Vol. 13, No. 2 (2025), pp.133–141

CHITKARA UNIVERSITY

Journal of Pharmaceutical Technology Research and Management



2D-QSAR Modeling of Quinazoline Derivatives as EGFR Inhibitors for Colorectal Cancer

Rishab Pathak[1], Mayank Garhewal[1], Shani Yadav[1], Jyoti Prakash Panda[1], Achal Mishra[1], Ayush Tiwari[1], Avantika Agrawal[1], Shekhar Verma[1] and Yogesh Vaishnav[1*]

[1]*Department of Pharmacy, Guru Ghasidas Vishwavidyalaya (A Central University), Bilaspur 495009, Chhattisgarh, India*

[*]yogesh446688@gmail.com (Corresponding Author)

RESEARCH ARTICLE **Open Access**

ARTICLE INFORMATION

Received: June 13, 2026
Accepted: July 28, 2026
Published Online: August 13, 2026

Keywords:
Quantitative structure-activity relationship, Partial least squares, Cross-validated coefficient, Quinazoline, Colon cancer

DOI: 10.15415/jptrm.2025.132008

ABSTRACT

Background: Globally, colorectal cancer represents one of the most common types of cancer, along with being one of the top 10 causes of cancer-related deaths. Epidermal Growth Factor Receptors (EGFRs) are a major target for colorectal cancer, and quinazoline derivatives have shown anticancer activity as EGFR inhibitors.

Purpose: The objective of this research was the development of a two-dimensional quantitative structure–activity relationship model (2D-QSAR); that is, to develop a simple way to identify the structural components that help predict a potent compound with the highest predicted activity among the series of quinazoline derivatives for EGFR inhibitors for colorectal cancer.

Methods: A dataset of 21 quinazoline derivatives was collected from the literature, molecular descriptors were calculated, and Partial Least Squares (PLS) regression was used to develop statistically significant QSAR models. Generated and validated using internal and external validation parameters.

Results: Model-1 (RANDOM_70_30_SFB_PLS_TRIALS_2) showed the best statistical performance with $r^2 = 0.6538$, $q^2 = 0.5314$, pred-$r^2 = 0.6940$, and an F-test value of 22.6638. The descriptors SsssNcount and SaaCHE-index were identified as significant contributors to EGFR inhibitory activity, indicating that tertiary nitrogen count and charge distribution play crucial roles in determining biological potency.

Conclusion: The developed 2D-QSAR model identified a potent compound with the highest predicted biological activity as an EGFR inhibitor for colorectal cancer.

1. Introduction

Cancer is a major global health problem affecting both developed and developing countries. Despite substantial improvements in cancer treatment, the disease keeps appearing as the most challenging illness in the world (Ryad *et al.*, 2024). According to the World Health Organization (WHO), cancer is the second leading cause of death worldwide, accounting for around 9.6 million fatalities in 2018. According to projections, the annual incidence of new cancer cases is expected to increase significantly, from 14.1 million in 2012 to roughly 21.6 million by 2030 (Pravin & Sudhir, 2018).

Cancer comprises more than 200 different diseases and is actually a collection of diseases characterized by unregulated and uncontrollable cell proliferation opposed to a single disease (Mohamed *et al.*, 2024). An imbalance between cell division and apoptosis triggers inappropriate cellular proliferation, which is identified as cancer (Mansouri *et al.*, 2025). Colorectal cancer (Figure 1) is the fourth leading cause of cancer-related death worldwide and the third most prevalent type of cancer to be diagnosed. Due to poor dietary habits, high-income nations have the greatest incidence and fatality rates. Other variables include a history of colon polyps and being older than 50 (Figure 2). Colon cancer risk factors include a personal history of inflammatory bowel disease, a family history of colon cancer, obesity, a lack of exercise, red meat consumption, using tobacco, and alcohol use. In underdeveloped nations, colorectal cancer is particularly more common. According to reports, around 1 million deaths and nearly 2 million new cases are anticipated in 2018 (Bentharavithana *et al.*, 2025). For advanced-stage colorectal cancer, systemic therapies are recommended. These usually involve first- and second-line

©Author(s) 2025. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are also included in the article's Creative Commons licence. To view a copy of this licence, visit https://creativecommons.org/licenses/by/4.0/

medications such as cetuximab, irinotecan, bevacizumab, oxaliplatin, and fluorouracil. These treatments are available, but their efficacy is severely limited by the development of drug resistance and adverse responses, which frequently result in recurrence and the spread of metastases (Wang *et al.*, 2025).

