ACS OMEGA

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

Article

Machine Learning-Assisted Prediction of the Biological Activity of Aromatase Inhibitors and Data Mining to Explore Similar Compounds

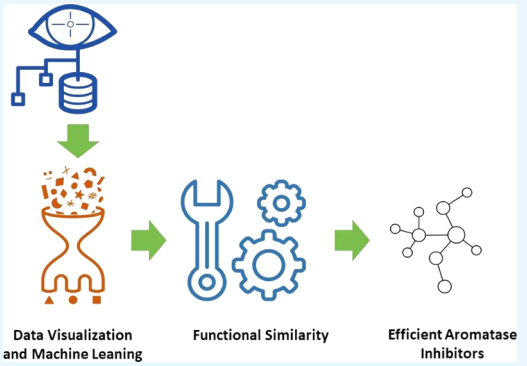
Muhammad Ishfaq,* Muhammad Aamir,* Farooq Ahmad, Abdelazim M Mebed, and Sayed Elshahat

Cite This: *ACS Omega* 2022, 7, 48139−48149

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Designing molecules for drugs has been a hot topic for many decades. However, it is hard and expensive to find a new molecule. Thus, the cost of the final drug is also increased. Machine learning can provide the fastest way to predict the biological activity of druglike molecules. In the present work, machine learning models are trained for the prediction of the biological activity of aromatase inhibitors. Data was collected from the literature. Molecular descriptors are calculated to be used as independent features for model training. The results showed that the R^2 values for linear regression, random forest regression, gradient boosting regression, and bagging regression are 0.58, 0.84, 0.77, and 0.80, respectively. Using these models, it is possible to predict the activity of new molecules in a short period of time and at a reasonable cost. Furthermore, Tanimoto similarity is used for similarity analysis, as well as a chemical database is mined to search for similar molecules. Nonetheless, this study provides a framework for repurposing other effective drug molecules to prevent cancer.



1. INTRODUCTION

Recently, aromatase inhibitors have gained considerable attention in the field of drug design and other pharmacological applications because of several distinctive characteristics, such as high enzyme specificity, prolonged inhibitory action, and minimal toxicological effects.[1] They have been developed to exhibit competitive, mechanism-based, and irreversible kind of inhibition in various pathologies such as breast cancer.[2] The number of breast cancer patients is increasing in many countries, leading to an economic burden. One-third of carcinomas are found to be hormone-dependent where the cell proliferation is affected directly by the estrogen hormones. Either targeting the estrogen receptor directly (first) or inhibiting the aromatase (second) activity are the two approaches that have been used to control or block the tumor progression of the said hormone.[3] An important approach to reducing tumor growth is the inhibition of the enzyme aromatase (the key enzyme (CYP19) for estrogen biosynthesis), a member of subfamilies of cytochrome P450s (a family containing more than 60 important metabolizing enzymes).[4] It is responsible for the catalytic conversion of androgen to estrogen, where the reaction is progressed in the active sites of this enzyme using ferric ions in haem.[5] Considered to be the mainstream treatment method for the estrogen receptor-positive breast cancer treatment regimen, aromatase inhibitors have been continuously used under the title of first-, second-, and third-generation aromatase inhibitors as approved by the FDA. The third-generation inhibitors comprise of letrozole, anastrozole, and exemestanea, which are used in the standard treatment of postmenopausal breast cancer these days.[6,7] In many studies, their use has been reported in reproductive technology,[8] endometriosis,[9] gynecomastia,[10] ovarian cancers,[11] male infertility,[12] and many others.[13]

Although the treatment using both steroidal and nonsteroidal aromatase inhibitors as third-generation aromatase inhibitors has gained tremendous attention, a few major side effects such as arthralgia, myalgia, hot flashes, night sweats, loss of sex drive, and vaginal dryness were observed in case of their prolonged clinical usage.[14] Furthermore, the situation gets worse in patients with liver, kidney, or adrenal insufficiency, leading to excessive hair loss.[15] Therefore, it is urgent to develop a new kind of efficient aromatase inhibitor with minimum side effects. Therefore, it is important for the researchers to investigate some more structural properties of these enzymes to get a better understanding of the quantitative structure−activity relationships (QSARs) to open horizons for new drug discovery. Machine learning seems to be famous in biological science.[16] Quantitative structure−activity relation-

Received: September 24, 2022
Accepted: December 2, 2022
Published: December 13, 2022

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

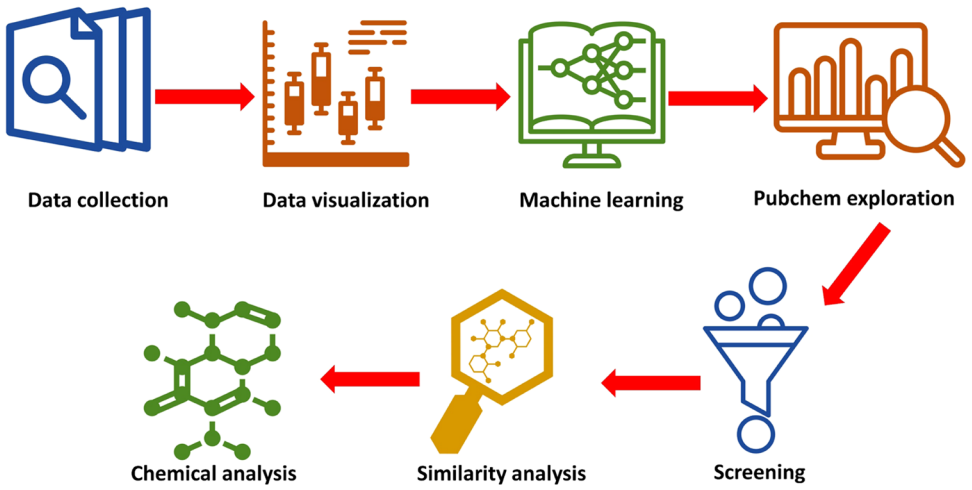


Figure 1. General framework of the current study.

ship (QSAR) is considered one among these. Compared with density functional theory calculations, machine learning requires much less time for prediction.[17−19]

Virtual design and prediction of single- or multitarget inhibitors of cancer-related proteins, including aromatase is also a hot topic.[20] A molecule able to simultaneously inhibit many different cancer cells will be better than a molecule that inhibits a few cancer cells.[21] In addition, the heterogeneity of the tumor is also a well-established fact and varies from person to person and lays the foundation of precision medicine. Of note, during the drug discovery process, we always move from one common target to a complicated multitarget strategy. Therefore, in this preliminary study, we focus on one target rather than multiple targets for the sake of simplicity.

