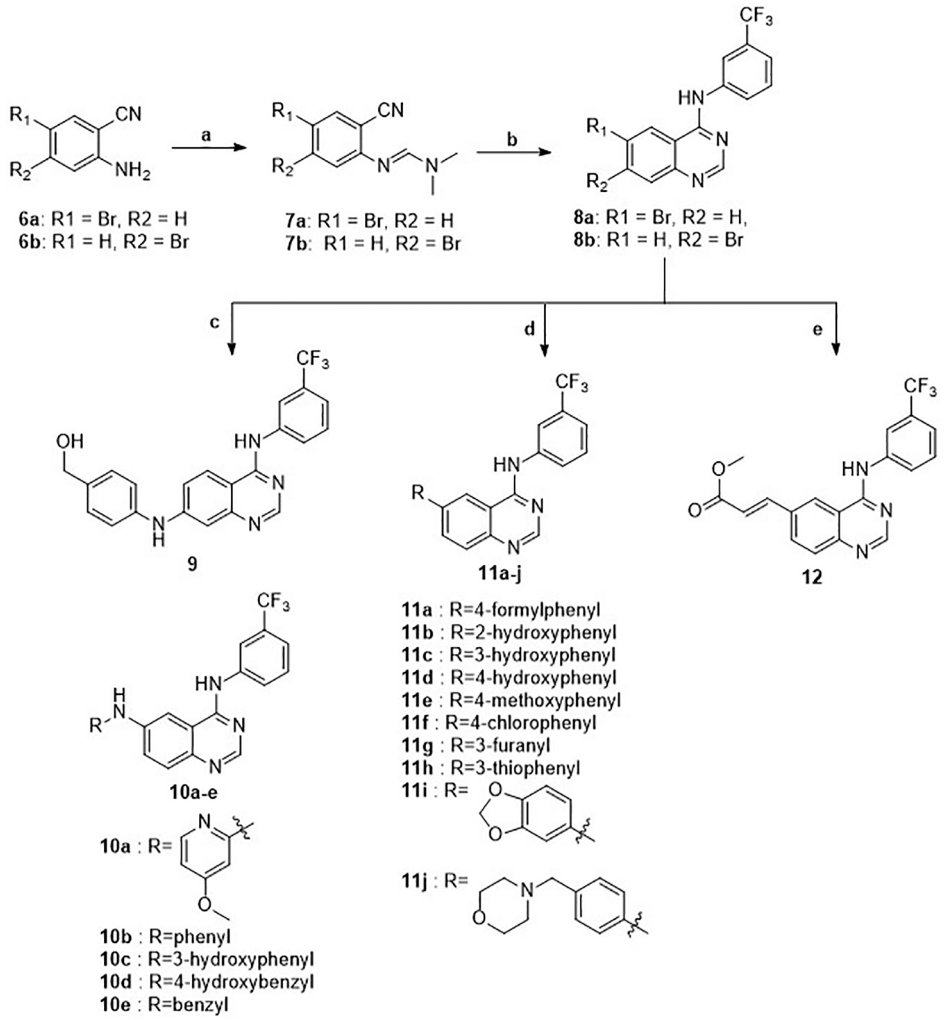
https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

Scheme 1. Synthesis of Compound 5[a]



[a]Reagents and conditions: (a) dimethylformamide dimethyl acetal, DCM, reflux, 1.5 h; (b) glacial acetic acid, 3-(trifluoromethyl) aniline, reflux, 2 h; (c) (i) aqueous EtOH(70% v/v), acetic acid, iron, reflux, 2 h; (ii) NH_3 solution, rt, 2 h; (d) ACN, aqueous NaOH, chloroacetyl chloride, rt, 3 h

Scheme 2. Synthesis of Compounds 9, 10a−e, 11a−j, and 12[a]



[a]Reagents and conditions: (a) dimethylformamide dimethyl acetal, reflux, 1.5 h; (b) glacial acetic acid, 3-(trifluoromethyl) aniline, reflux, 2 h; (c) toluene, amine derivatives, *t*-butyl XPhos, $Pd_2(dba)_3$, K_2CO_3, N_2 gas, 120 °C, 14 h, sealed tube; (d) dioxane, water, boronic acid derivatives, [1′1′-bis(diphenylphosphino)ferrocene]-dichloropalladium(II), K_2CO_3, 100 °C, 2 h, sealed tube; (e) DMF, tri(*o*-tolyl) phosphine, $Pd(OAc)_2$, Et_3N, methyl acrylate, reflux, 12 h.

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

significant difference in performance between the complete data set and the quinazolin-4-amine derivatives. In the permutation test, the hypothesis was as follows: null hypothesis (H0): the performance of the best ML model on the whole data set is not significantly different from its performance on the quinazoline data set. Alternative hypothesis (H1): the performance of the best ML model on the whole data set is significantly different from its performance on the quinazoline data set. The *p*-value calculated in the permutation test is used to assess the evidence against the null hypothesis. If the *p*-value is below a predefined significance level (e.g., 0.05), it would suggest that the observed difference in model performance is unlikely to occur by chance alone, providing evidence to reject the null hypothesis in favor of the alternative hypothesis. However, if the *p*-value is higher than the significance level (e.g., 0.05), it would suggest that the observed difference is likely due to random chance and there is insufficient evidence to reject the null hypothesis.

The model demonstrated a mean squared error (MSE) of 0.0302 on the entire data set, indicating a good overall performance. The permutation test yielded a high *p*-value of 1.0, suggesting that there was no statistically significant difference in the model's performance between the whole data set and the quinazoline subset. These results imply that the ML model performed consistently well on both the full data set and the quinazolin-4-amine derivatives, indicating the absence of bias or substantial performance discrepancy.

2.4. Chemistry. Scheme 1 shows the synthetic route of compound **5**. Synthesis of the intermediate compound **2** involved the reaction of 2-amino-4-nitrobenzonitrile (**1**) with dimethylformamide dimethyl acetal under a reflux condition.[34] The reaction of 3-(trifluoromethyl) aniline and **2** in glacial acetic acid provided ring-cyclized compound **3**. Reduction of compound **3** was performed using iron powder to obtain compound **4**. The final compound **5** was synthesized with chloroacetyl chloride and compound **4**.

The synthetic route used to obtain the desired compounds **9**, **10a−e**, **11a−j**, and **12** is outlined in Scheme 2. Refluxing the mixture of 2-amino-5-bromobenzonitrile (**6a**) or 2-amino-4-bromobenzonitrile (**6b**) with dimethylformamide dimethyl acetal afforded compounds **7a** and **7b**. Ring cyclization was performed to obtain compound **8a−b** using 3-(trifluoromethyl) aniline. The Buchwald−Hartwig reaction was carried out with the $Pd_2(dba)_3$ catalyst to afford compounds **9**, **10a−e** from reacting compounds **8a−b** with the appropriate anilines. Meanwhile, Suzuki cross-coupling of compound **8a** and the appropriate boronic acid derivatives was achieved using [1′1′-bis(diphenylphosphino)ferrocene]-dichloropalladium(II) as a catalyst to obtain compounds **11a−j**. Heck coupling was carried out to synthesize compound **12** via **8a** and palladium diacetate. The spectral NMR, high-resolution mass spectrometry (HRMS), and high-performance liquid chromatography (HPLC) data were examined in order to confirm the structure and purity of the target compounds. With the exception of compounds **11h** and **11j**, all synthesized compounds had a purity of at least 95%. Unfortunately, it was not possible to further purify compounds **11h** and **11j**, which had purities of 81.70 and 91.50%, respectively.

2.5. Biological Evaluation. *2.5.1. MTT Assay.* In this study, cell assays were carried out first to select compounds with good anticancer activity for the EGFR assay. This methodology was adopted because, while cell assays were conducted in-house, the EGFR assay was performed by Reaction Biology Corporation. Therefore, it was judged as more practical and cost-effective to first concentrate on compounds with notable anticancer activity and then conduct further assessments on their inhibitory activity against EGFR. This strategy enabled the prioritization of compounds with a higher probability of displaying both anticancer and EGFR inhibitory activities, thus optimizing the utilization of resources and efforts toward the primary objective of the study. To evaluate the impact of the target compounds on cell proliferation, an MTT assay was performed against two breast cancer cell lines: MCF-7 and MDA-MB-231. These cell lines were chosen because both cells are known to overexpress EGFR and are recognized for their high malignant potential compared to other common breast cancer cell lines. Additionally, MCF-7 cells are both multidrug-resistant and hormone-independent. The resulting IC_{50} values are summarized in Table 2.

Table 2. Cytotoxicity Displayed by the Synthesized Compounds against the Breast Cancer Cell Lines MCF-7 and MDA-MB-231[a]

compound	IC_{50} value against MCF-7 (μM)	±SE	IC_{50} value against MDA-MB-231 (μM)	±SE
5	7.777	0.007	6.877	0.035
9	2.497	0.105	1.954	0.032
10a	5.841	0.020	6.624	0.080
10b	11.4	0.042	15.45	0.085
10c	20.47	0.100	3.341	0.109
10d	3.654	0.105	2.295	0.075
10e	6.835	0.073	12.6	0.022
11a	15.45	0.064	8.67	0.096
11b	7.131	0.071	8.584	0.021
11c	18.12	0.131	14.01	0.084
11d	6.186	0.039	3.958	0.038
11e	11.67	0.053	6.391	0.004
11f	12.26	0.140	11.36	0.016
11g	4.852	0.100	3.734	0.097
11h	4.392	0.286	2.959	0.031
11i	6.736	0.022	7.244	0.086
11j	4.049	0.057	8.563	0.042
12	4.872	0.051	1.936	0.026

[a]The values represent the mean ± SE (standard error) of three independent experiments. Standard error (±SE) of the mean was also obtained for all experiments along with the IC_{50} values.

The analysis of the IC_{50} values against MCF-7 and MDA-MB-231 cell lines revealed that the synthesized compounds showed varying degrees of cytotoxicity. Among the tested compounds, the variability was relatively low, indicating the validity and reproducibility of the designed system. The IC_{50} values against the MCF-7 cell line ranged from 2.497 μM (compound **9**) to 20.47 μM (compound **10c**), whereas the IC_{50} values against the MDA-MB-231 cell line ranged from 1.936 μM (compound **12**) to 15.45 μM (compound **10b**). The results indicate that some compounds, such as compounds **9**, **10d**, **11g−h**, and **12**, exhibit potent antiproliferative activity against both breast cancer cell lines with IC_{50} values below 5 μM. Conversely, compounds **10c**, **11a**, **11c**, and **11e−f** showed lowered antiproliferative activity against MCF-7 and MDA-MB-231 cell lines. Compound **10c** exhibited the least potency, with an IC_{50} value of 20.47 μM against the MCF-7 cell line, while compound **10b** showed the lowest potency against the

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

MDA-MB-231 cell line, with an IC_{50} value of 15.45 μM. Interestingly, some compounds exhibited selective antiproliferative activity against one of the two cell lines. The results suggest that the synthesized compounds have the potential to be developed as anticancer agents, particularly against breast cancer cells.

