Personalized Cancer Specific Molecule Design using Deep Reinforcement Learning

Aljohara Hani Moaiteq Aljahdali
King Abdulaziz and His Companions Foundation for Giftedness and Creativity (Mawhiba)

Salma Elhag
King Abdulaziz University

ABSTRACT
Drug discovery remains a slow and costly process, limiting the rapid development of effective cancer therapies. This study presents a computational framework that applies Deep Reinforcement Learning (DRL) to generate novel molecules targeting the Epidermal Growth Factor Receptor (EGFR), a key cancer related protein. Bioactive compounds and molecular data were retrieved from ChEMBL and represented in Simplified Molecular Input Line Entry System (SMILES) format. Molecular descriptors were extracted using RDkit, and a DRL model (Proximal Policy Optimization) was trained to propose drug candidates optimized for EGFR binding. Generated molecules were evaluated through molecular docking using AutoDock Vina and Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) profiles were predicted to assess therapeutic suitability. The top candidate exhibited strong binding affinity (-8.9 kcal/mol), ideal Root Mean Square Deviation (RMSD) (0.0), and favorable druglike properties. Incorporating patient specific data, including mutation type, HLA profile, and disease stage further improved binding affinity, demonstrating the value of personalized molecule optimization. This work demonstrates the potential of AI guided approaches to accelerate early-stage cancer drug discovery and provides a foundation for integrating computational and experimental methods within precision oncology.

Keywords
DRL, EGFR, Personalized Medicine, Molecular Docking, ADMET, AI Drug Design, Precision Oncology.

1. INTRODUCTION
Global cancer drug development continues to face major challenges, including high costs, lengthy timelines, and frequent clinical trial failures. Traditional discovery approaches often depend on trial-and-error experimentation, slowing the delivery of effective therapies to patients. DRL offers a data driven framework capable of analyzing complex biological information to design therapeutic compounds efficiently. Its ability to learn optimal strategies in changing environments makes it well suited for addressing tumor diversity and variability in patient drug response. Unlike prior DRL models that focus on general molecular design, this study integrates patient specific features such as mutation type, HLA profile, and disease stage into reward shaping for molecular generation. The central hypothesis of this study is that incorporating patient level data into DRL guided molecule design enhances therapeutic specificity while minimizing off target toxicity. Therefore, this research addresses the question: "How can DRL be applied to generate novel EGFR targeting molecules optimized for personalized cancer therapy?". EGFR was selected as the biological target due to its pivotal role in the progression of lung, breast, and colorectal cancers and its contribution to therapeutic resistance. By integrating patient specific biological and clinical information, this work aims to accelerate early phase drug discovery and improve the precision of candidate selection.

The remainder of this paper is structed as follows: Section 2 presents a literature review, summarizing current advancements in DRL for drug design and cancer therapy. Section 3 outlines the proposed methodology, including data collection, compound filtering, molecular descriptor extraction, DRL model design, and patient specific optimization steps. Section 4 presents the results and discussion, including docking results, binding affinity scores, and ADMET evaluations of the generated molecules, emphasizing their potential effectiveness and relevance for personalized treatment. Finally, section 5 concludes by summarizing key findings, discussing the impact of DRL on personalized drug discovery, and outlining directions for future research, including experimental validation.

2. LITERATURE REVIEW
Recent advances in DRL have demonstrated its significant potential in optimizing cancer therapy and accelerating the process of drug discovery.

Engelhardt, D. (2020) [1] introduces CelluDose, a DRL framework designed to adaptively control emergent drug resistance in unpredictable cell populations. Unlike traditional dosing strategies, CelluDose is trained through stochastic simulations reflecting realistic mutation driven cell evolution. The DRL agent learned dosing policies that dynamically balanced efficacy and toxicity, achieving high suppression of harmful cells while maintaining minimal baseline dosing when resistance does not emerge. This study demonstrates how model free DRL guided by trajectory information for reward shaping can outperform traditional control methods in highly stochastic biological systems.

Gallagher, Kit, et al. (2023) [2] proposed a DRL based adaptive cancer therapy framework for treatment resistant prostate cancer. Rather than attempting to eliminate all tumor cells, their method maintained a stable population of drug sensitive cells to suppress resistant ones. DRL agents learned dosing strategies that extended treatment duration by up to 25% compared to existing clinical approaches. Despite its complex models, the approach produced simple interpretable rules based on tumor size and behavior, demonstrating adaptability to changing disease conditions.