The molecular docking studies were quite important and helpful in understanding the structural and functional properties of human aromatase because the information about the three-dimensional (3D) structure of this enzyme was also a result of its molecular docking studies.[22] Its three-dimensional structural information was reported based on a hypothetical theoretical 3D model of the enzyme aromatase.[23] 3D-QSAR studies on nonsteroidal aromatases as aromatase inhibitors to analyze their synthesis, structural features, and inhibitory activities have been performed.[3,24]

Lone et al. synthesized novel testololactam and testolactam (nitrogen congeners), whose structural and electronic properties were studied by theoretical density functional theory (DFT) studies. Although the computational and molecular docking studies predicted a relatively lower therapeutic efficacy, they could appreciably be used as steroidal aromatase inhibitors.[25] Another study was carried out by Banjare et al., who utilized structure-guided molecular docking-assisted alignment-dependent three-dimensional QSAR to analyze a set of 22 compounds to search novel, less toxic, and potent molecules.[26] The compounds having aromatase inhibitory activity were studied for antibreast cancer properties. In another study, steroidal aromatase inhibitors were evaluated using docking studies to rationalize the quantitative structure−activity relationships.[27] Recently, Giampietro et al. performed computational studies to design and prepare novel phenyldiazenyl sulfonamides and provided a sound rationale at a molecular level.[28] In addition, Osmaniye et al. used molecular docking and molecular dynamic studies to design and synthesize novel furan or thiophene ring containing triazolothiazine derivatives, which could have been used as anticancer agents.[29] Moreover, these computational studies are important to save time and resources, whereby the expected outcomes are predicted theoretically rather than directly engaging with the experimental complications.[30,31] Machine learning as a subfield of computer science and statistics provides a platform for strong artificial intelligence and optimization concepts (delivering methods, theory, and domain of a wide range of applications), with the main focus on providing data to improve patient outcomes.[32−34]

In the present work, multiple machine learning models have been trained for the prediction of the biological activity of aromatase inhibitors. Molecules from PubChem are extracted and their biological activity is predicted through machine learning models. In addition, the ChEMBL database is explored to find similar molecules using RDkit. The framework of the present study is given in Figure 1.

2. METHODOLOGY

2.1. Data Collection. The data for machine learning is collected from research papers and contains more than 400 data points. The smiles of molecules, aromatase inhibitor activity, and DOI of papers from where the data has been collected constitute the data. The acquired data is given in Table S1.

2.2. Molecular Descriptor Calculation. Various types of molecular descriptors for molecules are calculated using Dragon software.[35] The 3D geometries of compounds in structure data file (SDF) format are used as input. About 4000 descriptors are generated. These descriptors are exported in comma-separated values (.csv) file. Best descriptors are shortlisted using univariate regression. These descriptors are used for training machine learning models.

2.3. Training the Model. We have imported the necessary packages of Python such Scikit-leran, Pandas, Scipy, Numpy, Seaborn, and Matplotlib. These packages are necessary for data visualization and analysis. The molecular descriptors and biological activity in the comma-separated values (.csv) file are imported with the help of the Pandas module. Linear regression, random forest regression, gradient boosting regression, and bagging regression are used for machine learning analysis. The linear regression model predicts the target variable by analyzing the relationship between the target variable and independent variables. The random forest model

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

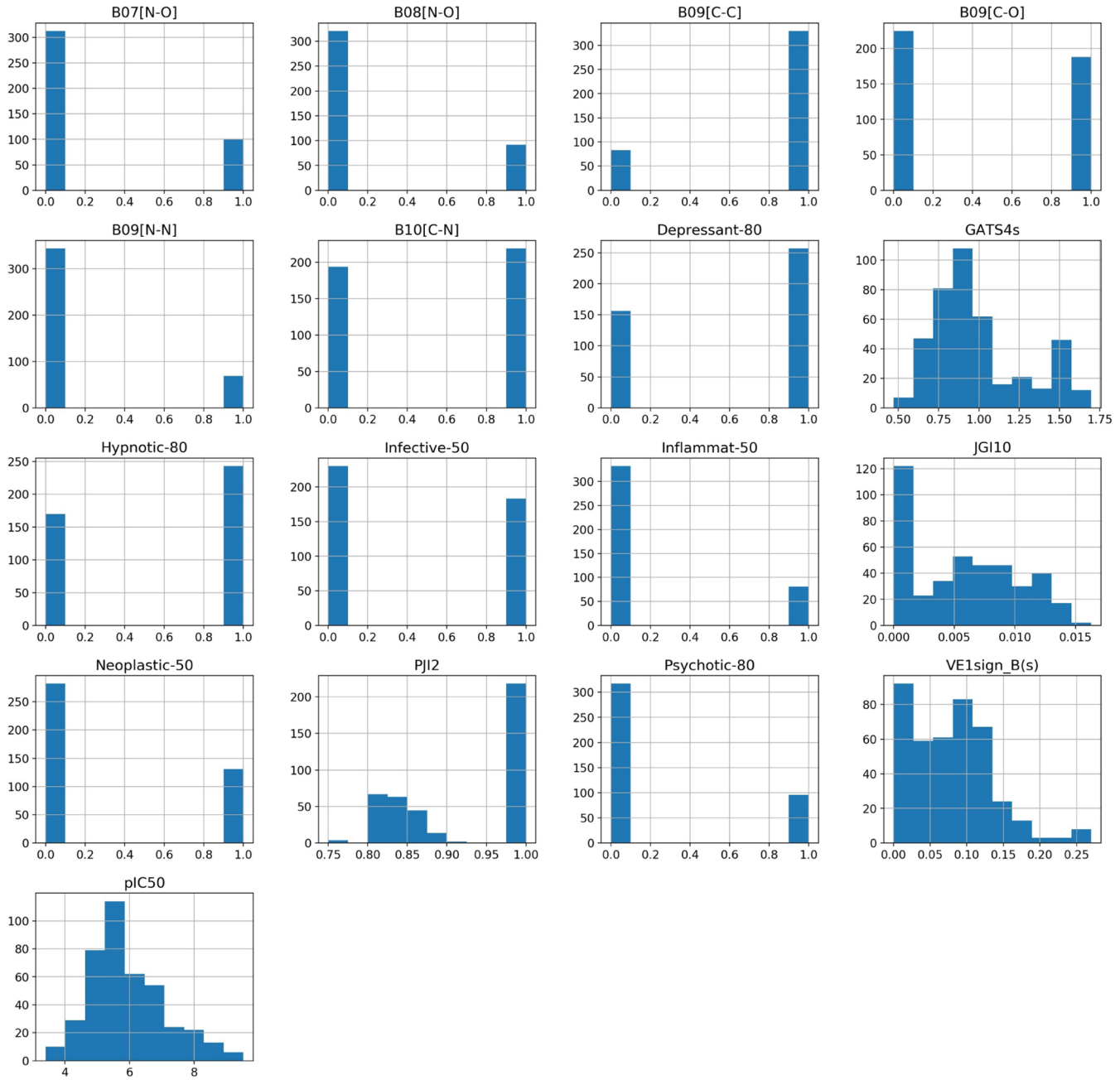


Figure 2. Distribution plot of descriptors (features) and dependent variable (pIC_{50}).

uses multiple decision trees to make a prediction. The results from individual trees are averaged to provide output predictions from the whole forest. The gradient boosting model also uses multiple decision trees. Compared to random forests, it builds relatively simple trees, which are sequentially incorporated into the ensemble. Bagging regression consists of two parts: bootstrapping and aggregation. In bootstrapping, multiple subsets are derived from the whole data set using the replacement procedure. In aggregation, all possible outcomes of the prediction are combined. The cross_val_score function of Scikit-leran is used for cross-validation. The GirdSearchCV library in Scikit-leran is used to tune hyperparameters.

2.4. Similarity Analysis. Similarity analysis is performed using RDKit, which is a cheminformatics software.[36] Many types of operations can be performed on chemical compounds using this software. Similarity analysis is a straightforward method to find similarities between reference structure and structure in the database.[37,38] For this purpose, pharmacophores, distances, fingerprints, etc. can be used. In our work, Tanimoto similarity is used. For this purpose, extended connectivity fingerprints (ECFP4) are selected. RDkit compares the fingerprint of the query structure (reference structure) with the fingerprints of each compound within the database and calculates the Tanimoto index.