2.5.2. EGFR Assay. Compounds **9**, **10d**, **11g**, and **12**, which demonstrated the most significant antiproliferative activity with IC_{50} values below 5 μM against both breast cancer cell lines, were selected to determine their EGFR inhibitory activity in vitro. However, it is essential to note that although compound **11h** exhibited potent antiproliferative activity, its IC_{50} value against MCF-7 cells had a relatively high standard error of 0.286. This may be due to its low purity, as it was the only compound tested with a purity of less than 90%. Hence, they were excluded from the EGFR assay. The tested compounds displayed potent EGFR inhibitory activity, with compound **9** exhibiting EGFR inhibitory activity of 2.53 nM. Compounds **10d** and **11g** possessed slightly lower EGFR inhibitory effects of 8.67 and 4.38 nM, respectively. Meanwhile, compound **12** possessed the lowest IC_{50} value of 19.1 nM (Table 3). For this series, the ML models showed promising ability to predict the activity of compounds with the tendency to slightly underestimate the activity of the compounds, indicating further room for future improvements.

Table 3. EGFR Inhibitory Activity of Compounds 9, 10d, 11g, and 12

compound	IC_{50} (nM)	pIC_{50}	predicted pIC_{50}
9	2.53	8.60	7.34
10d	8.67	8.06	7.00
11g	4.38	8.35	7.09
12	19.10	7.72	6.63
control (staurosporine)	51.60		

2.5.3. Analysis of the Cell Cycle Distribution. The capacity to trigger apoptosis in cancer cell lines is a crucial trait of many anticancer medications.[35] Additionally, agents that inhibit antiapoptotic activity have been shown to enhance the likelihood of overcoming resistance to EGFR inhibitors.[36] Accordingly, compound **9**, which exhibited the highest antiproliferative activity and EGFR inhibition, was selected for cell cycle and apoptosis analyses in this study. The results of the cell cycle analysis are demonstrated in Table 4.

The effect of compound **9** on the different phases of the cell cycle was examined in two breast cancer cell lines, MCF-7 and MDA-MB-231. The results showed that in MCF-7 cells, compound **9** exhibited cell growth arrest at the S phase of the cell cycle, increasing the percentage of cells in this phase from 15.32% in the control group to 65.12% at a concentration of $7.19 \times 10 \times 10^4$ cells/mL. Additionally, there was a decrease in the percentage of cells in the G0/G1 phase from 83.26% in the control group to 5.00% at the same concentration.

In MDA-MB-231 cells, compound **9** had a significant effect on the G0/G1 phase of the cell cycle, decreasing the percentage of cells in this phase from 59.01% in the control group to 22.26% at a concentration of $9.83 \times 10 \times 10^4$ cells/mL. There was also an increase in the percentage of cells in the S phase from 25.36% in the control group to 65.90%. These results suggest that compound **9** has a different effect on the cell cycle of MCF-7 and MDA-MB-231 cells, potentially due to differences in the genetic makeup of these cell lines. The increase in the percentage of cells in the S phase in both cell lines may indicate an apoptotic activity. The effect of inhibitors on the phases of the cell cycle compound **9**, compared with control MCF-7 and MDA-MB-231 cells, is illustrated in Figure 5.

2.5.4. Apoptosis Study. The results of the apoptosis analysis for compound **9** and control cells after 24 h are presented in Table 5.

In the MCF-7 cell line, the percentage of apoptosis induction by compound **9** was 62.18%, which was significantly higher than that of the control cells that showed only 96.72% live cells. The percentage of early apoptosis induction in compound **9**-treated MCF-7 cells was 12.78%, while it was only 3.06% in the control cells. The percentage of late apoptosis induction was 16.78% for compound **9**-treated MCF-7 cells compared to 0.04% in control cells. In contrast, the percentage of necrosis induction was lower in compound **9**-treated MCF-7 cells (8.27%) than in the control cells (0.18%). The apoptosis results of compound **9** over the two breast cancer cell lines are depicted as dotted plots in Figure 6.

Similarly, in the MDA-MB-231 cell line, compound **9** induced a higher percentage of apoptosis (67.72%) compared to control cells (90.90%). The percentage of early apoptosis induction in compound **9**-treated MDA-MB-231 cells was 10.12%, while it was 8.66% in control cells. The percentage of late apoptosis induction was 5.02% for compound **9**-treated MDA-MB-231 cells compared to 0.16% for the control cells. However, the percentage of necrosis induction was higher in compound **9**-treated MDA-MB-231 cells (12.31%) compared to that in control cells (1.63%). The results of the apoptosis analysis suggest that compound **9** can induce apoptosis in both MCF-7 and MDA-MB-231 cell lines. The induction of early and late apoptosis suggests that compound **9** may be acting through multiple mechanisms to induce cell death. However, the higher percentage of necrosis induction in compound **9**-treated MDA-MB-231 cells warrants further investigation to determine the underlying mechanism of cell death. Overall, these results suggest that compound **9** has potential as a therapeutic agent for breast cancer treatment. Figure 7 illustrates the images of intracellular fluorescence in MDA-MB-231 and MCF-7 cell lines following a 24 h treatment with compound **9**.

Table 4. Effect of Compound 9 on the Different Phases of the Cell Cycle

	MCF-7				MDA-MB-231			
	compound **9**		control		compound **9**		control	
	conc. (cells/mL)	percent (%)	conc. (cells/mL)	percent (%)	conc. (cells/mL)	percent (%)	conc. (cells/mL)	percent (%)
G0/G1 phase	$5.52 \times 10 \times 10^3$	5.00	$2.91 \times 10 \times 10^5$	83.26	$9.83 \times 10 \times 10^4$	22.26	$4.55 \times 10 \times 10^5$	59.01
S phase	$7.19 \times 10 \times 10^4$	65.12	$5.36 \times 10 \times 10^4$	15.32	$2.91 \times 10 \times 10^5$	65.90	$1.96 \times 10 \times 10^5$	25.36
G2/M phase	$3.15 \times 10 \times 10^4$	28.57	$4.87 \times 10 \times 10^3$	1.39	$3.40 \times 10 \times 10^4$	7.71	$8.02 \times 10 \times 10^4$	10.40

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

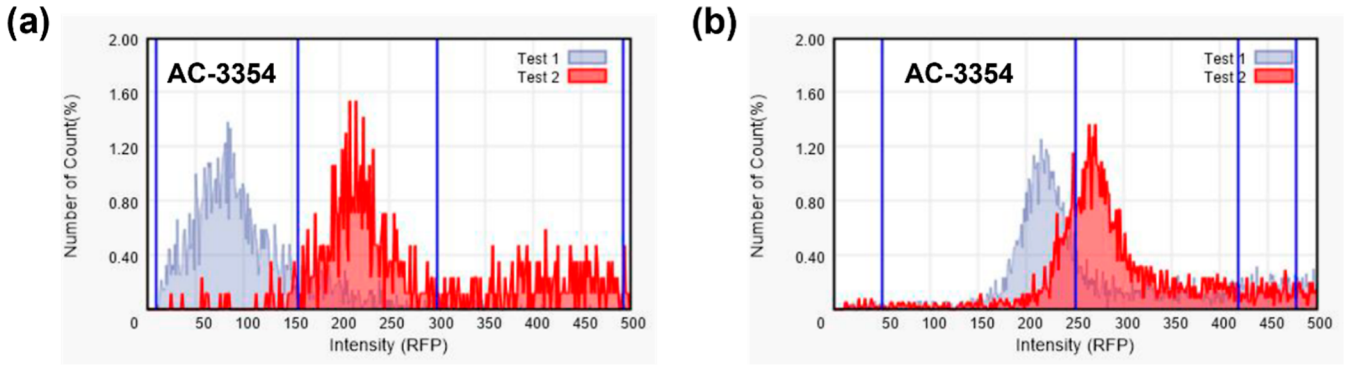


Figure 5. Effect of inhibitors on the phases of the cell cycle compound **9** (test 2), compared with control (test1) MCF-7 cells (a) and MDA-MB-231 (b).

Table 5. Comparison of Apoptosis Induction by Compound 9 and Control Cells after 24 h

cell line		apoptosis %			necrosis %
		live cells (L)	early (E.A)	late (L.A)	
MCF-7	compound **9**	62.18	12.78	16.78	8.27
	control	96.72	3.06	0.04	0.18
MDA-MB-231	compound **9**	67.72	10.12	5.02	12.31
	control	90.90	8.66	0.16	1.63

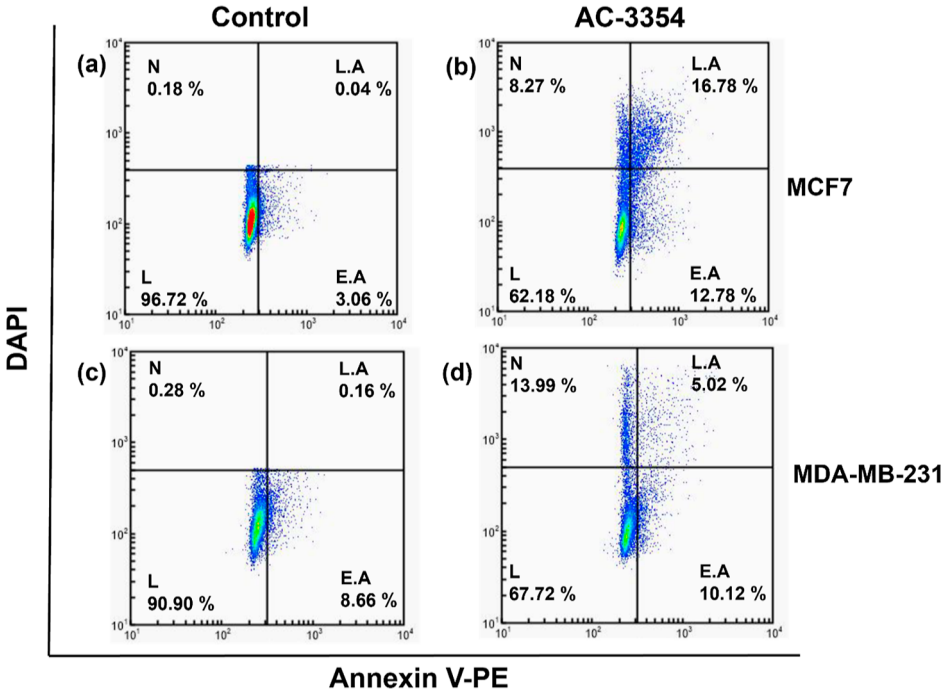


Figure 6. Apoptosis analysis of MCF-7 cells (a,b) and MDA-MB-231 (c,d) induced by compound **9** along with their controls represented in dot plots.