Horwood and Noutahi (2020) [3] focused on improving molecular design by training DRL models to prioritize compounds based on both biological activity and synthetic accessibility. Their system guided molecule generation toward compounds that were not only bioactive but also synthetically feasible, a crucial factor in early-stage drug development where time and cost constraints are significant. This study highlighted the capacity of DRL to balance chemical creativity with real world feasibility.

Madondo et al. (2025) [4] developed a patient specific DRL system for automatic replanning of proton therapy in head and neck cancer. Incorporating individual patient data allowed the model to optimize treatment schedules, improving therapeutic efficacy and minimizing side effects. This study showcased the potential of DRL to personalize cancer therapy and move toward AI guided precision oncology.

The Mathematical Model Team (2024) [5] integrated mechanistic tumor growth modelling with DRL to guide adaptive therapy dosing decisions. Their hybrid system dynamically adjusted drug dosing based on real time tumor data aiming to maximize efficacy while reducing toxicity. The integration of mechanistic models with data driven reinforcement learning represented a significant step toward clinical decision support systems capable of personalized treatment.

Korshunova et al. (2022) [6] combined generative neural networks with reinforcement learning to optimize the de novo design of EGFR inhibitors. By employing strategies such as policy gradients and experience replay, the study addressed sparse reward issues commonly encountered in drug discovery. The resulting molecules demonstrated strong predicted activity and favorable druglike properties, reinforcing DRL's potential in cancer targeted molecular generation.

Liu et al. (2022) [7] introduced a DRL framework termed Proximal Policy Optimization Ranking (PPORank), which modelled treatment recommendation as a Markov Decision Process. Using patient specific clinical data, PPORank learned optimal ranking policies for drug selection and demonstrated superior performance over supervised learning methods in precision oncology applications.

Eckardt et al. (2021) [8] reviewed the application of reinforcement learning in precision oncology, highlighting its ability to model complex treatment scenarios ad adapt dynamically to patient response. Although they noted challenges such as limited clinical data and workflow integration, the review emphasized DRL's success in tumor modelling and adaptive planning.

Pandiyan et al. (2022) [9] provided a comprehensive review of AI approaches for anti-cancer drug discovery, including DRL techniques for de novo molecular design. They identified DRL's capability to efficiently explore chemical space and generate compounds optimized for pharmacological activity.

Li et al. (2024) [10] demonstrated the use of DRL in radiation therapy planning, automating complex treatment designs while preserving clinical accuracy. This study underscored the broader applicability of DRL in adaptive oncology decision making beyond drug discovery.

Mashayakhi et al. (2024) [11] developed a model free DRL framework for closed loop chemotherapy control. By dynamically adjusting drug dosing according to tumor response, the system improved treatment outcomes and minimized toxicity, suggesting DRL's potential to enhance clinical treatment scheduling.

Popova et al. (2018) [12] introduced ReleaSE, a hybrid DRL framework that combined generative and predictive models for de novo molecular design. The system successfully produced novel chemical structures optimized for biological activity, providing evidence that reinforcement learning can accelerate early-stage drug discovery.

Özçelik et al. (2025) [13] explored generative deep learning and DRL techniques for de novo drug design using graph transformer models. Their work demonstrated how reinforcement learning can refine molecular generation to preserve drug likeness while optimizing target specificity.

Wang et al. (2023) [14] analyzed AI frameworks, including DRL throughout the oncology drug development pipeline. Their findings indicated that reinforcement learning enhances both efficiency and compound selectivity.

Albani et al. (2025) [15] reviewed AI driven strategies in oncology drug discovery, showing that DRL integration can reduce experimental workloads and improve the prioritization of promising compounds.

Svensson et al. (2023) [16] proposed a comprehensive reinforcement learning framework for molecular generation, testing multiple algorithmic approaches to optimize chemical structures. The results demonstrated DRL's effectiveness in simultaneously improving predicted activity and drug likeness.

Ünlü et al. (2023) [17] developed DrugGEN, a graph transformer based generative adversarial network enhanced with reinforcement learning to design protein specific molecules. The generated compounds showed high predicted binding affinity, confirming the potential of DRL in structure-based drug design.