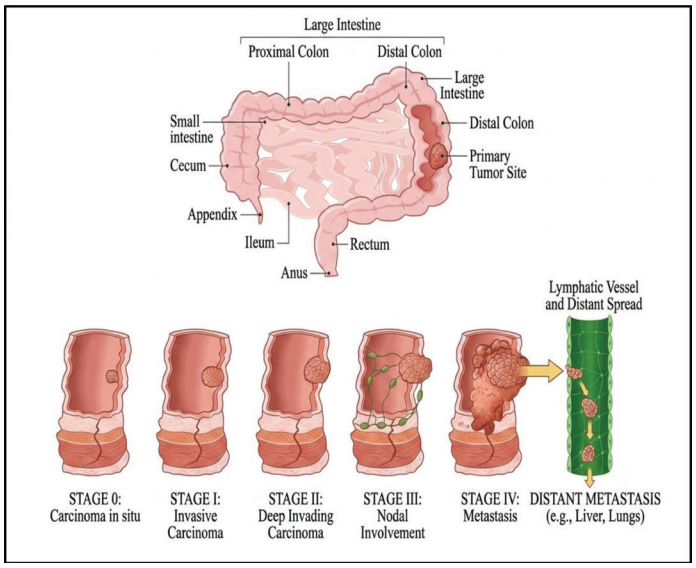


Figure 1: Colon Cancer Pathology and Staging

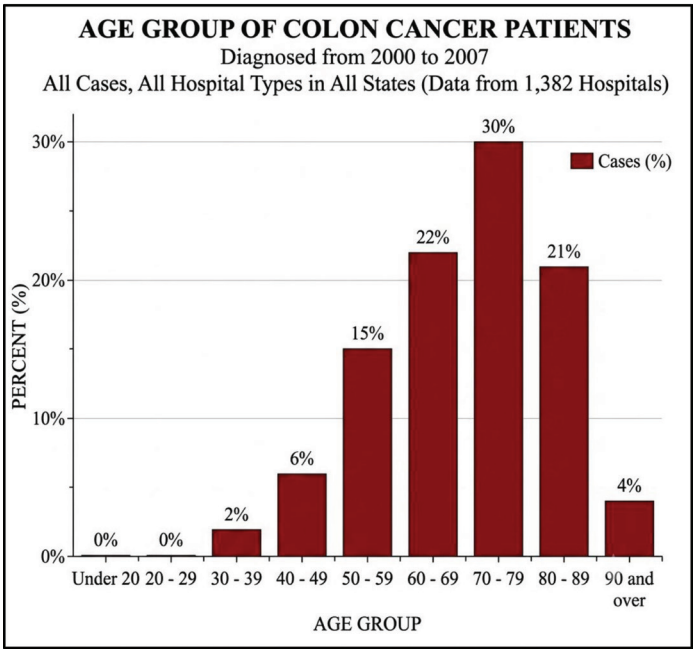


Figure 2: Age Group of Colon Cancer

Quinazoline is a bicyclic heterocyclic chemical that is produced when two aromatic rings with six members—pyrimidine and benzene—combine. With two nitrogen atoms, these compounds are classified as 1,3-benzodiazine (diazonaphthalene). There are four isomeric versions of the quinazoline ring, including quinazoline, quinoxaline, cinnolines, and pthalazines, depending on where the nitrogen atoms are located (Figure 1). Depending on where the keto group is located, quinazolinones, 2,4(1H,3H)-quinazolinedione, 4(3H)-quinazolinones, and 2(1H,3H)-quinazolinone are further subtypes of quinazolines in their oxidized form (Dhuguru & Ghoneim, 2022) (Figure 3).

Quinazolines can inhibit functions of the following: tubulin blockers, adenosine 5′-triphosphate-binding cassette (ABC) transporters, receptor tyrosine kinases (RTKs), phosphatidylinositol-3-kinase (PI3K), Heat Shock Protein 90 (Hsp90), and the Epidermal Growth Factor Receptor (EGFR) (Radzikowska-Bűchner *et al.*, 2025).