2.6. Molecular Docking. A molecular docking study was performed to investigate the interaction mechanisms of compounds **9**, **10d**, **11g**, and **12** with EGFR. The docking scores of each compound against EGFR are demonstrated in Table 6. The docking scores represent the binding energy of the compounds with EGFR, where lower scores indicate stronger binding.

All four compounds exhibited a strong binding pattern (Figure 8) with the EGFR binding site with all four compounds, establishing at least two hydrogen bonds with the amino acid residues of the binding cavity. The compounds established a hydrogen bond with Met769 and a carbon hydrogen interaction with the Gln767 amino acid residue of the binding site. This shared binding pattern among the four compounds validates the chosen docking approach and indicates a consistent mode of interaction with the EGFR binding site. Similarly, the compounds displayed a shared pattern of interaction with the Ala719 and Leu820 amino acid residues, which could further enhance the stability of the complexes. Compounds **9**, **10d**, and **11g** each established a hydrogen bond with the Thr766 amino acid residue. Compound **12** was the only compound to establish halogen

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

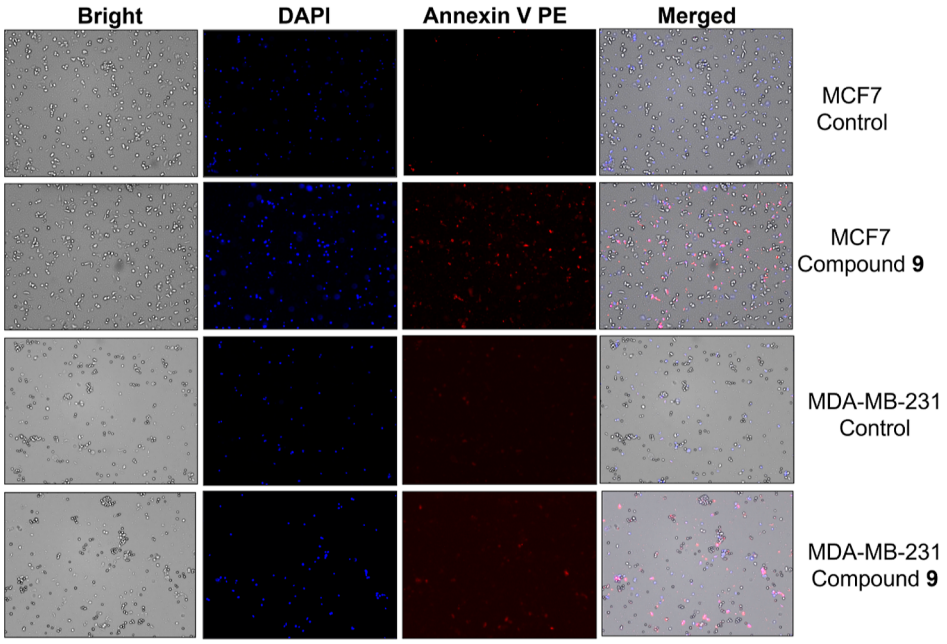


Figure 7. Cellular fluorescence images of MDA-MB-231 and MCF-7 cell lines treated with compound **9** for 24 h. Bright-field images, fluorescence images (DAPI:4′,6-diamidino-2-phenylindole, Annexin V PE), and merged images were assigned to the MDA-MB-231 and MCF-7 breast cancer cells with control (without any compound treatment) and compound **9**-treated, respectively, showing apoptotic cells.

Table 6. Docking Scores of Compounds 9, 10d, 11g, and 12 for the ATP and Allosteric EGFR Sites

compound	docking score (kcal/mol)	interactions
9	−8.70	hydrogen bond: Lys692, Thr766 and Met769
		$\pi-\pi$ stacking: Leu694, Ala719, Lys721, Met742, Leu764 and Leu820
		C−H bond: Gln767
10d	−7.77	hydrogen bond: Thr766, Met769 and Asp779
		$\pi-\pi$ stacking: Leu694, Ala719, Lys721, Met742, Leu764, Cys773 and Leu820
		C−H bond: Gln767
11g	−7.96	hydrogen bond: Thr766 and Met769
		$\pi-\pi$ stacking: Leu694, Ala719, Lys721, Met742, Leu764 and Leu820
		C−H bond: Gln767
12	−8.69	hydrogen bond: Lys721, Glu738 and Met769
		halogen bond: Asp831
		$\pi-\pi$ stacking: Phe699, Leu694, Val702, Ala719 and Leu820
		π−s interaction: Met742
		C−H interaction: Gln767

and π−s interactions with Asp831 and Met742 amino acid residues, respectively. This unique interaction pattern suggests that compound **12** may have a different mode of action compared to the other compounds.

In conclusion, the molecular docking study revealed that all four compounds had potential inhibitory activity on EGFR, with compound **9** exhibiting the strongest binding affinity. The interactions formed between the compounds and EGFR, including hydrogen bonding, $\pi-\pi$ stacking, halogen bonding, and C−H interaction, could play a vital role in stabilizing the complexes and enhancing the inhibitory activity of the compounds.

3. METHODOLOGY

3.1. Machine Learning Models of EGFR and Bioactivity Prediction Application. *3.1.1. Data Set Preparation and Descriptor Generation.* All of the ML construction processes were performed using the python programming language in the Jupyter notebook software. A data set of inhibitors against human EGFR erbB1 (Target ID CHEMBL203) was compiled from the ChEMBL, which comprised a total number of 13914 compounds.[37] SMILES notations and corresponding IC_{50} values of the compounds were curated and extracted into a CSV sheet.[29] The initial data set was further curated by discarding 4175 compounds which had no reported IC_{50} values, leaving the final data set to consist of 9019 compounds. The compounds were then further classified into active, inactive, and intermediate inhibitors based on their activity. Active compounds were characterized as the compounds exhibiting IC_{50} of less than or equal to 1 μM, while compounds were labeled as inactive when their activity was more than or equal to 10 μM. Subsequently, the IC_{50} values were converted to pIC_{50} and descriptors were calculated employing the RDKit AllChem software.[38]

3.1.2. QSAR Model Generation. The first step was reading in the data from the curated CSV file using pandas.[39] Next, the SMILES strings were converted into RDKit molecules, and a set of molecular descriptors were calculated using the RDKit library.[40] These descriptors include molecular weight, number of rotatable bonds, TPSA, and Morgan fingerprints. Eleven ML models were then constructed. The models included Random Forest Regression, Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, K-NN Regression, SVM Regression, MLP Regression, XGBoost, LightGBM, and CatBoost. The parameters for the models were set as follows: for Random Forest Regression, *n*_estimators was set to 100 and the random state was set to 42; for MLP Regression, the hidden_layer_sizes was set to (100,50), the activation function was set to "relu", the solver was set to "adam", and the

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

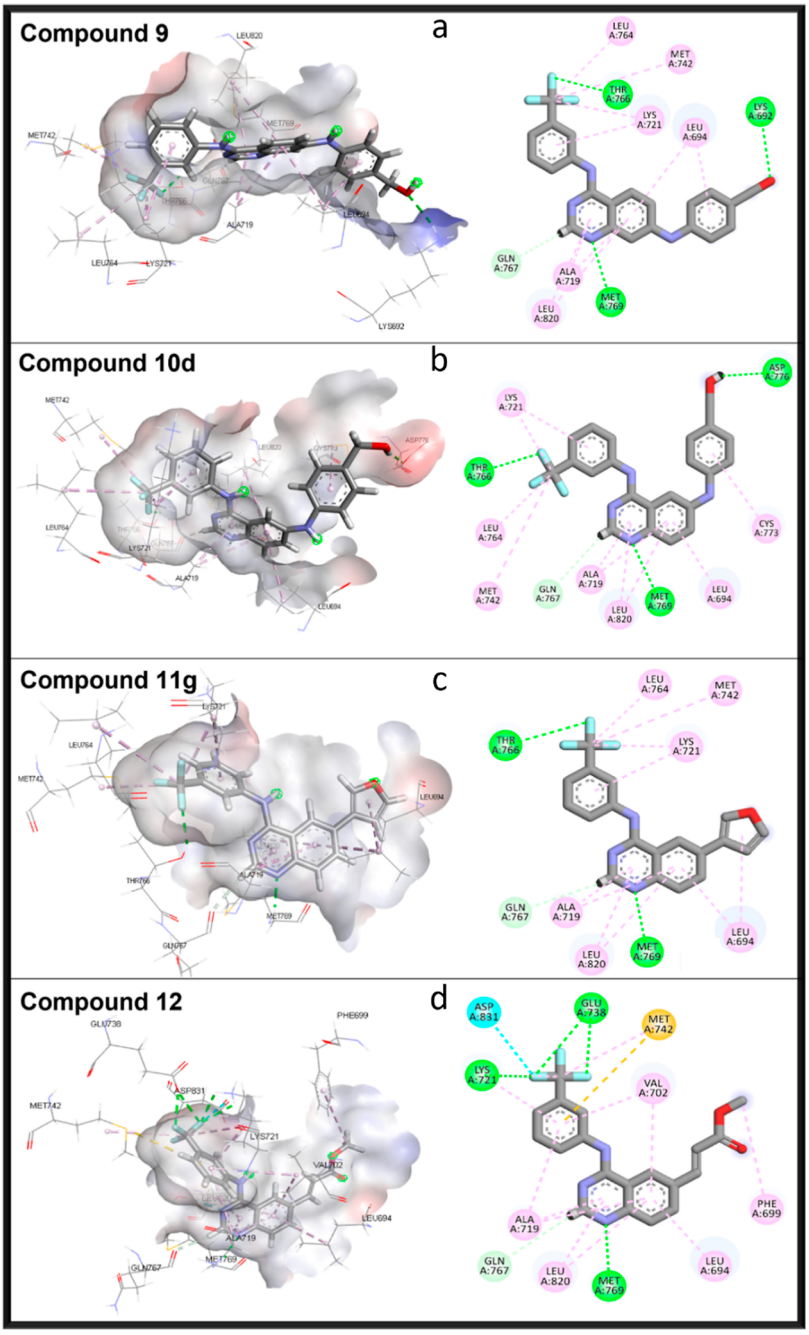


Figure 8. (a) Docked complexes of compounds **9** (docking score: −8.70 kcal/mol), (b) **10d** (docking score: −7.77 kcal/mol), (c) **11g** (docking score: −7.96 kcal/mol), and (d) **12** (docking score: −8.69 kcal/mol) with the EGFR binding site (PDB: 1M17).

maximum number of iterations was set to 500; for XGBoost, the objective was set to "reg:squarederror" and the random state was set to 42; for LightGBM, *n*_estimators was set to 100 and the random state was set to 42; and for CatBoost, the number of iterations was set to 100, the learning rate was set to 0.1, and the random seed was set to 42.

k-Fold cross-validation was then performed with five splits on the models to assess their performance.[41,42] The data were randomly divided into five sets, with four sets being used for training and one set being used for testing in each fold. The *R*-squared score was used as the performance metric, with higher scores indicating better performance. The mean *R*-squared score and standard deviation were calculated for each model.