Le et al. (2025) [18] reviewed emerging AI applications in cancer drug discovery, emphasizing the role DRL in lead optimization and molecular prioritization. Their analysis highlighted AI's ability to streamline candidate selection and accelerate development.

Herráiz-Gil et al. (2025) [19] discussed reinforcement learning for drug repurposing and novel molecule generation, demonstrating its adaptability across multiple stages of drug development.

As Table 1 shows, prior research has demonstrated the potential of DRL in drug design and adaptive cancer therapy, but most studies focused on generalized simulations or model driven predictions. Few considered patient specific factors when generating molecules. This study builds on these efforts by applying DRL to design EGFR targeting compounds while incorporating individual patient characteristics to improve binding and predicted effectiveness.

Table 1. Comparative summary of previous DRL studies showing their aims, methods, limitations, and key results.

Study	Aim	Method	Limitation	Key Result
Engelhardt (2020)	Adaptive control of emergent drug resistance	DRL (Celludose) trained on random cell population models	Computational only, no experimental testing	High suppression of harmful cells with minimal dosing. Demonstrates DRL in stochastic biological systems.
Gallagher et al. (2023)	Adaptive therapy for treatment Resistant prostate cancer	Allegiant maintains population of sensitive cells	Focused on only prostate cancer, limited general use	Extended time to treatment failure by 25%
Horwood & Noutahi (2020)	Molecular design focused on chemical realism	DRL model with design limitations	Early stage with limited lab evaluation	Generated realistic and chemically valid drug like molecules
Madondo et al. (2025)	Personalized proton therapy replanning	DRL with patient specific data	Applied only to proton therapy, not yet clinically proven	Improved treatment timing and reduced potential side effects
Mathematical Model Team (2024)	Adaptive therapy dosing decisions	Biological tumor growth model combined with DRL	Simulation based, lacks real world validation	Merged biological modelling and RL for more personalized dosing
Korshunova et al. (2022)	De novo EGFR inhibitor design	Generative neural networks with RL	Based on computer models, not experimentally confirmed	Create new compounds with strong predicted activity
Liu et al. (2022)	Personalized cancer treatment recommendations	DRL (PPORank) using MDP	Needs larger clinical validation	Outperformed supervised learning for precision oncology
Eckardt et al. (2021)	RL in precision oncology	Review of RL applications	Limited use of real patient data	Highlighted the role of RL in adaptive and patient specific therapy
Pandiyan & Wang (2022)	AI approaches for anticancer drug discovery	Review including DRL for de Novo molecular design	Mostly computational, minimal experimental testing	Showed that DRL can explore chemical spaces efficiently
Li et al. (2024)	Improving radiation therapy planning	DRL for dosing	Focused on radiation therapy only	Automated planning while keeping clinical accuracy
Mashayekhi et al. (2024)	Closed loop chemotherapy control	Model free DRL	Needs clinical validation	Improved treatment timing and reduced toxicity in simulations
Popova et al. (2018)	De novo molecular design	DRL (ReleaSE) Combining generative and predictive models	Fully computational approach	Generated molecules with strong activity and druglike properties
Özçelik et al. (2025)	Generative deep learning in drug design	Graph transformer with RL	Results limited to computer testing	Explored large chemical space while keeping drug properties stable
Wang et al. (2023)	AI frameworks for oncology drug development	Review of AI and DRL across the drug development process	General review, few detailed case examples	Showed DRL can improve drug activity and selectivity
Albani et al. (2025)	AI driven strategies in oncology drug	Review including DRL	Theoretical focus, not tested experimentally	Demonstrated higher efficiency and reduced workload in early research
Svensson et al. (2023)	Molecular generation using RL	Multiple RL approaches for molecule design	Based only on simulations, lacks lab testing	Generated molecules meeting multiple

				design criteria
Ünlü et al. (2023)	Target specific molecule design	Graph transformer GAN with RL	Computational predictions only	Designed high affinity molecules for protein targets
Le et al. (2025)	AI in cancer drug discovery	Review of AI and DRL techniques	Broad overview, limited experimental data	Showed how AI can speed up identification of potential drugs
Herráiz-Gil et al. (2025)	AI assisted drug repurposing	RL guided molecule generation	Computational predictions, not yet validated	Suggested new possible uses for existing drugs

Although DRL has shown success in molecular design and adaptive therapy, few studies integrate patient-specific data such as mutation profiles or disease stage. This study addresses this gap by designing EGFR-targeting molecules that incorporate individual patient characteristics to improve binding affinity and drug-like properties.