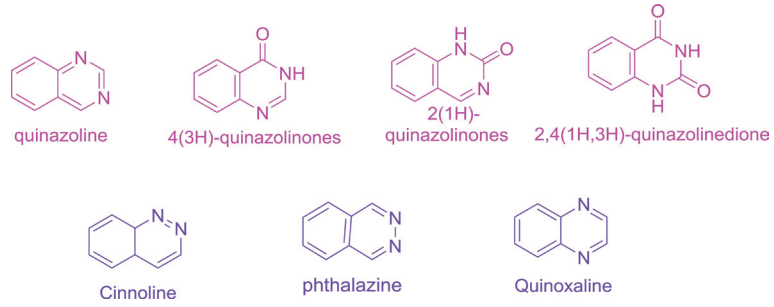
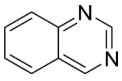


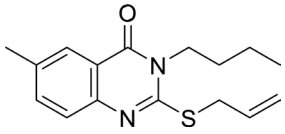
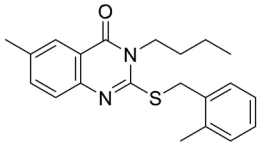
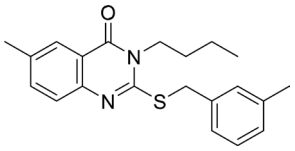
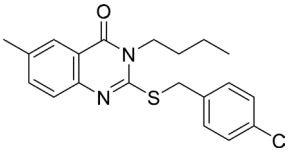
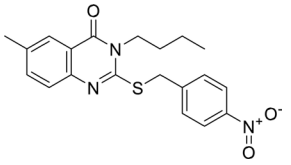
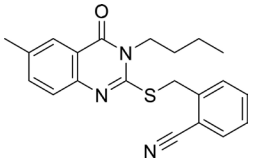
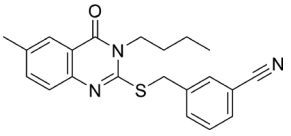
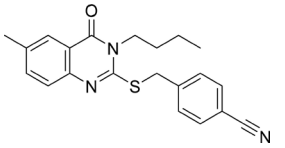
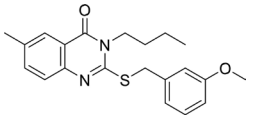
Figure 3: Isomers of Quinazolines and their Oxidized Forms Structures

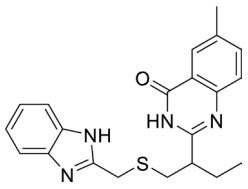
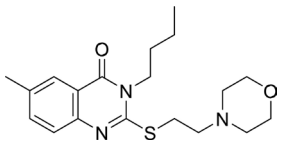
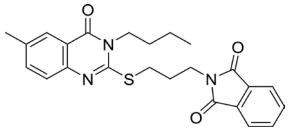
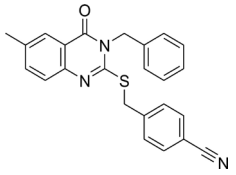
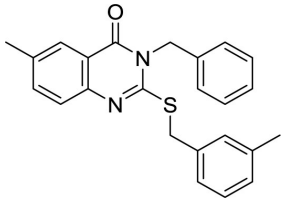
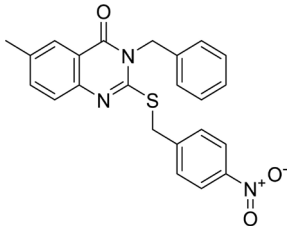
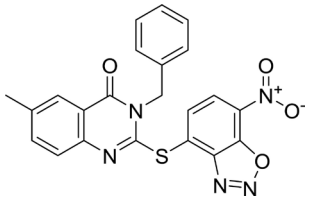
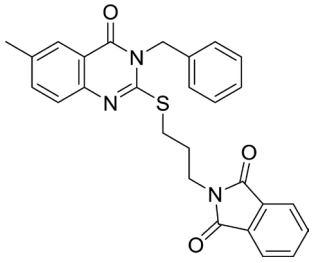
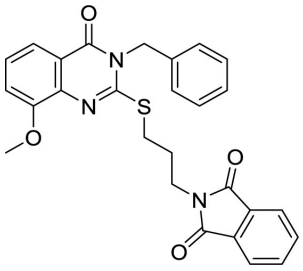
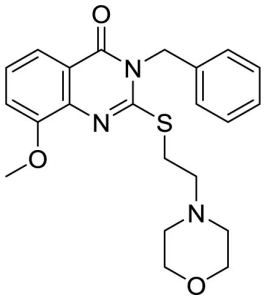
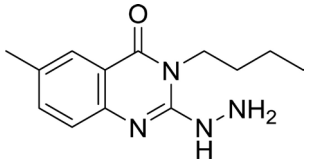
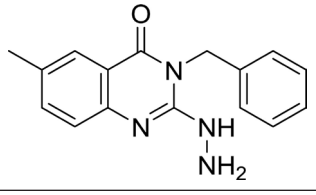
2. 2D-QSAR Analysis