3.1.3. Web Application Development. A Python code for a web app was built using Streamlit and RDKit libraries that predicts the biological activity of a molecule based on its structure.[40,43] The app first allows the user to upload a CSV file with SMILES strings for multiple molecules or enter a single SMILES string. The SMILES strings are then converted to molecular descriptors by using RDKit functions. These descriptors are then used to make predictions using a pretrained model, which is loaded from a saved pickle file. The app displays the predicted biological activity (pIC_{50}) of the molecules in a sorted DataFrame format. The Supporting Information includes both the application and the best ML model file (.pkl format).

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

3.1.4. Permutation Test. A permutation test was employed to assess the best model's performance between the overall data set and the quinazolin-4-amine derivatives' subset. The MSE metric was used to assess the performance of the ML model. The test involved loading the trained ML model via the joblib.load() function, which was followed by loading of the two data sets (full data set and the quinazoline only data set). The descriptor columns were then extracted from the data sets by removing the target variable (pIC_{50}). The expected number of features was calculated and compared to the actual number of features in the loaded descriptor data, ensuring consistency. Next, categorical variables in the descriptor data were encoded using the Label Encoder function from the sklearn.preprocessing module. Any missing values in the data sets were replaced with 0. The whole data set was split into training and testing sets using train_test_split() from sklearn.model_selection. A new Random Forest Regressor model was created with the criterion set to squared_error. The model was then fitted to the training data. The model's performance on the whole data set was evaluated by predicting the target variable for the testing features and calculating the mean squared error using mean_squared_error() from sklearn.metrics. The permutation test was performed by shuffling the target variable for the quinazoline data set, and for each permutation, predictions were made using the testing features. The mean square error was calculated for each permutation and stored in a list. The *p*-value was computed by counting the number of permuted mean squared errors that were greater than or equal to the mean squared error of the whole data set and dividing by the total number of permutations plus one. The calculated *p*-value indicates the significance of the observed difference in model performance between the full data set and the quinazoline data set.[44] The results can be used to assess the bias or generalization ability of the model toward the quinazolin-4-amine derivatives' subset compared to the overall data set. The code used to perform the permutation test and the curated databases is available in the Supporting Information.

3.2. Chemistry. The experiments were carried out by using the purchased reagents and solvents without further purification. The 1H NMR spectra were captured by using a Varian 400 MHz spectrometer (Varian Medical Systems, Inc., Palo Alto, CA, USA), with chemical shifts being recorded in parts per million (ppm) and coupling constants in Hz. HR electrospray ionization (ESI) MS data were collected using either a G2 QTOF mass spectrometer or a JMS-700 mass spectrometer (both from Jeol, Japan). Thin-layer chromatography was employed to monitor the reactions on 0.25 mm silica plates (E. Merck; silica gel 60 F254). Using an HPLC system from Waters Corp. with a UV detector set at 254 nm, reversed-phase HPLC was used to assess the purity of the products. Both A and B, which were mobile phases, contained 0.05% TFA in water. HPLC was used with a YMC Hydrosphere C18 (HS-302) column that was 4.6 mm in diameter and 150 mm in length and had a flow rate of 1.0 mL/min. The column had a 5 M particle size and a 12 nm pore size. Using either a gradient of 75% B or 100% B in 30 min, the resulting compounds' purity was determined. To measure the melting points, a Fisherbrand digital melting point instrument was employed.

3.2.1. General Procedure for the Synthesis of Compound ***2****.* A mixture of 2-amino-4-nitrobenzonitrile (**1**) (18.4 mmol) and dimethylformamide dimethyl acetal (40 mL) was stirred at reflux for 1.5 h with 10 mL of DCM to increase the solubility. Then, it was washed with water and ether to yield the product. The crude product was used in the next step directly without further purification.

3.2.2. General Procedure for the Synthesis of Compound ***3****.* A mixture of (*E*)-*N*′-(2-cyano-5-nitrophenyl)-*N*,*N*-dimethylformimidamide (**2**) (4.6 mmol) and 3-(trifluoromethyl)-aniline (4.6 mmol) in glacial acetic acid (20.0 mL) was refluxed for 2 h at 90 °C. The acetic acid was evaporated and extracted with ethyl acetate (EA). The extract was purified with MPLC to yield the product at 20% of EA/hexane.

3.2.3. General Procedure for the Synthesis of Compound ***4****.* A mixture of 7-nitro-*N*-(3-(trifluoromethyl)phenyl)-quinazolin-4-amine (**3**) (1.5 mmol) and iron (10.5 mmol) was suspended in aqueous ethanol (60.0 mL, 70% v/v) containing acetic acid (10.0 mL). The mixture was refluxed for 2 h. After 2 h, the mixture was cooled and basified with concentrated ammonia solution and stirred at rt for 2 h. The insoluble precipitate was removed by filtering. The filtrate was then evaporated under reduced pressure. The crude residue was purified by silica gel column chromatography (methylene chloride/methanol; 50:1 to 10:1 v/v) to yield the desired product.

3.2.4. General Procedure for the Synthesis of Compound ***5****.* *N*4-(3-(Trifluoromethyl)phenyl)quinazoline-4,7-diamine (**4**) (0.66 mmol) was dissolved in a mixture of acetonitrile, NaOH (1:1 ratio), and chloroacetyl chloride (1.32 mmol) for 3 h. The mixture was then extracted with EA and purified with MPLC to get the desired product.

3.2.5. General Procedure for the Synthesis of Compounds ***7a−b****.* A mixture of 2-amino-5-bromobenzonitrile (**6a**) or 2-amino-4-bromobenzonitrile (**6b**) (1.52 mmol) and dimethylformamide dimethyl acetal (10 mL) was stirred under reflux for 1.5 h. The mixture was cooled to room temperature and refrigerated. The solid was filtered, washed with several portions of ether, and dried to yield the desired compound (90%), which was used in the next step without further purification.

3.2.6. General Procedure for the Synthesis of Compounds ***8a−b****.* The mixture of (*E*)-*N*′-(5-bromo-2-cyanophenyl)-*N*,*N*-dimethylformimidamide (**7a**) or (*E*)-*N*′-(4-bromo-2-cyanophenyl)-*N*,*N*-dimethylformimidamide (**7b**) (2.6 mmol) and 3-(trifluoromethyl) aniline (2.6 mmol) in glacial acetic acid (15.0 mL) was refluxed for 2 h. The acetic acid was evaporated, and the solid was washed with water and diethyl ether and dried to afford the title compounds, which were used in the next step without further purification.

3.2.7. General Procedure for the Synthesis of Compounds ***9, 10a−e****.* The appropriate amine (4 mmol), *t*-butyl XPhos (0.15 mmol), pd_2dba_3 (0.15 mmol), and potassium carbonate (3 mmol) were added to a solution of **8a** or **8b** (1 mmol) and toluene (20 mL) in a sealed tube. The mixture was purged for 15 min with a N_2 gas and sealed. The sealed mixture was stirred at 120 °C for 14 h. The reaction mixture was subsequently poured into ice-cold water (25 mL) and extracted with EA (200 mL). The organic layer was dried over anhydrous $MgSO_4$ and purified by silica gel chromatography to yield the desired product.

3.2.8. General Procedure for the Synthesis of Compounds ***11a−j****.* **11a−j** were synthesized by Suzuki coupling with **8a** and boronic acid derivatives. A mixture of 15 mL of dioxane and 3 mL of water in a seal tube was purged for 15 min with a N_2 balloon. Compound **8a** (1 mmol), boronic acid derivative (1.5 mmol), K_2CO_3 (4 mmol), and [1,1′-bis-

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

(diphenylphosphino)ferrocene]-dichloro palladium(II) (0.05 mmol) were dissolved in the prepared mixture. The reaction mixture was heated in a sealed tube at 105 °C for 4 h and then cooled to room temperature and extracted with EA. Then, the organic layer was dried with $MgSO_4$ and evaporated. Column chromatography was then carried out to purify the mixture, yielding the pure product.

3.2.9. General Procedure for the Synthesis of Compound ***12.*** A suspension of **8a** (0.41 mmol) in DMF (20 mL) was stirred with tri-*o*-tolylphosphine (0.41 mmol), palladium acetate(II) (0.12 mmol), triethylamine (1.2 mmol), and methyl acrylate (4.1 mmol). The reaction mixture was then refluxed for 12 h. The resulting solution was poured into ice-cold water (100 mL) and extracted with EA (300 mL). The organic layer was dried over anhydrous $MgSO_4$, concentrated, and purified by silica gel chromatography to afford the titled compound.

*3.2.9.1. 7-Nitro-N-(3-(trifluoromethyl)phenyl)quinazolin-4-amine (**3**).* Yellow powder, yield: 52.03%, ^{1}H NMR (400 MHz, $CDCl_3$-d_1): δ 8.92 (s, 1H), 8.81 (d, J = 2.2 Hz, 1H), 8.37 (dd, J = 9.1, 2.2 Hz, 1H), 8.10 (d, J = 8.8 Hz, 2H), 8.01 (d, J = 8.3 Hz, 1H), 7.64 (s, 1H), 7.59 (t, J = 7.9 Hz, 1H), 7.49 (d, J = 7.8 Hz, 1H); ^{13}C NMR (101 MHz, $CDCl_3$-d_1): δ 157.10, 157.09, 156.69, 156.55, 150.51, 150.42, 138.04, 125.12, 125.03, 124.92, 122.28, 122.17, 120.36, 118.64, 118.13.

*3.2.9.2. N^4-(3-(Trifluoromethyl)phenyl)quinazoline-4,7-diamine (**4**).* Off-white solid, yield: 85%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.54 (s, 1H), 8.40 (s, 1H), 8.30 (s, 1H), 8.19 (d, J = 8.9 Hz, 2H), 7.56 (t, J = 8.1 Hz, 1H), 7.36 (d, J = 7.8 Hz, 1H), 6.92 (dd, J = 9.0, 2.1 Hz, 1H), 6.72 (d, J = 2.1 Hz, 1H), 6.07 (s, 2H).

*3.2.9.3. 2-Chloro-N-(4-((3-(trifluoromethyl)phenyl)-amino)quinazolin-7-yl)acetamide (**5**).* Yellow powder, yield: 16%, mp: 324.1 °C, HPLC purity: 6 min, 100%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.77 (s, 1H), 9.96 (s, 1H), 8.63 (s, 1H), 8.53 (d, J = 9.0 Hz, 1H), 8.34 (s, 1H), 8.24 (d, J = 7.9 Hz, 1H), 8.16 (d, J = 2.0 Hz, 1H), 7.77 (dd, J = 9.0, 2.1 Hz, 1H), 7.63 (t, J = 7.9 Hz, 1H), 7.46 (d, J = 7.7 Hz, 1H), 4.36 (s, 2H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 165.91, 157.59, 155.29, 151.21, 142.97, 140.62, 130.09, 129.81, 125.82, 124.48, 120.00, 119.46, 118.33, 118.29, 115.65, 111.77, 44.09; HRMS (ESI) m/z calcd for $C_{17}H_{12}ClF_3N_4O$: $[M + H]^+$, 381.0724; found, 381.0718.