3. Methodology

This section will present the methodology steps used in the study.

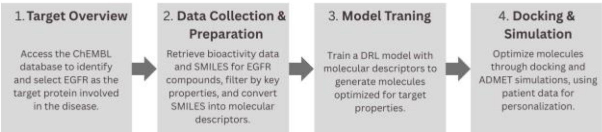

Figure 1. Workflow of the EGFR targeted drug design methodology showing the steps from target identification and data collection to DRL model training and personalized docking.

3.1 Target Identification

The target selected for this study was EGFR, a cell surface protein known to play a critical role in cancer cell growth and survival. EGFR was chosen due to its importance in cancer progression and its involvement in drug resistance across multiple tumor types. Information about this target was obtained from the ChEMBL database, a well-established resource containing structured data on bioactive molecules. The target was identified using its specific ChEMBL ID: CHEMBL4523998.

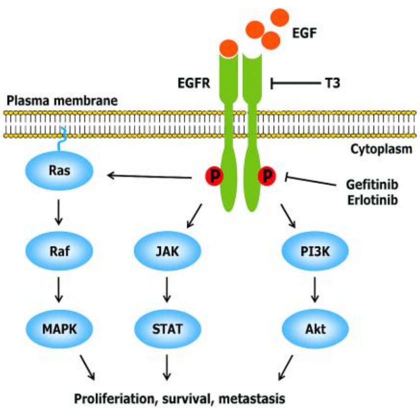

Figure 2. EGFR structure and signaling pathway involved in cancer cell growth and survival.

3.2 Compound and Bioactivity Retrieval

After selecting the target, a search was conducted in the ChEMBL database to collect compounds associated with EGFR. For each compound its SMILES representation and its binding strength data such as IC_{50} were retrieved. IC_{50} refers to the concentration of a compound required to inhibit 50% of the target activity, with lower values indicating stronger inhibition. These data formed as the foundation for analyzing compound target interactions and were used to train the molecule generation model.

3.3 Compound Filtering

At this stage, the extracted compounds were screened using predefined research-based standards. Compounds with weak binding affinity or limited experimental evidence were excluded. The section chose molecules demonstrating strong activity, suitable druglike properties, and reliable experimental validation, ensuring the dataset contained only potent and structurally appropriate compounds.

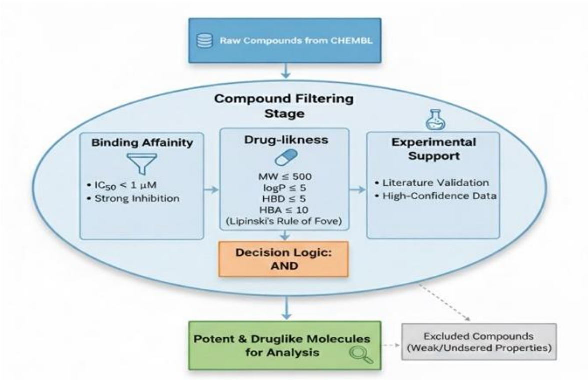

Figure 3. Compound filtering workflow based on binding affinity and drug likeness requirements to select potent molecules for further analysis.

3.4 Feature Extraction

In this phase the SMILES strings of the selected compounds were converted into molecular descriptors using cheminformatics libraries such as RDKit. These descriptors are calculated features that represent chemical and structural characteristics including molecular weight, shape, and hydrogen bonding capacity. They served as numerical input features for model training.

3.5 DRL Based Molecule Generation

During this step, a DRL model was trained to design new molecules capable of effectively binding to EGFR. DRL combines reinforcement learning with deep neural networks to optimize decision making processes. The model used the molecular descriptors from the filtered compounds as input. It received rewards for molecules predicted to have strong binding affinity, favorable druglike properties, and low toxicity as shown in the equation below.

Reward=α×Binding Affinity+β×Drug Likeness − γ × Toxicity

Equation 1. reward function in DRL assigning higher scores to molecules that bind EGFR strongly, have good druglike properties, and show low toxicity.

Patient specific data such as mutation type and cancer stage were incorporated to enhance therapeutic relevance. Training was performed in python using RDKit and a DRL framework designed for molecular generation. Generated molecules were sorted as SMILES and their predicted pIC_{50} values were calculated for prioritization. (pIC_{50} is the negative logarithm of the IC_{50} value where higher pIC_{50} indicates greater inhibitory potency and makes data easier to compare.)