3. RESULTS AND DISCUSSION

3.1. Molecular Descriptors. The chemical structures of the molecules determine their role in various applications.[39−41] Machine learning through molecular descriptors is a good way to link the chemical structure of molecules with biological activities. Molecular descriptors are calculated to feed the machine learning models.[42] These descriptors are easy and fast

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

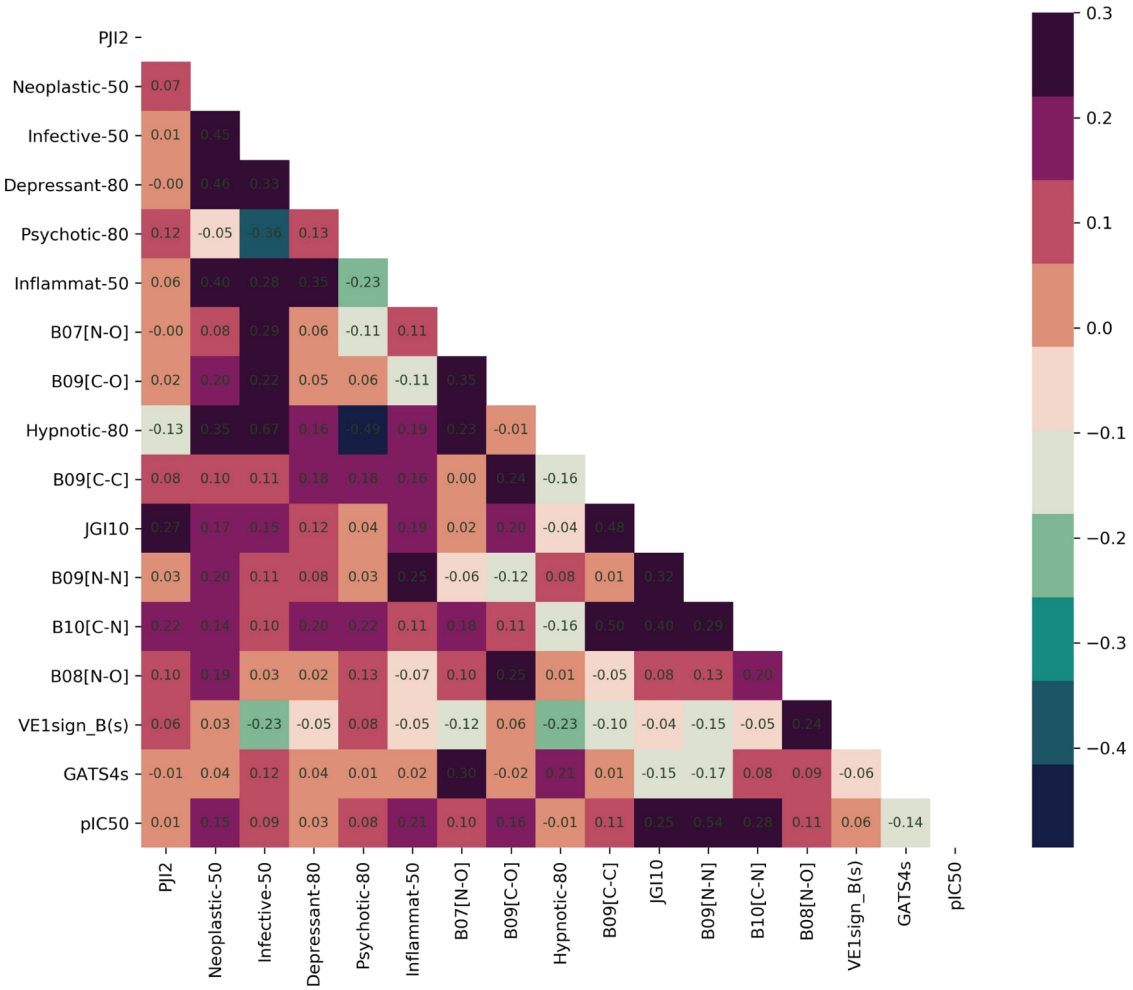


Figure 3. Heatmap of the Pearson correlation between descriptors and pIC_{50}.

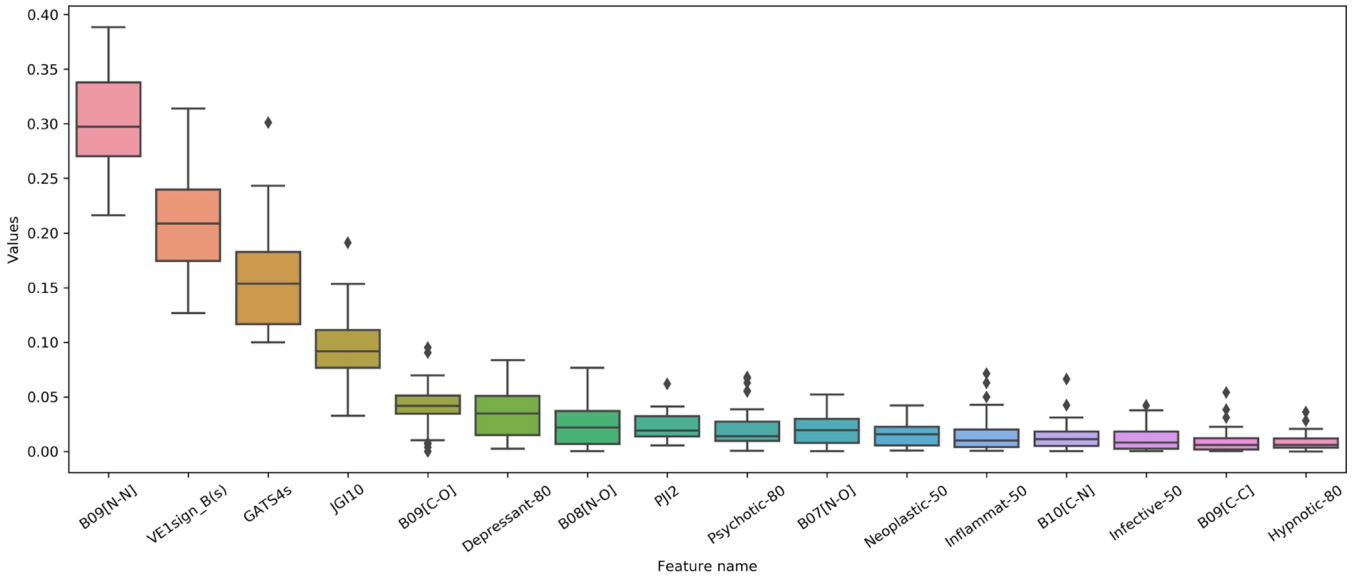


Figure 4. Feature importance calculated using the random forest model.

to calculate compared with quantum chemical descriptors.[43−45] The distribution plots of descriptors and pIC_{50} are given in Figure 2. Many descriptors have two types of values: 0 and 1.

The Pearson correlation between different parameters is calculated and their heatmap is plotted. The obtained graph is given in Figure 3. The correlation between different parameters is not high. The role of different descriptors in model training

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

Table 1. RMSE and R^2 Values for Different Machine Learning Models

model	RMSE	R^2
linear regression	2.07	0.71
random forest regression	1.14	0.93
gradient boosting regression	1.76	0.85
bagging regression	1.45	0.91

is determined using feature importance. It is done using the random forest model. The majority of descriptors have less importance (Figure 4). B09[N−N] is the most important descriptor. It is a topological distance descriptor. B09[N−N] encodes the presence or absence of nitrogen atoms with a topological distance of 9. VE1sign_B(s) is the second important descriptor. It is a two-dimensional (2D) matrix-based descriptor. VE1sign_B(s) represents the coefficient sum of the last eigenvector from the Burden matrix weighed by I-State. GATS4s is the third important descriptor. It is a 2D autocorrelation-based descriptor. GATS4s represent the Geary autocorrelation of lag 4 weighed by I-state.