In order to identify new drug candidates, virtual libraries and chemical databases must be screened for compounds with known activity using reliable and effective methods. For the purpose of creating new drug candidates, the QSAR model offers a practical means of investigating and taking advantage of the connection between chemical structures and their biological activities (Vaishnav *et al.*, 2026; Zhang *et al.*, 2024). By using 2D-QSAR techniques, it is possible to investigate a compound's physicochemical characteristics and how well they inhibit particular biological targets (Vaishnav *et al.*, 2018). The primary focus of this work is the 2D-QSAR analysis of many substituted quinazoline compounds designed as EGFR inhibitors. Our research aims to find structural variables that influence the biological activity of novel therapeutic drugs against cancer in order to provide insights that help guide their rational design.

Table 1: Structural Information of Binding Affinities of Quinazoline Derivatives



S. No.	Structure	IC_{50} Value	pIC_{50} value
1		5.65	5.24
2		6.3	5.20
3		4.45	5.35
4		3.7	5.43
5		4.56	5.34
6		3.7	5.43
7		6.1	5.21
8		7.5	5.12
9		6.79	5.16

10		7.8	5.10
11		4.11	5.38
12		5.75	5.38
13		3.9	5.40
14		3.04	5.51
15		5.6	5.25
16		4.14	5.38
17		1.85	5.73
18		2.5	5.60
19		2.6	5.58
20		5.39	5.26
21		4.77	5.32

3. Materials And Methods

3.1. Data Collection/Data Set

A dataset comprising twenty-one molecules was obtained from the study by Abuelizz *et al.* (2017), which reported the synthesis and anticancer activity of novel quinazoline derivatives. Synthesis and anticancer activity of new quinazoline derivatives. *Saudi Pharmaceutical Journal*, *25*(7), 1047-1054]. The biological activities of a particular data collection are shown in Table 1. The biological activity of each compound was expressed as an IC_{50} value. Which were subsequently transformed into pIC50 (–logIC50) values and used as a dependent variable. The entire dataset was separated into two parts: a test set for validating the created models and a training set for building the QSAR models. The formula for converting IC_{50} (in molar concentration) to pIC_{50} is as follows: $pIC_{50} = -\log_{10}(IC_{50})$.

3.2. 2D-QSAR Study

V Life MDS 2D Draw was used to illustrate the compounds' structures from a few selected series, and the resulting images were then exported to the V Life MDS window. Before being employed in 2D QSAR studies, by adopting the UFF force field approach, the energy of the compounds was reduced until the Root Mean Square (RMS) gradient was 0.01 kcal/mol Å. Afterwards, batch optimization was done. Physicochemical, alignment-independent, and atom-type descriptors were calculated after the molecules in the data set were optimized. Computed descriptors were regarded as dependent factors and biological activity as independent variables. The following physicochemical descriptors were computed: Atomic Valence Connectivity Index (ChiV); Path Count, Chi Chain, Chiv Chain, Chain Path Count, Cluster, Path Cluster, Kapa; Element Count (Hydrogen, Nitrogen, Carbon, Sulfur, Oxygen, Chlorine, Bromine, Iodine); retention index (Chi); individual (Hydrogen-Acceptor Count, Hydrogen-Donor Count, X logP, SMR, Polarizability, etc.); and Polar Surface Area. The following characteristics were also utilized to generate more than 150 independent descriptors of alignment (Alsfouk *et al.*, 2026) (Table 2).

Table 2: Selected Descriptors and Attributes

Structural descriptors	Selected Attributes
*Topological Range Min–0 Max.-7	2 3 T (any) Carbon Nitrogen Oxygen Fluorine Sulfur Clorine Bromine Iodine

3.3. Creation of the Sets of Training and Test Compounds

To externally validate the QSAR model, the dataset was divided into training and test sets using techniques such as random selection, manual data selection, and sphere exclusion. Using the training set, the quantitative structure-activity relationship model is created for which biological activity data are available. The test set is used to test the quantitative structure activity relationship model formed using the training set in order to assess the predictive effectiveness of the model (Raut *et al.*, 2023).