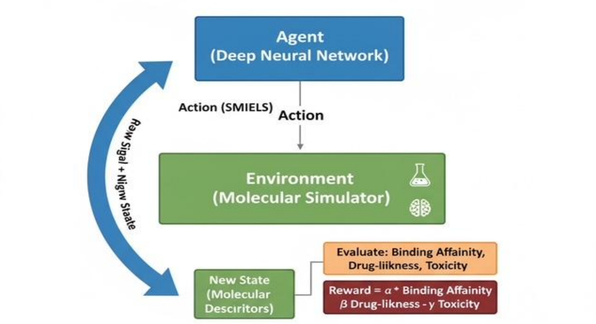

Figure 4. A basic overview of DRL showing how the model learns by interacting with an environment and receiving rewards to improve its decisions.

3.6 ADMET Optimization

At this stage the generated molecules were evaluated for their ADMET profiles using predictive computational tools. The goal was to optimize drug likeness and minimize potential toxicity. When possible, patient specific data were incorporated to enhance the biological relevance of these predictions.

3.7 Molecule Docking and Simulation

In this final phase molecular docking simulations (e.g. using AutoDock Vina) were followed by molecular dynamics to validate binding stability and behavior under physiological conditions. The predicted pIC_{50} values were used to assess the relative potency and efficacy of the generated compounds., a scale representing the negative logarithm of IC_{50} (half maximal inhibitory concentration) where higher pIC_{50} values indicate more potent inhibition making data easier to interpret and compare. Lipiniski's Rule of Five then evaluated the oral bioavailability of the compounds: optimal candidates typically have molecular weight ≤ 500 Da, ≤ 5 hydrogen bond donors, ≤ 10 acceptors, and a log P ≤ 5. Finally, ADMET properties were predicted using computational tools to ensure the molecules exhibit favorable pharmacokinetics and safety profiles before experimental validation.

4. RESULTS AND DISCUSSION

The generated molecules demonstrated strong predicted binding affinity toward EGFR and exhibited favorable docking conformations as summarized in table 2. These results indicate that the designed compounds were able to effectively occupy the active binding pocket of the target receptor with stable orientations.

Table 2. Binding affinity score and RMSD values of top candidate molecules.

	Binding Affinity Score (kcal/mol)	Ds form (RMSD)	Best Mode (RMSD)
1	-8.9	0.0	0.0
2	-8.3	1.234	2.50
3	-7.3	2.110	3.098

Molecule 1 exhibited the highest binding affinity (-8.9 kcal/mol) and an ideal RMSD of 0, indicating the most stable docking conformation. Molecule 2 and 3 also showed favorable binding scores (-8.3 kcal/mol) and (-7.3 kcal/mol) with slightly higher RMSD values that still reflect acceptable docking precision.

Figure 5 compares the predicted binding affinity and RMSD values, Ds form and Best Mode for the three generated molecules. Molecule 1 shows the most favorable profile with the lowest binding affinity (-8.9 kcal/mol) and minimal RMSD, indicating a highly stable and well oriented binding orientation within the EGFR active site. Molecules 2 and 3 exhibit slightly weaker binding and higher RMSD confirms that stronger binding corresponds to more precise and consistent molecular positioning during docking.

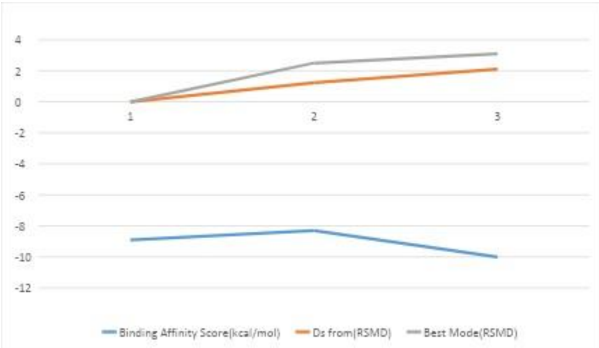

Figure 5. Binding affinity score and RMSD for a three generated molecule.

We analyzed the SMILES and IC50 values of the molecules supporting their biological activity and potential effectiveness.

Table 3. SMILES and IC50 values of the top candidates.