3.2. Regression Analysis. Classification and regression are two important categories of machine learning. In classification, the data set is divided into predefined groups. The range of a group controls the classification accuracy.[46−48] Classification only predicts the group in which the biological activity of a particular molecule will fall. To predict the biological activity value of a molecule, regression analysis is performed. For this purpose, multiple regressors are used. Various machine learning models have been tried. A 10-fold CV shows higher performance. Table 1 presents the performance parameters of different models, including the root mean square errors (RMSEs) and the r-square values. It is clear that the random forest regressor and the bagging regressor showed higher performance. The hyperparameters of these models were optimized. The accurate prediction can decrease the dependence on expensive experimental methods.[49−52] The scatter plot between true and predicted values for different models are given in Figure 5. Several approaches are reported in the literature to check the reliability of machine learning models.[53−55] We have checked the reliability of machine learning models using the prediction on an external data set.

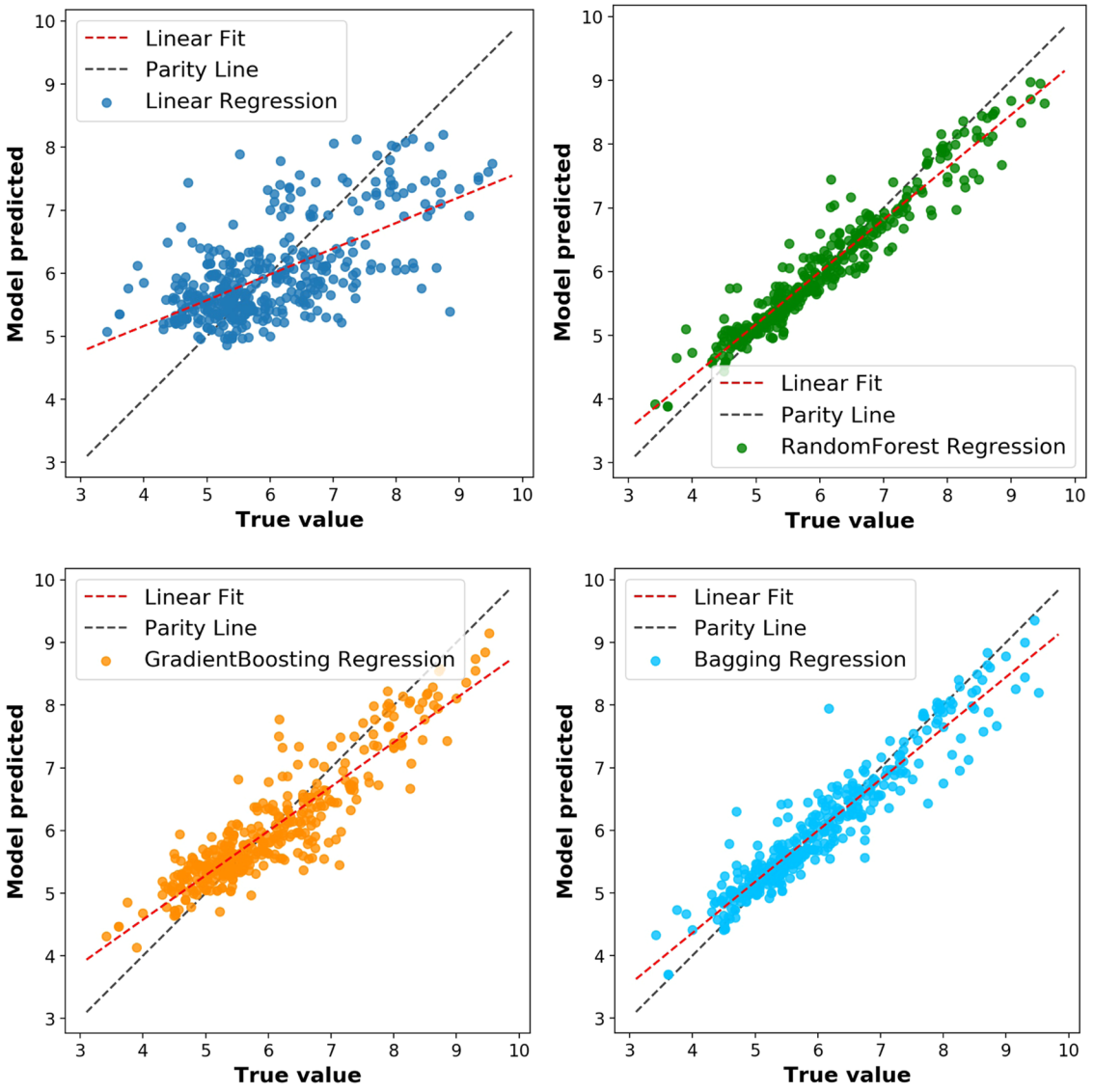


Figure 5. Scatter plot between true and predicted values (pIC_{50}).

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

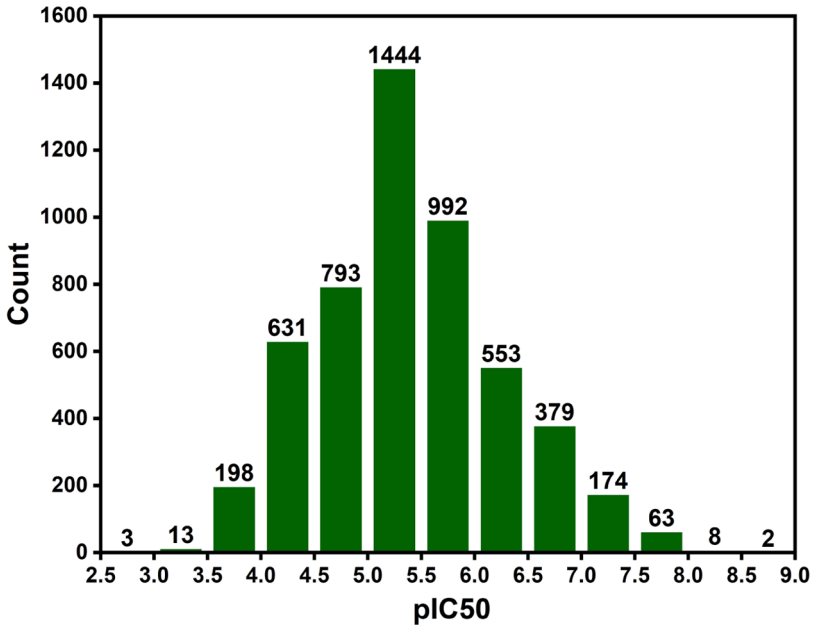


Figure 6. Distribution of the predicted pIC_{50}.