Example: The entire dataset was separated into two parts: a test set for validating the created models and a training set for building the QSAR models.

3.4. Random Selection

Training data sets were divided into 90%–60% (90%, 85%, 80%, 75%, 70%, 65%, and 60%) of the entire data set at random. test [10%–40% (10%, 15%, 20%, 30%, 35%, and 40%) of the entire data set] to develop and validate the QSAR models both internally and externally. Ten trials were carried out in each case for the development and validation of the QSAR models.

3.5. Manual Data Selection

The full range of activities was ordered in both ascending and descending order, with the test set receiving each of the following compounds: 4th, 5th, 6th, 7th, 8th, 9th, and 10th.

3.6. Statistical Calculation

To develop QSAR models, statistical methods such as partial least squares regression (PLSR), stepwise multiple linear regression (MLR), and partial component regression (PCR) were applied. The following statistical features were considered in order to assess the statistical significance of QSAR models: F-Test (F-test for statistical significance of the model), cross-validated squared correlation coefficient (q^2), squared correlation coefficient (r^2), and predicted correlation coefficient (Panda *et al.*, 2026).

4. Results And Discussion

After all 21 compounds of the selected series completed different regression analyses, the following significant QSAR models with equations for EGFR inhibitor activity were (Table 3).

Table 3: List of Predictive QSAR Models Generated from the PLS Regression Method

Model Number	Method	Equation
1	RANDOM_70_30_ SFB_PLS_ TRIALS_2	pIC_{50}=0.2108SsssNcount +0.0105 SaaCHE-index+4.9419 n=14, Degree Of Freedom=12, F Test =22.6638, r^2=0.6538,q^2=0.5314, predicted r^2=0.6940, r^2se(standard error) =0.0951, q^2 se=0.1106, predicted r^2 se=0.1051

r^2se is the standard error of estimate (lower is preferable), predicted r^2 is the predictive correlation coefficient, and n is the number of data points used in the aforementioned QSAR models.

Model-1 was selected as the optimal 2D-QSAR model. since it explains 65.38% (r^2 = 0.6538) of the total variance in the training set, has internal (q^2) and external (predicted r^2) predictive abilities of 53.14% and 69.4%, respectively, and an F-test of 22.6638.

Example: The current QSAR study yielded Model 1, which can be utilized to predict the activity of the newly created compounds in order to find some more potent compounds.

4.1. Y-randomization Analysis

Y-randomization (response permutation) analysis was performed to evaluate whether the developed QSAR model resulted from chance correlation as presented in Table 4. The obtained Z-score values for R^2 (3.68), Q^2 (3.93), and Pred-R^2 (3.76) were all greater than 3, indicating that the model is statistically significant and unlikely to have arisen by random chance. Furthermore, the randomized models exhibited substantially lower R^2 (0.243) and Q^2 (0.255) values than the original model, confirming the robustness of the developed QSAR model. The low alpha values (0.001 for R^2 and 0.010 for Q^2) further support the statistical reliability of the model. These results demonstrate that the developed PLS-QSAR model possesses good robustness and predictive reliability within the studied chemical space.

Table 4: Y-randomization (Response Permutation) Validation of the Developed PLS-QSAR Model

Dependent Variable	**ZScore** R^2	**ZScore** Q^2	**Best Rand R^2**	**Best Rand Q^2**	**Alpha Rand R^2**	**Alpha Rand Q^2**	**ZScore Pred-R^2**	**Best Rand Pred-R^2**	**Alpha Rand Pred-R^2**
pIC_{50}	3.68293	3.93128	0.24315	0.25513	0.001	0.01	3.75899	0.25461	0.05

Figure 4 compares the actual and predicted biological activity of the training and test sets were made for model 1.

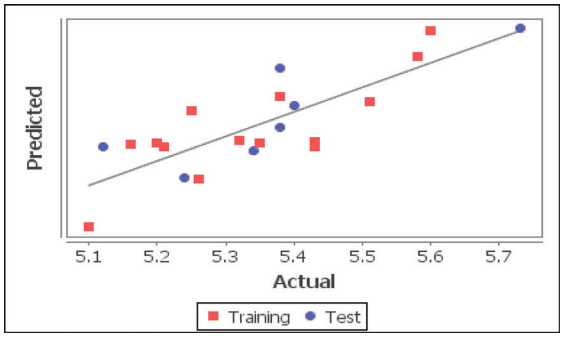


Figure 4: Fitness Plot of Model 1's Test and Training Sets

Example: Using the training set, the QSAR model was developed based on available biological activity data.