SMILES	IC50 (nM)
CC1=CC=C(C=C1)NC(=O)C2=CC=CC=C2	154.6546234527607
COC1=CC=C(C=C1)C(=O)N2CCCCC2	77.64416595447096
CN(C)CCOC1=CC=CC=C1C	50.70760441055491
C1=CC=C(C=C1)C(CNC2=CC=CC=C2)=O	97.1988046319471
CC(C)(C)NC(=O)C1=CC=CC=C1	176.05255988630864
COC1=CC=CC=C1OC	201.39989974418162
CCC1=CC=CC=C1NC(=O)C	271.7142833316617
CCOC(=O)C1=CC=CC=C1Br	212.50128462963093
CN(C)C1=CC=CC=C1	344.18365269639594
C1=CC=CC=C1C(=O)NC	106.20386361710113

To further refine the generated molecules, patient specific data was incorporated. This personalized optimization influenced predicted binding affinity, ADMET properties, and RMSD, ensuring that molecules were tailored for individual therapeutic contexts.

Patient_ID: P001

Tumor_Type: NSCLC

Mutation: EGFR L858R + PD-L1 overexpression

HLA_Type: HLA-A*02:01

Preferred_Binder: PD1_MOL_2

Therapeutic_Suggestion: Immune checkpoint inhibitor + targeted therapy

Figure 6. Patient data was used for personalized drug optimization.

Figure 7 present toxicity levels and Lipinski compliance for the top candidates confirming that the molecules meet drug likeness criteria well minimizing predicted toxicity. Figure 8 shows the distribution of binding affinity relative to hepatotoxicity risk indicating that molecules with higher binding affinity generally maintain acceptable liver toxicity profiles. Figure 9 correlates binding affinity with patient age, treatment response, and disease stage illustrating the impact of personalization on predicted molecular efficacy.

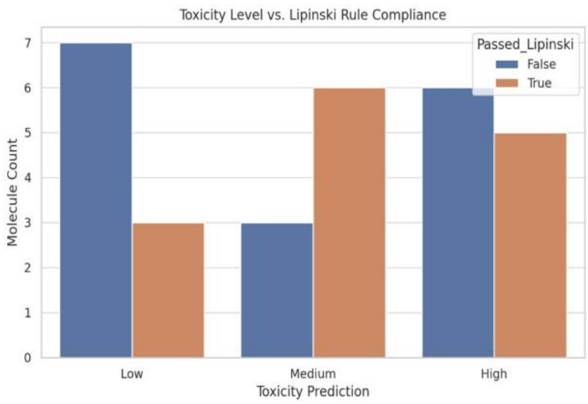

Figure 7. Toxicity levels Lipinski compliance.

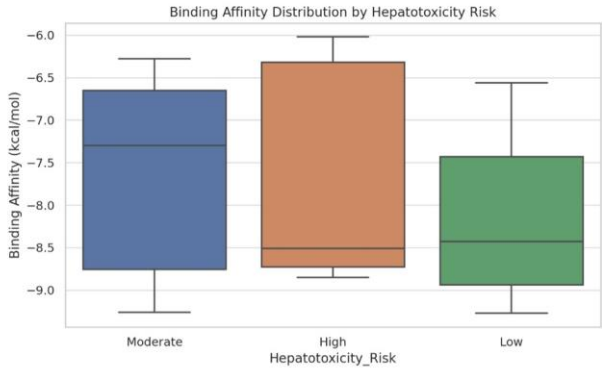

Figure 8. Binding affinity distribution by hepatotoxicity risk.