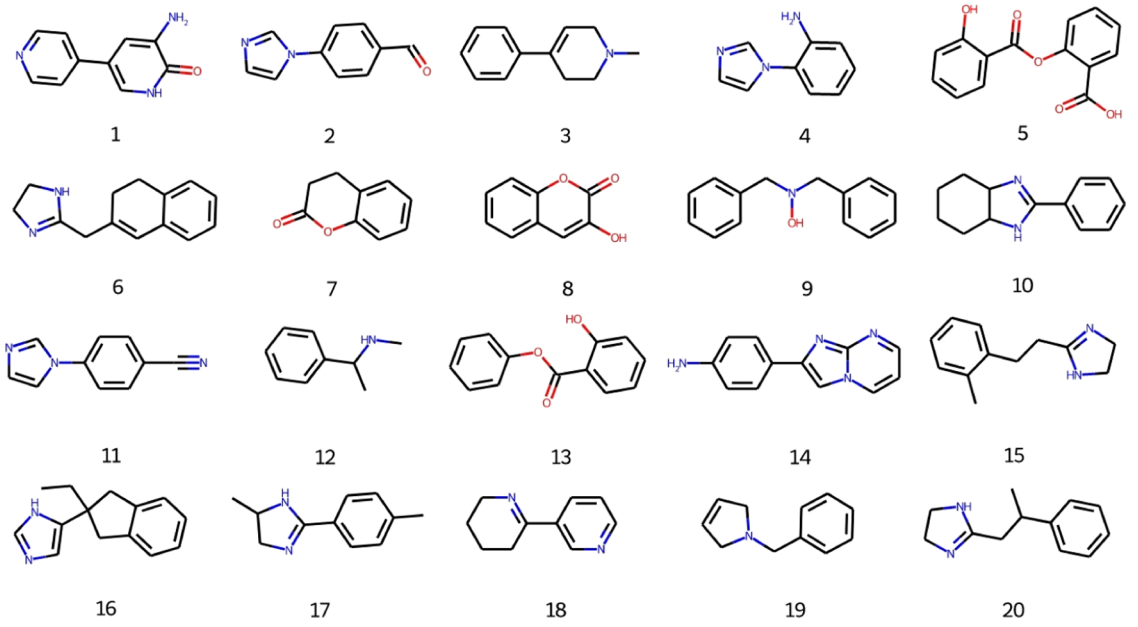


Figure 7. Top 20 molecules from the PubChem database.

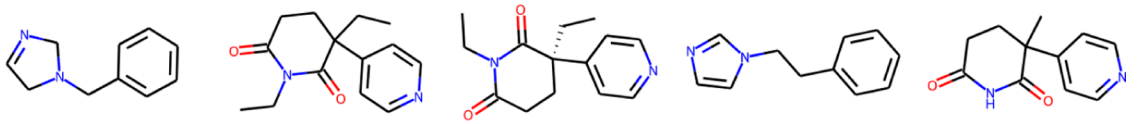


Figure 8. Top five molecules in the training set.

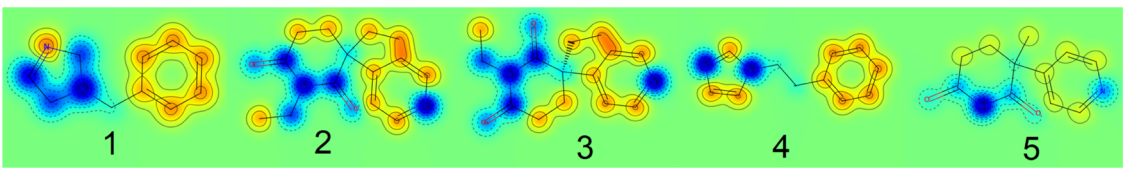


Figure 9. Gasteiger atomic charges of the top five molecules from the collected data (training set).

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

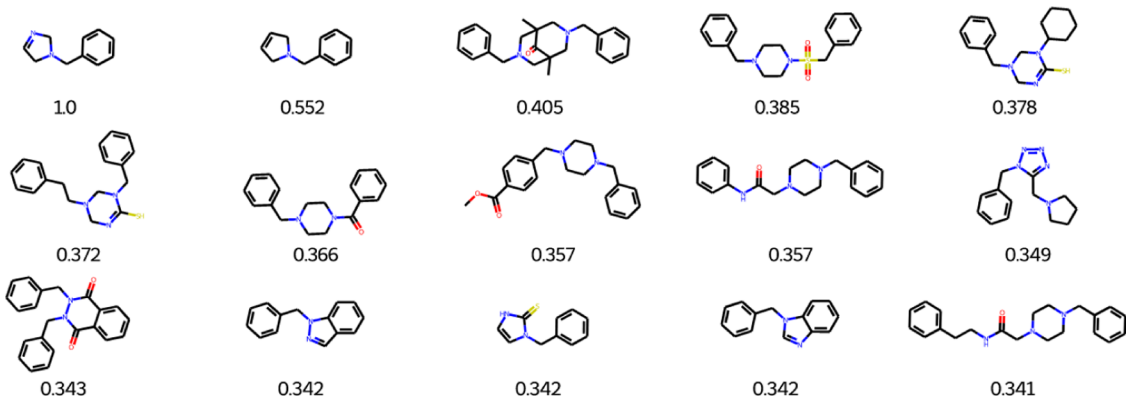


Figure 10. Top 14 molecules similar to compound 1.

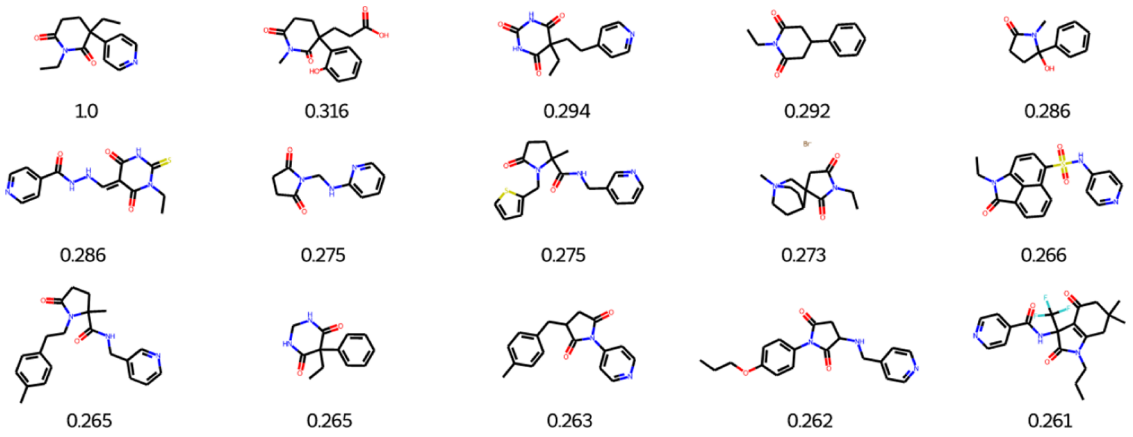


Figure 11. Top 14 molecules similar to compound 2.

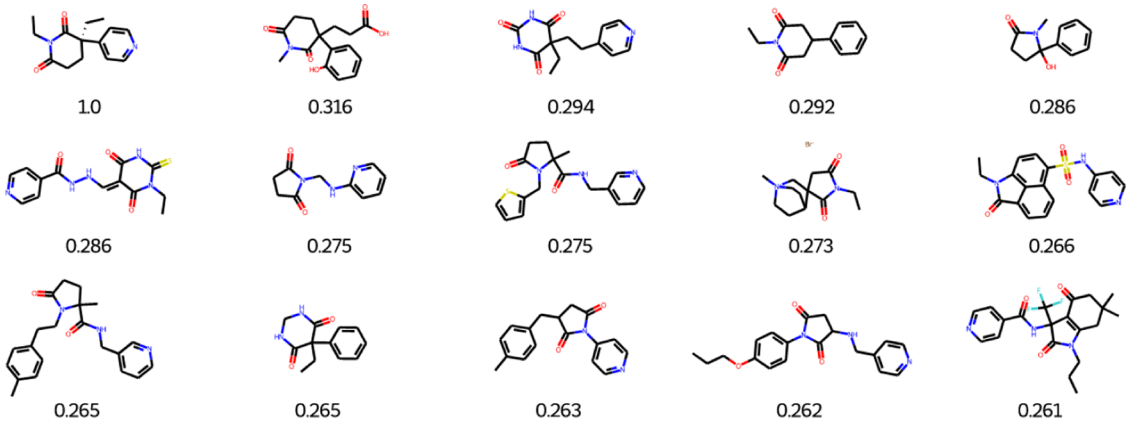


Figure 12. Top 14 molecules similar to compound 3.