*3.2.9.4. 6-Bromo-N-(3-(trifluoromethyl)phenyl)-quinazolin-4-amine (**8a**).* Yellow solid, yield: 43%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.93 (s, 1H), 8.80 (s, 1H), 8.64 (s, 1H), 8.17 (d, J = 6.7 Hz, 1H), 7.98 (d, J = 8.8 Hz, 1H), 7.85−7.77 (m, 1H), 7.73 (d, J = 8.8 Hz, 1H), 7.43 (t, J = 9.1 Hz, 1H).

*3.2.9.5. 7-Bromo-N-(3-(trifluoromethyl)phenyl)-quinazolin-4-amine (**8b**).* Yellow solid, yield: 86%. It was used in the crude form and was not purified.

*3.2.9.6. (4-((4-((3-(Trifluoromethyl)phenyl)amino)-quinazolin-7-yl)amino)phenyl)methanol (**9**).* Yellow powder, yield: 22%, mp: 234 °C, HPLC purity: 6 min, 97.47%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.74 (s, 1H), 8.86 (s, 1H), 8.49 (s, 1H), 8.39−8.32 (m, 2H), 8.23 (d, J = 8.0 Hz, 1H), 7.60 (t, J = 7.9 Hz, 1H), 7.41 (d, J = 7.3 Hz, 1H), 7.33 (d, J = 8.2 Hz, 2H), 7.24 (dd, J = 19.8, 11.4 Hz, 4H), 5.12 (t, J = 5.5 Hz, 1H), 4.48 (d, J = 5.7 Hz, 2H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 157.26, 154.99, 152.31, 149.12, 141.03, 140.12, 139.83, 137.12, 129.97, 129.44, 128.20, 125.41, 124.65, 120.19, 119.45, 118.07, 117.92, 108.19, 106.77, 63.12. HR MS (ESI) m/z calcd for $C_{22}H_{17}F_3N_4O$: $[M + H]^+$, 411.1427; found, 411.1423.

*3.2.9.7. N^6-(4-Methoxypyridin-2-yl)-N4-(3-(trifluoromethyl)phenyl)quinazoline-4,6-diamine (**10a**).* Yellow powder, yield: 12.5%, mp: 290.8 °C, HPLC purity: 4 min, 97.95%, ^{1}H NMR (400 MHz, DMSO-d_6) 1H NMR (400 MHz, dmso): δ 9.87 (s, 1H), 9.33 (s, 1H), 8.67 (s, 1H), 8.53 (s, 1H), 8.29 (s, 1H), 8.19 (d, J = 8.8 Hz, 1H), 8.08−8.04 (m, 1H), 8.00−7.96 (m, 1H), 7.75 (d, J = 9.1 Hz, 1H), 7.61 (t, J = 8.0 Hz, 1H), 7.43 (d, J = 7.5 Hz, 1H), 6.49 (d, J = 2.6 Hz, 2H), 3.81 (s, 3H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 166.78, 157.55, 157.25, 152.14, 149.16, 145.69, 141.03, 140.45, 129.95, 129.76, 128.70, 127.67, 125.89, 119.66, 118.31, 118.27, 116.28, 109.04, 104.32, 94.53, 55.42. HRMS (ESI) m/z calcd for $C_{21}H_{16}F_3N_5O$: $[M + H]^+$, 412.1380; found, 412.1375.

*3.2.9.8. N^6-Phenyl-N4-(3-(trifluoromethyl)phenyl)-quinazoline-4,6-diamine (**10b**).* Yellowish-brown powder, yield: 46%, mp: 86 °C, HPLC purity: 9 min, 95.71%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 1H NMR (400 MHz, dmso): δ 9.81 (s, 1H), 8.61 (s, 1H), 8.49 (s, 1H), 8.29−8.05 (m, 3H), 7.74 (d, J = 9.0 Hz, 1H), 7.61 (t, J = 7.7 Hz, 2H), 7.43 (d, J = 7.7 Hz, 1H), 7.37−7.17 (m, 4H), 6.92 (t, J = 7.2 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 156.81, 154.87, 151.63, 143.23, 142.37, 140.92, 129.92, 129.80, 129.47, 127.02, 126.05, 120.94, 119.76, 119.74, 118.46, 118.41, 117.38, 116.61, 105.23. HRMS (ESI) m/z calcd for $C_{21}H_{15}F_3N_4$, 381.1322: $[M + H]^+$; found, 381.1313.

*3.2.9.9. 3-((4-((3-(Trifluoromethyl)phenyl)amino)-quinazolin-6-yl)amino)phenol (**10c**).* Yellow powder, yield: 43%, mp: 245 °C, HPLC purity: 7 min, 100%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.82 (s, 1H), 9.27 (s, 1H), 8.49 (s, 2H), 8.25 (s, 1H), 8.20 (d, J = 8.0 Hz, 1H), 8.11 (s, 1H), 7.72 (d, J = 9.0 Hz, 1H), 7.60 (dd, J = 13.1, 7.5 Hz, 2H), 7.42 (d, J = 7.3 Hz, 1H), 7.07 (t, J = 8.3 Hz, 1H), 6.66 (s, 2H), 6.32 (d, J = 7.7 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 166.54, 156.75, 151.54, 145.04, 142.75, 141.74, 140.93, 135.29, 129.94, 129.53, 128.36, 126.68, 126.68, 126.05, 118.43, 118.38, 117.57, 117.35, 116.69, 104.46, 63.20. HRMS (ESI) m/z calcd for $C_{21}H_{15}F_3N_4O$: $[M + H]^+$, 397.1271; found, 397.1266.

*3.2.9.10. (4-((4-((3-(Trifluoromethyl)phenyl)amino)-quinazolin-6-yl)amino)phenyl)methanol (**10d**).* Yellow powder, yield: 24%, mp: 143 °C, HPLC purity: 6 min, 100%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.80 (s, 1H), 8.57 (s, 1H), 8.48 (s, 1H), 8.24 (s, 1H), 8.17 (d, J = 7.9 Hz, 1H), 8.05 (s, 1H), 7.72 (d, J = 9.0 Hz, 1H), 7.59 (d, J = 7.2 Hz, 2H), 7.42 (d, J = 8.4 Hz, 1H), 7.23 (dd, J = 22.0, 8.4 Hz, 4H), 5.07 (s, 1H), 4.44 (d, J = 5.4 Hz, 2H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 156.73, 142.71, 141.75, 140.98, 135.29, 129.69, 129.31, 128.57, 128.14, 127.89, 126.99, 125.94, 125.80, 123.33, 121.54, 120.44, 118.35, 117.73, 116.70, 116.69, 63.20. HRMS (ESI) m/z calcd for $C_{22}H_{17}F_3N_4O$: $[M + H]^+$, 411.1427; found, 411.1425.

*3.2.9.11. N^6-Benzyl-N^4-(3-(trifluoromethyl)phenyl)-quinazoline-4,6-diamine (**10e**).* Yellow powder, yield: 23.4%, mp: 186.8 °C, HPLC purity: 98.17%, 9.127 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.52 (s, 1H), 8.40 (s, 1H), 8.27 (s, 1H), 8.22 (d, J = 8.4 Hz, 1H), 7.60 (dd, J = 19.1, 8.5 Hz, 2H), 7.50−7.33 (m, 7H), 7.27 (t, J = 7.3 Hz, 1H), 6.75 (t, J = 5.8 Hz, 1H), 4.47 (d, J = 5.7 Hz, 2H); ^{13}C NMR (100 MHz, DMSO-d_6): δ156.13, 150.05, 147.92, 143.53, 141.11, 139.69, 129.95, 129.77, 129.46, 129.10, 128.84, 128.23, 127.44, 126.09, 125.72, 124.51, 123.38, 119.48, 118.06, 116.94, 97.43,

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

47.25. HRMS (ESI) m/z calcd for $C_{22}H_{17}F_3N_4$: $[M + H]^+$, 395.1484; found, 395.1491.

*3.2.9.12. 4-(4-((3-(Trifluoromethyl)phenyl)amino)-quinazolin-6-yl)benzaldehyde (**11a**).* Off-white solid, yield: 32.5%, mp: 207.1 °C, HPLC purity: 10.2 min, 98.27%, ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.20 (s, 1H), 10.11 (s, 1H), 8.98 (d, J = 1.7 Hz, 1H), 8.71 (s, 1H), 8.32 (dt, J = 14.9, 7.5 Hz, 3H), 8.13 (dd, J = 21.0, 8.4 Hz, 4H), 7.94 (d, J = 8.7 Hz, 1H), 7.68 (t, J = 7.9 Hz, 1H), 7.51 (d, J = 7.9 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 193.15, 158.14, 155.14, 150.10, 145.12, 140.29, 137.22, 135.84, 132.48, 130.62, 130.11, 129.15, 128.24, 126.09, 123.29, 121.89, 120.36, 120.33, 118.59, 115.73. HRMS (ESI) m/z calcd for $C_{22}H_{14}F_3N_3O$: $[M + H]^+$, 394.1167; found, 394.1163.

*3.2.9.13. 2-(4-((3-(Trifluoromethyl)phenyl)amino)-quinazolin-6-yl)phenol (**11b**).* Off-white powder, yield: 7.3%, mp: 186.9 °C, HPLC purity: 100%, 9.050 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.98 (s, 1H), 8.71 (s, 1H), 8.65 (s, 1H), 8.37 (s, 1H), 8.30 (s, 1H), 8.27 (d, J = 8.1 Hz, 1H), 8.18 (dd, J = 8.7, 1.7 Hz, 1H), 7.91−7.80 (m, 3H), 7.66 (dd, J = 17.8, 9.8 Hz, 2H), 7.50 (d, J = 7.8 Hz, 1H), 7.18 (d, J = 1.0 Hz, 1H); ^{13}C NMR (101 MHz, CD_3OD-d_4): δ 159.74, 155.12, 149.64, 145.68, 141.26, 141.16, 133.22, 132.81, 132.29, 131.97, 130.61, 128.79, 127.06, 127.00, 121.73, 121.70, 120.24, 120.20, 119.41, 117.05, 109.65. HRMS (ESI) m/z calcd for: $C_{21}H_{14}F_3N_3O$: $[M + H]^+$, 382.1167; found, 382.1159.