Figure 5 compares the training set with data collection of model-01's actual and predictive biological activity.

Training Set Model 1 Test Set Model 1

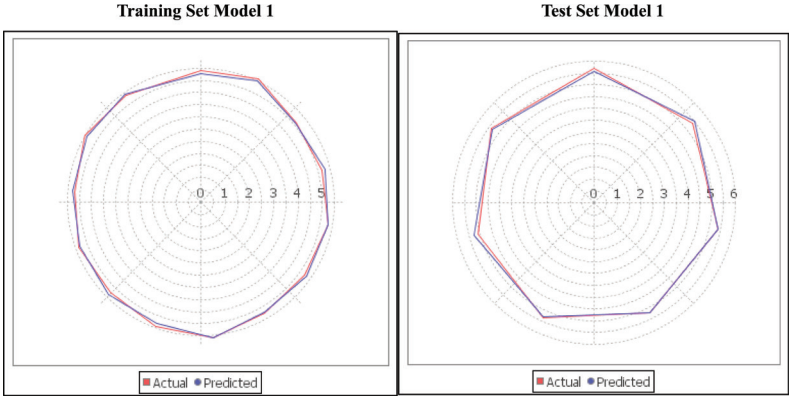


Figure 5: Comparison of the Training Set with Data Collection of Model-01's Actual and Predictive Biological Activity

Example: The model›s parameters are correlated with one another, as shown by the correlation matrix.

Model 1's correlation matrix was provided in Table 5. The model's parameters are correlated with one another, as shown by the correlation matrix.

Table 5: Model-01 Correlation Matrix for Descriptors

SsssNcount	SaaCHE-index
1	13.662
1	13.703
1	13.71
1	14.212
0	14.048
2	6.049
2	12.844
1	23.724
1	24.878
1	22.423
1	18.492
2	22.758
2	22.242
2	15.743
1	5.649
1	15.364
1	5.957
1	14.603
1	14.739
1	15.031
1	12.552

The contribution charts represent contribution of different descriptors in biological activity for model 1 and shown in Figure 6.

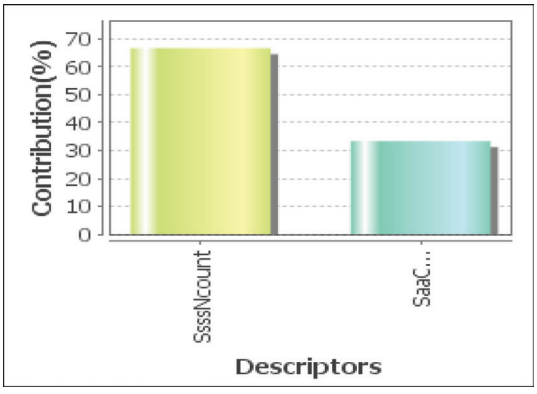


Figure 6: Model-01's Contribution Table for Different Biological Activity Descriptors

4.2. Interpretation of the result

Table 2 illustrates that PLS, in conjunction with stepwise forward-backward variable selection, produced a statistically significant model for the current investigation. SsssCount, which represents the number of tertiary nitrogen atoms, exhibited the highest positive contribution to the PLS model, indicating that tertiary nitrogen-containing substituents favor EGFR inhibitory activity by enhancing ligand-receptor interactions. SaaCHE-index, an electronic descriptor associated with the charge distribution of aromatic carbon atoms bonded to heteroatoms, also contributed positively, suggesting that an optimized electronic environment improves molecular recognition and binding affinity toward EGFR. Collectively, these descriptors indicate that both tertiary nitrogen substitution and favorable electronic characteristics are key structural determinants of the EGFR inhibitory activity of quinazoline derivatives.

Only two descriptors were retained in the final model because the stepwise forward–backward variable selection identified SsssNcount and SaaCHE index as the most statistically significant and nonredundant variables. Given the limited dataset size (n = 21), inclusion of additional descriptors did not improve predictive performance and could increase the risk of overfitting. Thus, a parsimonious two-descriptor model was selected to maximize robustness, interpretability, and predictive reliability.