The data collected for the external validation is given in Table S3 and is not part of training and test sets. Linear regression has shown the lowest value.1

More than 5000 molecules have been extracted from PubChem, which is a free chemical repository of small organic molecules.[56] It is maintained by the National Library of Medicine. The biological activity of the extracted molecules is predicted using the already trained random forest model. The distribution of the predicted pIC_{50} values is given in Figure 6. The best molecules are shortlisted. The top 20 molecules are given in Figure 7.

3.3. Similarity Analysis. A similarity analysis based on the chemical structure is a useful method to identify potential compounds in drug discovery. It is because two molecules with similar structures are likely to show similar bioactivities.[57] However, exceptions also cannot be ignored.[58] Once a lead compound has been found, a series of structural analogues also can be designed. In the present study, the five best molecules (with the lowest pIC_{50} values) from the training set are selected. These structures are given in Figure 8. These molecules are selected individually as a reference to search for similar compounds. The ChEMBL database is used to find similar compounds. The database is managed by EMBL's European Bioinformatics Institute.[59]

Gasteiger charges are simple and fast to compute, requiring only the knowledge of the topology of a molecule. The blue

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

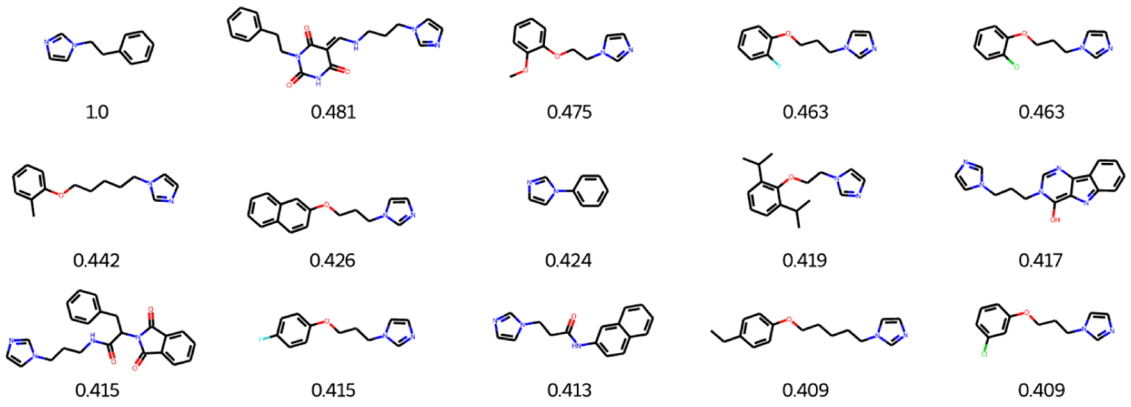


Figure 13. Top 14 molecules similar to compound 4.

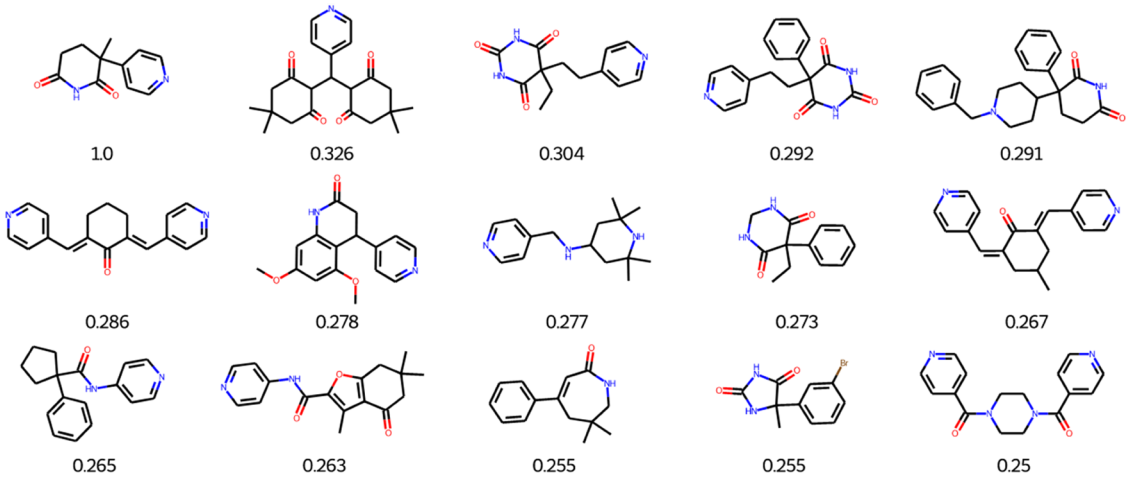


Figure 14. Top 14 molecules similar to compound 5.

color represents a negative charge and the yellow color represents a positive charge (Figure 9). The structures of the top five molecules suggest that they might develop strong interaction with the haem iron in the active site through a nitrogenous Sigma-donor ligand. This strong Fe−N interaction reduces the enzyme's intrinsic flexibility.[60] The loss of flexibility of the activity site also blocks the substrate channel, effectively putting a stop to the generation of a product responsible for tumor progression.[61] The prime advantage of these newly suggested molecules is the extremely lower systemic toxicity with elevated AI activity. Therefore, it is safe to say that the ML-based discovery and prediction of other drug molecules is an effective strategy.

The chemical and biological behavior of the molecules strongly depends on the chemical structure of the molecules.[62,63] The comparison of their structures can be used to design and screen better drugs. The similarity analysis is based on the comparison of structures. The highly similar molecules given as a reference (1−5) are given in Figures 10−14.

There is no doubt that the similarity score is not very high, but it is still much better than the random screening. Even the chance of finding a few potential candidates through this cheaper method is valuable.

4. CONCLUSIONS

Developing drugs requires a more time- and cost-efficient method and fast models to generate the best inhibitor for a given target protein.In the present study, molecular descriptors are calculated and shortlisted using various measures. Various machine learning models are trained. More than 5000 molecules from PubChem are extracted and their biological activities are predicted using already trained models. Using the collected data set, five of the best molecules are selected and their chemical properties are calculated. These selected five molecules are selected one by one to perform similarity analysis. Moreover, the present study provides new insight to find potential lead compounds for the targeted inhibition of aromatase-associated disorders.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c06174.