*3.2.9.14. 3-(4-((3-(Trifluoromethyl)phenyl)amino)-quinazolin-6-yl)phenol (**11c**).* Off-white powder, yield: 48.5%, mp: 221.9 °C, HPLC purity: 98.65%, 8.37 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.14 (s, 1H), 9.64 (s, 1H), 8.80 (s, 1H), 8.65 (s, 1H), 8.32−8.24 (m, 2H), 8.13 (dd, J = 8.0, 2.0 Hz, 1H), 7.87 (d, J = 8.0 Hz, 1H), 7.64 (t, J = 8.0 Hz, 1H), 7.47 (d, J = 8.0 Hz, 1H), 7.37−7.21 (m, 3H), 6.85 (dd, J = 2.0, 4.0 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 158.36, 158.12, 154.59, 149.54, 141.01, 140.53, 138.99, 132.47, 130.48, 130.07, 128.89, 126.12, 126.11, 126.02, 120.76, 118.61, 118.57, 118.43, 115.74, 115.40, 114.53. HRMS (ESI) m/z calcd for: $C_{21}H_{14}F_3N_3O$: $[M + H]^+$, 382.1167; found, 382.1156.

*3.2.9.15. 4-(4-((3-(Trifluoromethyl)phenyl)amino)-quinazolin-6-yl)phenol (**11d**).* Off-white powder, yield: 9.66%, mp: 282 °C, HPLC purity: 96.81%, 7.85 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ10.06 (s, 1H), 9.69 (s, 1H), 8.73 (s, 1H), 8.63 (s, 1H), 8.28 (d, J = 10.1 Hz, 2H), 8.14 (d, J = 8.9 Hz, 1H), 7.83 (d, J = 8.8 Hz, 1H), 7.72 (d, J = 8.5 Hz, 2H), 7.64 (t, J = 8.7 Hz, 1H), 7.46 (d, J = 7.3 Hz, 1H), 6.93 (d, J = 8.4 Hz, 2H); ^{13}C NMR (100 MHz, CD_3OD-d_4): δ 159.91, 159.10, 154.99, 149.33, 141.59, 141.25, 133.63, 132.25, 130.57, 129.66, 128.59, 127.04, 121.67, 121.63, 120.29, 120.25, 120.14, 117.03, 116.97. HRMS (ESI) m/z calcd for $C_{21}H_{14}F_3N_3O$: $[M + H]^+$, 382.1167; found, 382.1169.

*3.2.9.16. 6-(4-Methoxyphenyl)-N-(3-(trifluoromethyl)-phenyl)quinazolin-4-amine (**11e**).* White powder, yield: 46.5%, mp: 214 °C, HPLC purity: 97.44%, 10.658 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.09 (s, 1H), 8.77 (s, 1H), 8.63 (s, 1H), 8.31−8.24 (m, 2H), 8.17 (d, J = 8.0 Hz, 1H), 7.84 (dd, J = 8.0, 4.0 Hz, 3H), 7.64 (t, J = 8.0 Hz, 1H), 7.46 (d, J = 8.0 Hz, 1H), 7.11 (d, J = 8.0 Hz, 2H), 3.82 (s, 3H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 159.81, 158.01, 154.34, 149.15, 138.48, 132.19, 131.86, 130.09, 129.53, 128.79, 126.11, 119.87, 118.56, 115.87, 114.95, 55.74. HRMS (ESI) m/z calcd for: $C_{22}H_{16}F_3N_3O$: $[M + H]^+$, 396.1324; found, 396.1335.

*3.2.9.17. 6-(4-Chlorophenyl)-N-(3-(trifluoromethyl)-phenyl)quinazolin-4-amine (**11f**).* Off-white powder, yield: 34.6%, mp: 228.8 °C, HPLC purity: 96.60%, 13.05 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.17 (s, 1H), 8.87 (d, J = 1.7 Hz, 1H), 8.68 (s, 1H), 8.32−8.20 (m, 3H), 7.92 (dd, J = 16.6, 8.7 Hz, 3H), 7.66 (dd, J = 12.8, 8.3 Hz, 3H), 7.49 (d, J = 7.8 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 158.13, 154.84, 149.69, 140.62, 138.33, 137.25, 133.36, 132.22, 130.09, 129.84, 129.45, 129.32, 129.00, 126.17, 121.04, 120.19, 118.66, 118.62, 115.86. HRMS (ESI) m/z calcd for: $C_{21}H_{13}ClF_3N_3$: $[M + H]^+$, 400.0828; found, 400.0820.

*3.2.9.18. 6-(Furan-3-yl)-N-(3-(trifluoromethyl)phenyl)-quinazolin-4-amine (**11g**).* Off-white powder, yield: 13.02%, mp: 188.7 °C, HPLC purity: 98.36%, 9.3 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 9.96 (s, 1H), 8.69 (d, J = 1.5 Hz, 1H), 8.63 (s, 1H), 8.35 (s, 1H), 8.28 (s, 1H), 8.25 (d, J = 8.1 Hz, 1H), 8.16 (dd, J = 8.7, 1.7 Hz, 1H), 7.85 (t, J = 1.7 Hz, 1H), 7.81 (d, J = 8.7 Hz, 1H), 7.65 (t, J = 8.0 Hz, 1H), 7.48 (d, J = 7.7 Hz, 1H), 7.16 (dd, J = 1.7, 0.7 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 157.72, 154.26, 149.29, 145.11, 140.63, 140.36, 131.59, 130.80, 130.07, 129.89, 129.58, 128.83, 125.99, 125.78, 120.17, 118.98, 118.47, 115.74, 109.24. HRMS (ESI) m/z calcd for $C_{19}H_{12}F_3N_3O$: $[M + H]^+$, 356.1011; found, 356.1007.

*3.2.9.19. 6-(Thiophen-3-yl)-N-(3-(trifluoromethyl)phenyl)-quinazolin-4-amine (**11h**).* Off-white powder, yield: 22.3%, mp: 212.5 °C, HPLC purity: 81.70%, 10.097 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ10.03 (s, 1H), 8.83 (s, 1H), 8.64 (s, 1H), 8.26 (dd, J = 8.7, 7.0 Hz, 3H), 8.09 (dd, J = 2.8, 1.3 Hz, 1H), 7.85−7.73 (m, 3H), 7.65 (t, J = 8.0 Hz, 1H), 7.47 (d, J = 7.8 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 157.98, 154.42, 149.34, 140.99, 140.48, 133.80, 132.07, 130.11, 128.88, 127.97, 126.92, 126.13, 122.67, 120.26, 120.23, 119.71, 118.60, 118.56, 115.79. HRMS (ESI) m/z calcd for: $C_{19}H_{12}F_3N_3S$: $[M + H]^+$, 372.0782; found, 372.0772.

*3.2.9.20. 6-(Benzo[d][1,3]dioxol-5-yl)-N-(3-(trifluoromethyl)phenyl)quinazolin-4-amine (**11i**).* Off-white powder, yield: 36%, mp: 200.8 °C, HPLC purity: 100%, 10.65 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.05 (s, 1H), 8.74 (s, 1H), 8.64 (s, 1H), 8.32−8.22 (m, 2H), 8.16 (d, J = 8.7 Hz, 1H), 7.83 (d, J = 8.7 Hz, 1H), 7.64 (t, J = 7.9 Hz, 1H), 7.51−7.43 (m, 2H), 7.38 (d, J = 8.0 Hz, 1H), 7.09 (d, J = 8.0 Hz, 1H), 6.10 (s, 2H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 158.00, 154.45, 149.28, 148.61, 147.78, 140.51, 138.44, 133.69, 132.31, 130.09, 128.82, 126.14, 121.38, 120.17, 118.62, 115.74, 109.24, 107.91, 101.81. HRMS (ESI) m/z calcd for: $C_{22}H_{14}F_3N_3O_2$, 410.1116: $[M + H]^+$; found, 410.1130.

*3.2.9.21. 6-(4-(Morpholinomethyl)phenyl)-N-(3-(trifluoromethyl)phenyl)quinazolin-4-amine (**11j**).* Off-white powder, yield: 19.8%, mp: 232.7 °C, HPLC purity: 91.5%, 3.3 min; ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.12 (s, 1H), 8.85 (s, 1H), 8.68 (s, 1H), 8.30 (d, J = 11.8 Hz, 2H), 8.22 (d, J = 8.6 Hz, 1H), 7.88 (dd, J = 17.9, 8.4 Hz, 3H), 7.66 (t, J = 8.1 Hz, 1H), 7.50 (d, J = 7.5 Hz, 3H), 3.60 (s, 4H), 3.55 (s, 2H), 2.40 (s, 4H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 158.07, 154.57, 149.48, 140.53, 138.64, 138.32, 138.20, 132.39, 130.54, 130.04, 129.53, 128.93, 127.48, 126.03, 120.72, 118.54, 118.49, 115.79, 115.30, 66.65, 62.46, 53.61. HRMS (ESI) m/z calcd for: $C_{26}H_{23}F_3N_4O$: $[M + H]^+$, 465.1902; found, 465.1902.

*3.2.9.22. (E)-Methyl 3-(4-((3-(trifluoromethyl)phenyl)-amino)quinazolin-6-yl)acrylate (**12**).* Yellow powder, yield: 13%, mp: 186.7 °C, HPLC purity: 10 min, 97.45%, ^{1}H NMR (400 MHz, MeOH): δ 8.65 (d, J = 12.6 Hz, 2H), 8.28 (s, 1H),

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

8.13 (t, J = 7.5 Hz, 2H), 7.86 (d, J = 16.0 Hz, 2H), 7.60 (t, J = 8.0 Hz, 1H), 7.47 (d, J = 7.7 Hz, 1H), 7.22 (s, 1H), 6.77 (d, J = 16.0 Hz, 1H), 3.82 (s, 3H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 166.95, 158.14, 155.68, 155.66, 143.82, 132.74, 132.59, 130.18, 129.89, 129.58, 128.93, 126.04, 125.98, 124.12, 120.46, 119.50, 118.57, 118.52, 52.10. HRMS (ESI) m/z calcd for $C_{19}H_{14}F_3N_3O_2$: $[M + H]^+$, 374.1116; found, 374.1116.