5. Conclusion

A statistically significant 2D-QSAR model has been successfully developed to elucidate the structural features influencing the EGFR inhibitory activity of quinazoline derivatives. The Partial Least Squares (PLS) regression approach was used to evaluate a dataset that included twenty-one compounds. The best statistical performance was shown by Model-1 (RANDOM_70_30_SFB_PLS_TRIALS_2), which had an F-test value of 22.6638, a correlation coefficient (r^2 = 0.6538), a cross-validated correlation coefficient (q^2 = 0.5314), and a predicted correlation coefficient (pred-r^2 = 0.6940). The suggested model's internal consistency and dependability were validated by its low standard error of estimation. The descriptors SsssNcount and SaaCHE-index were identified as the most significant contributors to biological activity. The identified descriptors, SsssNcount and SaaCHE-index, suggest that tertiary nitrogen atoms and favorable electronic characteristics are associated with EGFR inhibitory activity in the studied quinazoline derivatives. However, these observations are based on statistical correlations obtained from the QSAR model and should not be interpreted as direct mechanistic evidence. The developed QSAR model provides useful structure–activity relationship

insights that can assist in the design and prioritization of new quinazoline derivatives targeting EGFR. Furthermore, the identified descriptors may serve as valuable parameters for virtual screening and future drug design. Additional studies, including molecular docking and experimental validation, are required to confirm the mechanistic role of these structural features in EGFR inhibition.

Acknowledgement

Authors sincerely thank the Department of Pharmacy at Guru Ghasidas Vishwavidyalaya for providing the laboratory facilities, computational resources, and infrastructure needed to complete this research.

Authorship Contribution

Rishab Pathak contributed to conceptualization, methodology, investigation, validation, writing–original draft, and writing–review & editing. Mayank Garhewal contributed to methodology, software, data curation, investigation, validation, writing–original draft, and writing–review & editing. Shani Yadav contributed to data curation, visualization, and formal analysis. Jyoti Prakash Panda contributed to data curation, visualization, and formal analysis. Achal Mishra contributed to conceptualization, methodology, software, visualization, formal analysis, and resources. Ayush Tiwari contributed to data curation, visualization, validation, and formal analysis. Avantika Agrawal contributed to data curation, visualization, validation, and formal analysis. Shekhar Verma contributed to data curation, visualization, validation, and formal analysis. Yogesh Vaishnav contributed to conceptualization, methodology, software, formal analysis, and resources. All authors reviewed the manuscript, approved the final version, and agreed to be accountable for the work.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Ethical Approvals No ethical approvals were required for this study.

Conflict of Interest

The authors declare no conflict of interest related to this study.

Declarations

The authors declare that this work is original and has not been submitted elsewhere for publication. All data, methodologies, and system components have been developed and reported in adherence to academic standards. All referenced materials have been duly cited, and the authors accept full responsibility for the integrity and accuracy of the findings presented.

Ethical Approval

Ethical approval was not required for this study as it involved only computational and in silico analyses and did not involve human participants, animals, or biological samples.

Figure Permissions

Authors declare that the entirety of this work, including text, figures, and tables, is original, has not been previously published, is in the public domain, or the authors are the exclusive copyright holders.

Data Availability Statement

The authors confirm that the data supporting the conclusions of this study are included within the article. Additional information is available from the corresponding author (Vaishnav Y) upon reasonable request.

Human and Animal Rights

No humans or animals were involved in this research.

AI Disclosure Statement

The authors declare that artificial intelligence (AI)-assisted tools were used only for language improvement, grammar refinement, and formatting assistance during manuscript preparation. All scientific content, literature interpretation, critical analysis, conclusions, and final manuscript review were performed by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the manuscript and confirm that it complies with the journal's plagiarism and publication ethics policies.

References

Abuelizz, H. A., Marzouk, M., Ghabbour, H., & Al-Salahi, R. (2017). Synthesis and anticancer activity of new quinazoline derivatives. *Saudi Pharmaceutical Journal*, *25*(7), 1047-1054.