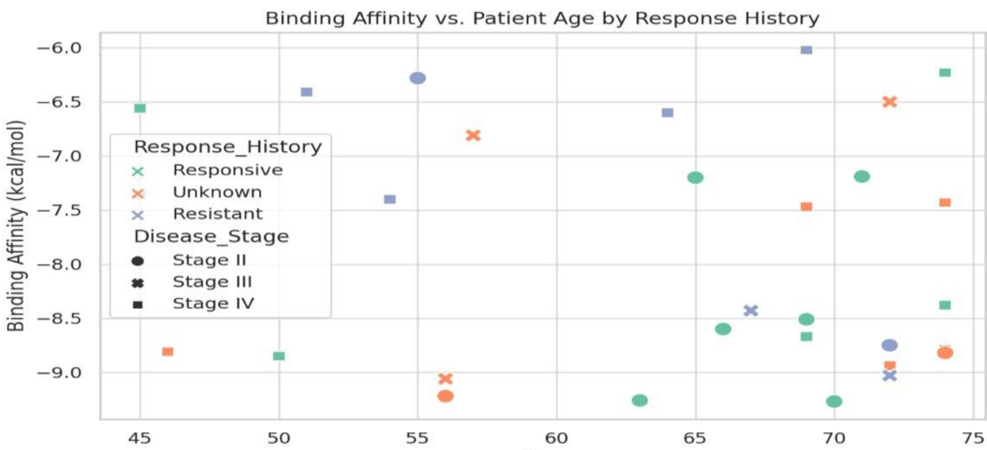

Figure 9. Binding affinity vs. patient age by response history and disease stage.

5. CONCLUSION

The findings of this study demonstrate that DRL can effectively generate novel molecules targeting EGFR achieving high predicted binding affinity and favorable biochemical profiles. The optimized compounds exhibited strong and stable binding orientations, met drug likeness criteria, and showed low predicted toxicity indicating their potential as promising leads for further cancer drug development. Incorporating patient specific information such as mutation type, HLA profile, and disease stage, enhanced molecular binding precision and emphasized the importance of personalization in therapeutic design. This research provides evidence that AI guided molecular generation can significantly accelerate early phase drug discovery and improve the specificity of cancer therapies. Future work should include lab experimental validation to test the ability of designed molecules to inhibit EGFR activity and to verify computational predictions. Expanding the model to include larger and more diverse datasets, additional cancer targets, and broader molecular descriptors would further strengthen its predictive capability. Additionally integrating multi objective optimization methods could enable simultaneous improvements of efficacy, safety, and synthesis feasibility paving the way for fully automated patient tailored drug discovery.

6. ACKNOWLEDGMENTS

This research was supported by the King Abdulaziz and His Companions Foundation for Giftedness and Creativity (Mawhiba). Sincere thanks to my school Al Andalus Mawhiba Al Shati Branch, for their continuous support, encouragement, and for providing a positive environment that made this research possible.

7. REFERENCES

[1] D. Engelhardt, J. Mach. Learn. Res., 21(203), 1–30, 2020.

[2] K. Gallagher et al., bioRxiv, 2023-04, 2023.

[3] J. Horwood and E. Noutahi, ACS Omega, 5(51), 32984–32994, 2020.

[4] M. Madondo et al., arXiv preprint, arXiv:2506.10073, 2025.

[5] K. Gallagher et al., Cancer Res., 84(11), 1929–1941, 2024.

[6] M. Korshunova et al., Commun. Chem., 5(1), 129, 2022.

[7] M. Liu, X. Shen, and W. Pan, Stat. Med., 41(20), 4034–4056, 2022.

[8] J. N. Eckardt et al., Cancers, 13(18), 4624, 2021.

[9] S. Pandiyan and L. Wang, Comput. Biol. Med., 150, 106140, 2022.

[10] C. Li et al., Phys. Med., 125, 104498, 2024.

[11] H. Mashayekhi et al., Comput. Methods Programs Biomed., 243, 107884, 2024.

[12] M. Popova, O. Isayev, and A. Tropsha, Sci. Adv., 4(7), eaap7885, 2018.

[13] R. Özçelik et al., J. Chem. Inf. Model., 65(14), 7352–7372, 2025.

[14] L. Wang et al., Pharmaceuticals, 16(2), 253, 2023.

[15] F. G. Albani et al., Drug Des. Dev. Ther., 5685–5707, 2025.

[16] H. G. Svensson et al., Mach. Learn., 113(7), 4811–4843, 2024.

[17] A. Ünlü et al., Nat. Mach. Intell., 1–17, 2025.

[18] M. H. N. Le et al., Biochim. Biophys. Acta, 167680, 2025.

[19] S. Herráiz-Gil et al., Appl. Sci., 15(5), 2798, 2025.

[20] Takahiro Eitsuka, Naoto Tatewaki, Hiroshi Nishida, Kiyotaka Nakagawa, and Teruo Miyazawa. 2016. *Synergistic anticancer effect of tocotrienol combined with chemotherapeutic agents or dietary components: A review.* *International Journal of Molecular Sciences* 17, 10 (2016), 1605.