3.3. Biological Evaluation. *3.3.1. Cell Culture.* MCF-7 and MDA-MB-231 breast cancer cell lines were obtained from the Korean Cell Center (KCL, Seoul, Kr). In addition to 1% antibiotic and 10% fetal bovine serum, Roswell Park Memorial Institute Medium (RPMI) 1640 was used to cultivate both cell lines (FBS). The cells were kept at 37 °C in an environment that contained 5% CO_2 and was 95% humid.[45]

3.3.2. MTT Assay. The MTT test was used to further examine the intracellular cytotoxicity effect of the compounds as they were produced. Moreover, the experiments were separated into the treated group and control group subcategories. MCF-7 and MDA-MB-231 cell lines were first seeded in 96-well plates (4×10^4 cells per well) in each group, and they were then incubated for 24 h at 37 °C in an environment that contained 5% CO_2. The culture plate's medium was then taken out, and 100 μL of RPMI media alone and media containing various concentrations of compounds (2.5, 5, 10, 20, and 40 μM) were added to the control and treatment cell lines in each well, respectively. The cells were then incubated for 24 h at 37 °C in an environment that contained 5% CO_2. Subsequently, an MTT solution (150 μL, 1 mg/mL) was added to each well in place of the compound-containing media. Once the MTT reagent had been incubated for 4 h, 200 μL of DMSO was added to dissolve the formazan crystals. A BioTek Synergy H1 analyzer was then used to evaluate the color intensity (BioTek, Winooski, VT, USA). (1) Using Graph Pad Prism 5, the IC_{50} was determined using the Boltzmann sigmoidal concentration−response equation.[45]

3.3.3. EGFR Assay. The 4 compounds displaying a cytotoxicity of <5 μM against both the tested cancer cell lines were subjected to EGFR inhibitory assay at Reaction Biology Corporation using the "HotSpot" assay platform in 10-dose IC_{50} mode with 3-fold serial dilutions starting at 10 μM. Compound **11h** was not among the tested compounds as it displayed high intrinsic variability in the cytotoxicity tests. Control compound, Staurosporine, was tested in 10-dose IC_{50} mode with 4-fold serial dilution starting at 20 μM, and reactions were carried out at 10 μM ATP. The Supporting Information includes raw data, % enzyme activity (relative to DMSO controls), and curve fits (curve fits were obtained where the enzyme activities at the highest concentration of compounds were less than 65%.)

3.3.4. Analysis of the Cell Cycle Distribution. Cell cycle analysis was carried out to ascertain the effect of compound 9 on the distribution of the cell cycle in the MCF-7 and MDA-MB-231 cell lines. Analyzing cell cycles involves comparing the change in the cell cycle to that of the control group. The cell cycle arrest phase for the test sample was calculated by using untreated control cells as a standard. The Cell Cycle Analysis Kit was used to perform cell cycle analysis (ADAMII LS, NanoEntek, Seoul, Korea). The cancer cells were initially plated in 6-well plates (0.8×10^6 cells per well), and then they were incubated for 24 h at 37 °C with 5% CO_2. After removal of the medium from the growth plate, 3 mL of RPMI media containing compound **9** concentrations at IC_{50} values was applied to the cells in each well. The cells were then incubated for 24 h in a 5% CO_2 humidified incubator at 37 °C. Following a single PBS wash, the medium was removed, and 25 μL of the cell sample mixes was stained with 25 μL of propidium iodide. The ADAMII LS assay slide was loaded with 25 μL of the sample mixture, and the slide was incubated at room temperature for 1 min in the dark before being run on the ADAMII LS fluorescent cell analyzer. The software ADAMII LS was used to calculate the cell cycle distribution (ADAMII LS, NanoEntek, Seoul, Korea).

3.3.5. Apoptosis Analysis. An apoptosis study with Annexin V-PE, DAPI solution, was used to evaluate the apoptotic effect of compound **9** on MCF-7 and MDA-MB-231 cell lines (ADAMII LS, NanoEntek, Seoul, Korea). In comparison to the control, early and late apoptotic effects were examined. On a 6-well plate with a cell density of 0.8×10^5 cells/well, a volume of 3 mL of RPMI and DMEM media was introduced, and the cells were then incubated for 24 h at 37 °C in a 5% CO_2 environment. Afterward, 3 mL of RPMI and DMEM media containing compound **9** was added to the culture plate's medium. The medium was then incubated for 24 h at 37 °C and with 5% CO_2 in place of the original medium. The ADAMII LS apoptosis analysis kit was used to conduct the apoptosis assay.

After the sample was washed with PBS, a cell scraper (Alfa Aesar) was used to prepare the apoptosis-induced cell sample (PBS). The cell was resuspended in 100 μL of 1× Annexin V binding buffer; 5 μL of Annexin V-PE reagent was then added; and the mixture was then incubated at room temperature for 15 min. After centrifuging the material, 1.25 μL of DAPI dye and 500 μL of 1× Annexin V binding buffer were added to the pallet for its resuspension. The prepared sample was put onto a slide, which was then incubated at room temperature for 1 min in the dark before being run through the ADAMII LS fluorescent cell analyzer. ADAMII LS software was used to calculate and evaluate the comparison data, including pictures and dot plot graphs (NanoEntek).

3.3.6. Statistical Analysis. The statistical analysis of the data was done by standard deviations, and all values represent the mean ± SD of three independent experiments.

3.4. Molecular Docking. The X-ray crystal structures were downloaded from the Protein databank (PDB ID:1M17), which was used to define the binding mode.[46,47] The protein structure were then prepared by removing the water molecules and utilizing Maestro Schrodinger's Ligprep module to add any missing residues or hydrogen atoms.[48] To predict the binding modes and biological activity of the synthesized drugs, a molecular docking investigation was carried out. The Maestro Schrodinger Glide extra precision module was used to dock the synthesized compounds, generating 32 poses for each ligand. The BIOVIA Discovery Studio Visualizer 2021 was used to visualize the posture with the lowest energy score.[31]

4. CONCLUSIONS

This study aimed to develop novel EGFR inhibitors for breast cancer treatment by integrating ML approaches with a rational drug design. From the ChEMBL database, a collection of EGFR inhibitors was gathered and employed to create QSAR models for projecting the EGFR activity. Among the models tested, the Random Forest algorithm demonstrated the best performance for EGFR activity estimation, and subsequently, a web application was constructed utilizing this model. Using this prediction model, a hybridization strategy was employed to combine the *N*-substituted quinazolin-4-amine scaffold with

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

various FDA-approved kinase inhibitors. Eighteen novel compounds were synthesized, and their IC_{50} values were tested against MCF-7 and MDA-MB-231 cell lines. Compounds **9**, **10d**, **11g**, and **12** demonstrated significant antiproliferative activity with IC_{50} values below 5 μM against both cell lines and were selected for further testing of their EGFR inhibitory activity in vitro. Among the four tested compounds, compound **9** exhibited the most potent activity against EGFR with IC_{50} of 2.53 nM. Cell cycle and apoptosis analyses of compound **9** were promising, marking it as a potential therapeutic candidate for breast cancer. However, further studies are required to establish its pharmacokinetic properties as well as its efficacy and safety for clinical use.

■ ASSOCIATED CONTENT

Data Availability Statement

The spectra data (^{1}H NMR, ^{13}C NMR, mass, and HPLC) for the synthesized compounds can be found in the Supporting Information. Additionally, the web-based application that features the best ML model and the various databases generated and utilized in this study can be accessed through the following data repository: https://osf.io/4ebnm/?view_only=5a34644fda7c444da6bcdd981739b3c5.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c02799.

^{1}H NMR and ^{13}C NMR data for all compounds ; HPLC purity data for all compounds; representative HRMS data for all compounds; machine learning databases, web-based application, permutation test code, and its related databases can be accessed on the online data repository: https://osf.io/4ebnm/?view_only=5a34644fda7c444da6bcdd981739b3c5 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Tae Jung Park − *Department of Chemistry, Chung-Ang University, Seoul 06974, South Korea;* orcid.org/0000-0001-8918-0957; Email: tjpark@cau.ac.kr

Kyeong Lee − *BK21 FOUR Team and Integrated Research Institute for Drug Development, College of Pharmacy, Dongguk University-Seoul, Goyang 10326, Republic of Korea;* orcid.org/0000-0002-5455-9956; Email: kaylee@dongguk.edu

Authors

Hossam Nada − *BK21 FOUR Team and Integrated Research Institute for Drug Development, College of Pharmacy, Dongguk University-Seoul, Goyang 10326, Republic of Korea;* orcid.org/0000-0002-4882-5621

Anam Rana Gul − *Department of Chemistry, Chung-Ang University, Seoul 06974, South Korea*

Ahmed Elkamhawy − *BK21 FOUR Team and Integrated Research Institute for Drug Development, College of Pharmacy, Dongguk University-Seoul, Goyang 10326, Republic of Korea; Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt*

Sungdo Kim − *BK21 FOUR Team and Integrated Research Institute for Drug Development, College of Pharmacy, Dongguk University-Seoul, Goyang 10326, Republic of Korea*

Minkyoung Kim − *BK21 FOUR Team and Integrated Research Institute for Drug Development, College of Pharmacy, Dongguk University-Seoul, Goyang 10326, Republic of Korea*

Yongseok Choi − *College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Republic of Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c02799

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) [no. 2018R1A5A2023127] and [no. 2023R1A2C3004599]. This work is also supported by the BK21 FOUR program, which was funded by the Ministry of Education of Korea through NRF.

■ REFERENCES

(1) Arnold, M.; Morgan, E.; Rumgay, H.; Mafra, A.; Singh, D.; Laversanne, M.; Vignat, J.; Gralow, J. R.; Cardoso, F.; Siesling, S.; Soerjomataram, I. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast* **2022**, *66*, 15−23.

(2) Nada, H.; Elkamhawy, A.; Lee, K. Structure Activity Relationship of Key Heterocyclic Anti-Angiogenic Leads of Promising Potential in the Fight against Cancer. *Molecules* **2021**, *26* (3), 553.

(3) Tzenios, N. A New Hallmark of Cancer: Stemness. *Special journal of the Medical Academy and other Life Sciences.* **2023**, *1* (1), 1−6.

(4) Skandalis, S. S.; Afratis, N.; Smirlaki, G.; Nikitovic, D.; Theocharis, A. D.; Tzanakakis, G. N.; Karamanos, N. K. Cross-talk between estradiol receptor and EGFR/IGF-IR signaling pathways in estrogen-responsive breast cancers: Focus on the role and impact of proteoglycans. *Matrix Biol.* **2014**, *35*, 182−193.

(5) Gialeli, C.; Theocharis, A. D.; Karamanos, N. K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *FEBS J.* **2011**, *278* (1), 16−27.

(6) Elkamhawy, A.; Hassan, A. H. E.; Paik, S.; Sup Lee, Y.; Lee, H.-H.; Shin, J.-S.; Lee, K.-T.; Roh, E. J. EGFR inhibitors from cancer to inflammation: Discovery of 4-fluoro-N-(4-(3-(trifluoromethyl)-phenoxy)pyrimidin-5-yl)benzamide as a novel anti-inflammatory EGFR inhibitor. *Bioorg. Chem.* **2019**, *86*, 112−118.

(7) Zubair, T.; Bandyopadhyay, D. Small Molecule EGFR Inhibitors as Anti-Cancer Agents: Discovery, Mechanisms of Action, and Opportunities. *Int. J. Mol. Sci.* **2023**, *24* (3), 2651.

(8) Ye, P.; Wang, Y.; Li, R.; Chen, W.; Wan, L.; Cai, P. The HER family as therapeutic targets in colorectal cancer. *Crit. Rev. Oncol. Hematol.* **2022**, *174*, 103681.

(9) Mourad, M. A. E.; Abo Elmaaty, A.; Zaki, I.; Mourad, A. A. E.; Hofni, A.; Khodir, A. E.; Aboubakr, E. M.; Elkamhawy, A.; Roh, E. J.; Al-Karmalawy, A. A. Novel topoisomerase II/EGFR dual inhibitors: design, synthesis and docking studies of naphtho[2′,3′:4,5]thiazolo-[3,2-a]pyrimidine hybrids as potential anticancer agents with apoptosis inducing activity. *J. Enzyme Inhib. Med. Chem.* **2023**, *38* (1), 2205043.