Alsfouk, A. A., Channar, P. A., Bilal, M. S., Ejaz, S. A., & Kim, S. J. (2026). Designing of promising VEGFR-2 inhibitors by using QSAR modeling, molecular

docking, and molecular dynamic simulation: Implications toward the treatment of cancer. *Journal of Chemistry*, *2026*(1), 9405335. https://doi.org/10.1155/joch/9405335

Bentharavithana, J., Islam, T., & Xu, B. (2025). Medicinal mushrooms in colon cancer therapy: Mechanisms of action of bioactive compounds and therapeutic potential. *International Journal of Molecular Sciences, 26*(11), 5304. https://doi.org/10.3390/ijms26115304

Dhuguru, J., & Ghoneim, O. A. (2022). Quinazoline based HDAC dual inhibitors as potential anti-cancer agents. *Molecules, 27*(7), 2294. https://doi.org/10.3390/molecules27072294

Mansouri, M. M., Emami, L., Rezaei, Z., & Khabnadideh, S. (2025). Design, synthesis, biological assessments and computational studies of 3-substituted phenyl quinazolinone derivatives as promising anti-cancer agents. *BMC Chemistry, 19*(1), 125. https://doi.org/10.1186/s13065-025-01492-4

Mohamed, A. M., Abou-Ghadir, O. M. F., Mostafa, Y. A., Almarhoon, Z. M., Bräse, S., & Youssif, B. G. M. (2024). Design, synthesis, and antiproliferative activity of new 1,2,3-triazole/quinazoline-4-one hybrids as dual EGFR/BRAF V600E inhibitors. *RSC Advances, 14*(52), 38403–38415. https://doi.org/10.1039/D4RA06694D

Panda, J. P., Garhewal, M., Jain, S. K., Yadav, S., Mishra, A., Pathak, R., Tiwari, A., Verma, S., & Vaishnav, Y (2026). *Designing novel indole-based anticancer drugs: Insights from 3D-QSAR and molecular field analysis.*

Pravin, S., & Sudhir, A. (2018). Integration of 3D printing with dosage forms: A new perspective for modern healthcare. *Biomedicine & Pharmacotherapy, 107*, 146–154. https://doi.org/10.1016/j.biopha.2018.07.167

Radzikowska-Bűchner, E., Radej, S., Niezabitowska, E., Sitarz, R., & Szewc, M. (2025). Potential of using new indole- and benzimidazo[1,2-C]quinazolines in anticancer therapy based on mesenchymal stem cells. *Cancer Management and Research, 17*, 1087–1097. https://doi.org/10.2147/CMAR.S516593

Raut, V. V., Bhandari, S. V., Patil, S. M., & Sarkate, A. P. (2023). A rational approach to anticancer drug design: 2D and 3D-QSAR, molecular docking and prediction of ADME properties using silico studies of thymidine phosphorylase inhibitors. *Letters in Drug Design & Discovery, 20*(2), 153–166. https://doi.org/10.2174/1570180819666220215115633

Ryad, N., Elmaaty, A. A., M Ibrahim, I., Ahmed Maghrabi, A. H., Yahya Alahdal, M. A., Saleem, R. M., Zaki, I., & Ghany, L. M. A. A. (2024). Harnessing molecular hybridization approach to discover novel quinoline EGFR-TK inhibitors for cancer treatment. *Future Medicinal Chemistry, 16*(11), 1087–1107. https://doi.org/10.1080/17568919.2024.2342201

Vaishnav, Y., Thakur, A., Kaur, C. D., Verma, S., Mishra, A., Jain, S. K., & Ghode, P. (2018). QSAR analysis of some N, N-diphenyl urea derivatives as CCR5 receptor antagonist. *Research Journal of Pharmacy and Technology, 11*(9), 3802. https://doi.org/10.5958/0974-360X.2018.00697.2

Vaishnav, Y., Yadav, S., Jain, S. K., Panda, J. P., Mishra, A., Garhewal, M., Pathak, R., & Verma, S. (2026). 2D-QSAR analysis and molecular descriptor insights for designing potent anticancer pyrazolopyrimidine derivatives. *RASAYAN Journal of Chemistry, 19*(02), 921–931. https://doi.org/10.31788/RJC.2026.1929617

Wang, R., Han, X., Zhu, L., Wang, W., Gu, Y., Zhang, W., & Guo, W. (2025). Combination of fruquintinib with venetoclax for the treatment of colorectal cancer. *Oncology Research, 33*(1), 225–234. https://doi.org/10.32604/or.2024.050047

Zhang, J., Liu, Q., Zhao, H., Li, G., Yi, Y., & Shang, R. (2024). Design and synthesis of pleuromutilin derivatives as antibacterial agents using quantitative structure–activity relationship model. *International Journal of Molecular Sciences, 25*(4), 2256. https://doi.org/10.3390/ijms25042256