Tables S1−S3 containing collected data and details of descriptors (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Muhammad Ishfaq − *College of Computer Science, Huanggang Normal University, Huanggang 438000, China;* orcid.org/0000-0003-1376-8986; Email: muhammad@hgnu.edu.cn

Muhammad Aamir − *College of Computer Science, Huanggang Normal University, Huanggang 438000, China;*

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

Phone: 008618580344982; Email: aamirshaikh86@hgnu.edu.cn

Authors

Farooq Ahmad − *Department of Biomedical Engineering, College of Engineering and Applied Sciences, School of Chemistry and Chemical Engineering, Chemistry and Biomedicine Innovation Center (ChemBIC), Nanjing University, Nanjing 210093, China;* orcid.org/0000-0001-6019-436X

Abdelazim M Mebed − *Physics Department, Faculty of Science, Assiut University, Assiut 71516, Egypt; Department of Physics, College of Science, Jouf University, Sakaka 72388, Saudi Arabia*

Sayed Elshahat − *Physics Department, Faculty of Science, Assiut University, Assiut 71516, Egypt; Beijing Key Lab of Nanophotonics and Ultrafine Optoelectronic Systems, Center for Micro-Nanotechnology; Key Lab of Advanced Optoelectronic Quantum Design and Measurement, Ministry of Education, School of Physics, Beijing Institute of Technology, Beijing 100081, P. R. China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c06174

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This project is supported by the Huanggang Normal University Project (No. 2042022005). Farooq Ahmad thanks the National Natural Science Foundation of China (22150410326) and the Jiangsu Province Postdoctoral Science Foundation (2021K301C) for financial support.

■ REFERENCES

(1) Séralini, G.-E.; Moslemi, S. Aromatase Inhibitors: Past, Present and Future. *Mol. Cell. Endocrinol.* **2001**, *178*, 117−131.
(2) Siraki, A. G. Free Radical Metabolites in Arylamine Toxicity. In *Advances in Molecular Toxicology*; Elsevier, 2013; Vol. 7, pp 39−82.
(3) Dai, Y.; Wang, Q.; Zhang, X.; Jia, S.; Zheng, H.; Feng, D.; Yu, P. Molecular Docking and Qsar Study on Steroidal Compounds as Aromatase Inhibitors. *Eur. J. Med. Chem.* **2010**, *45*, 5612−5620.
(4) Verma, S. K.; Ratre, P.; Jain, A. K.; Liang, C.; Gupta, G. D.; Thareja, S. De Novo Designing, Assessment of Target Affinity and Binding Interactions against Aromatase: Discovery of Novel Leads as Anti-Breast Cancer Agents. *Struct. Chem.* **2021**, *32*, 847−858.
(5) Amaral, C.; Trouille, F. M.; Almeida, C. F.; Correia-da-Silva, G.; Teixeira, N. Unveiling the Mechanism of Action Behind the Anti-Cancer Properties of Cannabinoids in Er+ Breast Cancer Cells: Impact on Aromatase and Steroid Receptors. *J. Steroid Biochem. Mol. Biol.* **2021**, *210*, No. 105876.
(6) Ratre, P.; Mishra, K.; Dubey, A.; Vyas, A.; Jain, A.; Thareja, S. Aromatase Inhibitors for the Treatment of Breast Cancer: A Journey from the Scratch. *Anticancer Agents Med. Chem.* **2020**, *20*, 1994−2004.
(7) Sayyad, N. B.; Sabale, P. M.; Umare, M. D.; Bajaj, K. K. Aromatase Inhibitors: Development and Current Perspectives. *Indian J. Pharm. Educ. Res.* **2022**, *56*, 311−320.
(8) Zupin, L.; Pascolo, L.; Luppi, S.; Ottaviani, G.; Crovella, S.; Ricci, G. Photobiomodulation Therapy for Male Infertility. *Lasers Med. Sci.* **2020**, *35*, 1671−1680.
(9) Dong, M.; Jiang, S.; Tian, W.; Yan, Y.; Gao, C.; Gao, J.; Sheng, Y.; Wang, Y.; Xue, F. Preliminary Clinical Application of an Aromatase Inhibitor and a Gonadotropin-Releasing Hormone Agonist Combination for Inoperable Endometrial Cancer Patients with Comorbidities: Case Report and Literature Review. *Cancer Biol. Ther.* **2018**, *19*, 956−961.
(10) Kanakis, G. A.; Nordkap, L.; Bang, A.; Calogero, A.; Bártfai, G.; Corona, G.; Forti, G.; Toppari, J.; Goulis, D.; Jørgensen, N. Eaa Clinical Practice Guidelines—Gynecomastia Evaluation and Management. *Andrology* **2019**, *7*, 778−793.
(11) Mitra, S.; Lami, M. S.; Ghosh, A.; Das, R.; Tallei, T. E.; Islam, F.; Dhama, K.; Begum, M. Y.; Aldahish, A.; Chidambaram, K.; et al. Hormonal Therapy for Gynecological Cancers: How Far Has Science Progressed toward Clinical Applications? *Cancers* **2022**, *14*, 759.
(12) Yang, C.; Li, P.; Li, Z. Clinical Application of Aromatase Inhibitors to Treat Male Infertility. *Hum. Reprod. Update* **2021**, *28*, 30−50.
(13) Karaer, Ö.; Oruç, S.; Koyuncu, F. M. Aromatase Inhibitors: Possible Future Applications. *Acta Obstet. Gynecol. Scand.* **2004**, *83*, 699−706.
(14) Din, O. S.; Dodwell, D.; Wakefield, R. J.; Coleman, R. E. Aromatase Inhibitor-Induced Arthralgia in Early Breast Cancer: What Do We Know and How Can We Find out More? *Breast Cancer Res. Treat.* **2010**, *120*, 525−538.
(15) Rossi, A.; Iorio, A.; Scali, E.; Fortuna, M. C.; Mari, E.; Maxia, C.; Gerardi, M.; Framarino, M.; Carlesimo, M. Aromatase Inhibitors Induce 'Male Pattern Hair Loss' in Women? *Ann. Oncol.* **2013**, *24*, 1710−1711.
(16) Ahmad, F.; Mahmood, A.; Muhmood, T. Machine Learning-Integrated Omics for the Risk and Safety Assessment of Nanomaterials. *Biomater. Sci.* **2021**, *9*, 1598−1608.
(17) Mahmood, A. Photovoltaic and Charge Transport Behavior of Diketopyrrolopyrrole Based Compounds with a−D−a−D−a Skeleton. *J. Cluster Sci.* **2019**, *30*, 1123−1130.
(18) Mahmood, A.; Abdullah, M. I.; Khan, S. U.-D. Enhancement of Nonlinear Optical (Nlo) Properties of Indigo through Modification of Auxiliary Donor, Donor and Acceptor. *Spectrochim. Acta, Part A* **2015**, *139*, 425−430.
(19) Mahmood, A.; Abdullah Muhammad, I.; Nazar Muhammad, F. Quantum Chemical Designing of Novel Organic Non-Linear Optical Compounds. *Bull. Korean Chem. Soc.* **2014**, *35*, 1391−1396.
(20) Kumar, V.; Saha, A.; Roy, K. In Silico Modeling for Dual Inhibition of Acetylcholinesterase (Ache) and Butyrylcholinesterase (Buche) Enzymes in Alzheimer's Disease. *Comput. Biol. Chem.* **2020**, *88*, No. 107355.
(21) Banerjee, A.; De, P.; Kumar, V.; Kar, S.; Roy, K. Quick and Efficient Quantitative Predictions of Androgen Receptor Binding Affinity for Screening Endocrine Disruptor Chemicals Using 2d-Qsar and Chemical Read-Across. *Chemosphere* **2022**, *309*, No. 136579.
(22) Numazawa, M.; Yamada, K.; Nitta, S.; Sasaki, C.; Kidokoro, K. Role of Hydrophilic Interaction in Binding of Hydroxylated 3-Deoxy C19 Steroids to the Active Site of Aromatase. *J. Med. Chem.* **2001**, *44*, 4277−4283.
(23) Nagaoka, M.; Watari, Y.; Yajima, H.; Tsukioka, K.; Muroi, Y.; Yamada, K.; Numazawa, M. Structure−Activity Relationships of 3-Deoxy Androgens as Aromatase Inhibitors. Synthesis and Biochemical Studies of 4-Substituted 4-Ene and 5-Ene Steroids. *Steroids* **2003**, *68*, 533−542.
(24) Cepa, M. M.; da Silva, E. J. T.; Correia-da-Silva, G.; Roleira, F. M.; Teixeira, N. A. Synthesis and Biochemical Studies of 17-Substituted Androst-3-Enes and 3, 4-Epoxyandrostanes as Aromatase Inhibitors. *Steroids* **2008**, *73*, 1409−1415.
(25) Lone, S. H.; Bhat, M. A.; Lone, R. A.; Jameel, S.; Lone, J. A.; Bhat, K. A. Hemisynthesis, Computational and Molecular Docking Studies of Novel Nitrogen Containing Steroidal Aromatase Inhibitors: Testolactam and Testololactam. *New J. Chem.* **2018**, *42*, 4579−4589.
(26) Banjare, L.; Verma, S. K.; Jain, A. K.; Thareja, S. Structure Guided Molecular Docking Assisted Alignment Dependent 3dqsar Study on Steroidal Aromatase Inhibitors (Sais) as Anti-Breast Cancer Agents. *Lett. Drug Des. Discovery* **2019**, *16*, 808−817.
(27) Roleira, F. M. F.; Varela, C.; Amaral, C.; Costa, S. C.; Correia-da-Silva, G.; Moraca, F.; Costa, G.; Alcaro, S.; Teixeira, N. rA.; Tavares da Silva, E. J. C-6α-Vs C-7α-Substituted Steroidal Aromatase