(10) Nicholson, R. I.; Gee, J. M. W.; Harper, M. E. EGFR and cancer prognosis. *Eur. J. Cancer* **2001**, *37*, 9−15.

(11) Ciardiello, F.; Tortora, G. Epidermal growth factor receptor (EGFR) as a target in cancer therapy: understanding the role of receptor expression and other molecular determinants that could influence the response to anti-EGFR drugs. *Eur. J. Cancer* **2003**, *39* (10), 1348−1354.

(12) Sigismund, S.; Avanzato, D.; Lanzetti, L. Emerging functions of the EGFR in cancer. *Mol. Oncol.* **2018**, *12* (1), 3−20.

(13) Elkamhawy, A.; Paik, S.; Hassan, A. H. E.; Lee, Y. S.; Roh, E. J. Hit discovery of 4-amino-N-(4-(3-(trifluoromethyl)phenoxy)-pyrimidin-5-yl)benzamide: A novel EGFR inhibitor from a designed small library. *Bioorg. Chem.* **2017**, *75*, 393−405.
(14) Uribe, M. L.; Marrocco, I.; Yarden, Y. EGFR in Cancer: Signaling Mechanisms, Drugs, and Acquired Resistance. *Cancers* **2021**, *13* (11), 2748.
(15) Herceg, Z.; Hainaut, P. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. *Mol. Oncol.* **2007**, *1* (1), 26−41.
(16) Zandi, R.; Larsen, A. B.; Andersen, P.; Stockhausen, M.-T.; Poulsen, H. S. Mechanisms for oncogenic activation of the epidermal growth factor receptor. *Cell. Signal.* **2007**, *19* (10), 2013−2023.
(17) Cheng, W.-L.; Feng, P.-H.; Lee, K.-Y.; Chen, K.-Y.; Sun, W.-L.; Van Hiep, N.; Luo, C.-S.; Wu, S.-M. The Role of EREG/EGFR Pathway in Tumor Progression. *Int. J. Mol. Sci.* **2021**, *22* (23), 12828.
(18) Holohan, C.; Van Schaeybroeck, S.; Longley, D. B.; Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* **2013**, *13* (10), 714−726.
(19) Foley, J.; Nickerson, N. K.; Nam, S.; Allen, K. T.; Gilmore, J. L.; Nephew, K. P.; Riese, D. J. EGFR signaling in breast cancer: Bad to the bone. *Semin. Cell Dev. Biol.* **2010**, *21* (9), 951−960.
(20) Ramani, S.; Samant, S.; Manohar, S. M. The story of EGFR: from signaling pathways to a potent anticancer target. *Future Med. Chem.* **2022**, *14* (17), 1267−1288.
(21) Li, T.; Fu, W.; Lei, C.; Hu, S. Chapter 1 - Current status of anti-EGFR agents. In *Novel Sensitizing Agents for Therapeutic Anti-EGFR Antibodies*; Hu, S., Ed.; Academic Press, 2023; pp 1−12.
(22) Ma, G.; Deng, Y.; Qian, L.; Vallega, K. A.; Zhang, G.; Deng, X.; Owonikoko, T. K.; Ramalingam, S. S.; Fang, D. D.; Zhai, Y.; Sun, S.-Y. Overcoming acquired resistance to third-generation EGFR inhibitors by targeting activation of intrinsic apoptotic pathway through Mcl-1 inhibition, Bax activation, or both. *Oncogene* **2022**, *41* (12), 1691−1700.
(23) Nada, H.; Kim, S.; Park, S.; Lee, M. Y.; Lee, K. Identification of Potent hDHODH Inhibitors for Lung Cancer via Virtual Screening of a Rationally Designed Small Combinatorial Library. *ACS Omega* **2023**, *8* (24), 21769−21780.
(24) Álvarez-Machancoses, Ó.; Fernández-Martínez, J. L. Using artificial intelligence methods to speed up drug discovery. *Expet Opin. Drug Discov.* **2019**, *14* (8), 769−777.
(25) Lin, X.; Li, X.; Lin, X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* **2020**, *25* (6), 1375.
(26) Nada, H.; Sivaraman, A.; Lu, Q.; Min, K.; Kim, S.; Goo, J.-I.; Choi, Y.; Lee, K. Perspective for Discovery of Small Molecule IL-6 Inhibitors through Study of Structure-Activity Relationships and Molecular Docking. *J. Med. Chem.* **2023**, *66*, 4417−4433.
(27) Achary, P. G. R. Applications of Quantitative Structure-Activity Relationships (QSAR) based Virtual Screening in Drug Design: A Review. *Mini Rev. Med. Chem.* **2020**, *20* (14), 1375−1388.
(28) Perkins, R.; Fang, H.; Tong, W.; Welsh, W. J. Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.* **2003**, *22* (8), 1666−1679.
(29) Simeon, S.; Anuwongcharoen, N.; Shoombuatong, W.; Malik, A. A.; Prachayasittikul, V.; Wikberg, J. E. S.; Nantasenamat, C. Probing the origins of human acetylcholinesterase inhibition via QSAR modeling and molecular docking. *PeerJ* **2016**, *4*, No. e2322.
(30) Melge, A. R.; Parate, S.; Pavithran, K.; Koyakutty, M.; Mohan, C. G. Discovery of Anticancer Hybrid Molecules by Supervised Machine Learning Models and in Vitro Validation in Drug Resistant Chronic Myeloid Leukemia Cells. *J. Chem. Inf. Model.* **2022**, *62* (4), 1126−1146.
(31) Nada, H.; Elkamhawy, A.; Lee, K. Identification of 1H-purine-2,6-dione derivative as a potential SARS-CoV-2 main protease inhibitor: molecular docking, dynamic simulations, and energy calculations. *PeerJ* **2022**, *10*, No. e14120.
(32) Zaretckii, M.; Bashkirova, I.; Osipenko, S.; Kostyukevich, Y.; Nikolaev, E.; Popov, P. 3D chemical structures allow robust deep learning models for retention time prediction. *Digit. Discovery* **2022**, *1* (5), 711−718.
(33) Randazzo, G. M.; Bileck, A.; Danani, A.; Vogt, B.; Groessl, M. Steroid identification via deep learning retention time predictions and two-dimensional gas chromatography-high resolution mass spectrometry. *J. Chromatogr. A* **2020**, *1612*, 460661.
(34) Elkamhawy, A.; Farag, A. K.; Viswanath, A. N. I.; Bedair, T. M.; Leem, D. G.; Lee, K.-T.; Pae, A. N.; Roh, E. J. Targeting EGFR/HER2 tyrosine kinases with a new potent series of 6-substituted 4-anilinoquinazoline hybrids: Design, synthesis, kinase assay, cell-based assay, and molecular docking. *Bioorg. Med. Chem. Lett.* **2015**, *25* (22), 5147−5154.
(35) Helmbach, H.; Rossmann, E.; Kern, M. A.; Schadendorf, D. Drug-resistance in human melanoma. *Int. J. Cancer* **2001**, *93* (5), 617−622.
(36) E Taylor, T.; B Furnari, F.; K Cavenee, W. Targeting EGFR for treatment of glioblastoma: molecular basis to overcome resistance. *Curr. Cancer Drug Targets* **2012**, *12* (3), 197−209.
(37) 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* (D1), D1100−D1107.
(38) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4* (11), 1465−1476.
(39) Murray, B.; Kerfoot, E.; Chen, L.; Deng, J.; Graham, M. S.; Sudre, C. H.; Molteni, E.; Canas, L. S.; Antonelli, M.; Klaser, K.; Visconti, A.; Hammers, A.; Chan, A. T.; Franks, P. W.; Davies, R.; Wolf, J.; Spector, T. D.; Steves, C. J.; Modat, M.; Ourselin, S. Accessible data curation and analytics for international-scale citizen science datasets. *Sci. Data* **2021**, *8* (1), 297.
(40) Bento, A. P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L. J.; De Veij, M.; Leach, A. R. An open source chemical structure curation pipeline using RDKit. *J. Cheminf.* **2020**, *12* (1), 51.
(41) Xiong, Z.; Cui, Y.; Liu, Z.; Zhao, Y.; Hu, M.; Hu, J. Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Comput. Mater. Sci.* **2020**, *171*, 109203.
(42) Wong, T. T.; Yeh, P. Y. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Trans. Knowl. Data Eng.* **2020**, *32* (8), 1586−1594.
(43) Khorasani, M.; Abdou, M.; Hernández Fernández, J. Streamlit at Work. In *Web Application Development with Streamlit: Develop and Deploy Secure and Scalable Web Applications to the Cloud Using a Pure Python Framework*; Apress: Berkeley, CA, 2022; pp 363−379.
(44) Filgueiras, P. R.; Alves, J. C. L.; Sad, C. M. S.; Castro, E. V. R.; Dias, J. C. M.; Poppi, R. J. Evaluation of trends in residuals of multivariate calibration models by permutation test. *Chemom. Intell. Lab. Syst.* **2014**, *133*, 33−41.
(45) Gul, A. R.; Shaheen, F.; Rafique, R.; Bal, J.; Waseem, S.; Park, T. J. Grass-mediated biogenic synthesis of silver nanoparticles and their drug delivery evaluation: A biocompatible anti-cancer therapy. *Chem. Eng. J.* **2021**, *407*, 127202.
(46) Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. Structure of the Epidermal Growth Factor Receptor Kinase Domain Alone and in Complex with a 4-Anilinoquinazoline Inhibitor*. *J. Biol. Chem.* **2002**, *277* (48), 46265−46272.
(47) To, C.; Beyett, T. S.; Jang, J.; Feng, W. W.; Bahcall, M.; Haikala, H. M.; Shin, B. H.; Heppner, D. E.; Rana, J. K.; Leeper, B. A.; Soroko, K. M.; Poitras, M. J.; Gokhale, P. C.; Kobayashi, Y.; Wahid, K.; Kurppa, K. J.; Gero, T. W.; Cameron, M. D.; Ogino, A.; Mushajiang, M.; Xu, C.; Zhang, Y.; Scott, D. A.; Eck, M. J.; Gray, N. S.; Jänne, P. A. An allosteric inhibitor against the therapy-resistant mutant forms of EGFR in non-small cell lung cancer. *Nat. Cancer* **2022**, *3* (4), 402−417.

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800

(48) Nada, H.; Kim, S.; Godesi, S.; Lee, J.; Lee, K. Discovery and optimization of natural-based nanomolar c-Kit inhibitors via in silico and in vitro studies. *J. Biomol. Struct. Dyn.* **2023**, 1−12.

https://doi.org/10.1021/acsomega.3c02799
ACS Omega 2023, 8, 31784−31800