Inhibitors: Which Is Better? Synthesis, Biochemical Evaluation, Docking Studies, and Structure−Activity Relationships. *J. Med. Chem.* **2019**, *62*, 3636−3657.
(28) Giampietro, L.; Gallorini, M.; Gambacorta, N.; Ammazzalorso, A.; De Filippis, B.; Della Valle, A.; Fantacuzzi, M.; Maccallini, C.; Mollica, A.; Cataldi, A.; et al. Synthesis, Structure-Activity Relationships and Molecular Docking Studies of Phenyldiazenyl Sulfonamides as Aromatase Inhibitors. *Eur. J. Med. Chem.* **2021**, *224*, No. 113737.
(29) Osmaniye, D.; Karaca, Ş.; Kurban, B.; Baysal, M.; Ahmad, I.; Patel, H.; Özkay, Y.; Kaplancıklı, Z. A. Design, Synthesis, Molecular Docking and Molecular Dynamics Studies of Novel Triazolothiadiazine Derivatives Containing Furan or Thiophene Rings as Anticancer Agents. *Bioorg. Chem.* **2022**, *122*, No. 105709.
(30) Mahmood, A.; Irfan, A. Effect of Fluorination on Exciton Binding Energy and Electronic Coupling in Small Molecule Acceptors for Organic Solar Cells. *Comput. Theor. Chem.* **2020**, *1179*, No. 112797.
(31) Mahmood, A.; Khan, S. U.-D.; Rana, U. A.; Tahir, M. H. Red Shifting of Absorption Maxima of Phenothiazine Based Dyes by Incorporating Electron-Deficient Thiadiazole Derivatives as Π-Spacer. *Arab. J. Chem.* **2019**, *12*, 1447−1453.
(32) Mahmood, A.; Irfan, A.; Wang, J.-L. Developing Efficient Small Molecule Acceptors with Sp2-Hybridized Nitrogen at Different Positions by Density Functional Theory Calculations, Molecular Dynamics Simulations and Machine Learning. *Chem. - Eur. J.* **2022**, *28*, No. e202103712.
(33) Mahmood, A.; Wang, J.-L. A Time and Resource Efficient Machine Learning Assisted Design of Non-Fullerene Small Molecule Acceptors for P3ht-Based Organic Solar Cells and Green Solvent Selection. *J. Mater. Chem. A* **2021**, *9*, 15684−15695.
(34) Mahmood, A.; Khan, S. U.-D.; Rehman, Fu. Assessing the Quantum Mechanical Level of Theory for Prediction of Uv/Visible Absorption Spectra of Some Aminoazobenzene Dyes. *J. Saudi Chem. Soc.* **2015**, *19*, 436−441.
(35) Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. Dragon Software: An Easy Approach to Molecular Descriptor Calculations. *MATCH Commun. Math. Comput. Chem.* **2006**, *56*, 237−248.
(36) Landrum, G. *Rdkit: Open-Source Cheminformatics*. http://www.rdkit.org.
(37) Khalid, M.; Ali, A.; Abid, S.; Tahir, M. N.; Khan, M. U.; Ashfaq, M.; Imran, M.; Ahmad, A. Facile Ultrasound-Based Synthesis, Sc-Xrd, Dft Exploration of the Substituted Acyl-Hydrazones: An Experimental and Theoretical Slant Towards Supramolecular Chemistry. *ChemistrySelect* **2020**, *5*, 14844−14856.
(38) Khalid, M.; Ali, A.; Khan, M. U.; Tahir, M. N.; Ahmad, A.; Ashfaq, M.; Hussain, R.; Morais, S. F. d. A.; Braga, A. A. C. Non-Covalent Interactions Abetted Supramolecular Arrangements of N-Substituted Benzylidene Acetohydrazide to Direct Its Solid-State Network. *J. Mol. Struct.* **2021**, *1230*, No. 129827.
(39) Najam, T.; Shah, S. S. A.; Ding, W.; Jiang, J.; Jia, L.; Yao, W.; Li, L.; Wei, Z. An Efficient Anti-Poisoning Catalyst against Sox, Nox, and Pox: P, N-Doped Carbon for Oxygen Reduction in Acidic Media. *Angew. Chem., Int. Ed.* **2018**, *57*, 15101−15106.
(40) Shah, S. S. A.; Najam, T.; Javed, M. S.; Rahman, M. M.; Tsiakaras, P. Novel Mn-/Co-Nx Moieties Captured in N-Doped Carbon Nanotubes for Enhanced Oxygen Reduction Activity and Stability in Acidic and Alkaline Media. *ACS Appl. Mater. Interfaces* **2021**, *13*, 23191−23200.
(41) Shah, S. S. A.; Najam, T.; Nazir, M. A.; Wu, Y.; Ali, H.; Rehman, A. U.; Rahman, M. M.; Imran, M.; Javed, M. S. Salt-Assisted Gas-Liquid Interfacial Fluorine Doping: Metal-Free Defect-Induced Electrocatalyst for Oxygen Reduction Reaction. *Mol. Catal.* **2021**, *514*, No. 111878.
(42) Mahmood, A.; Wang, J.-L. Machine Learning for High Performance Organic Solar Cells: Current Scenario and Future Prospects. *Energy Environ. Sci.* **2021**, *14*, 90−105.
(43) Mahmood, A.; Khan, S. U.-D.; Rana, U. A. Theoretical Designing of Novel Heterocyclic Azo Dyes for Dye Sensitized Solar Cells. *J. Comput. Electron.* **2014**, *13*, 1033−1041.
(44) Hussain, R.; Hassan, F.; Khan, M. U.; Mehboob, M. Y.; Fatima, R.; Khalid, M.; Mahmood, K.; Tariq, C. J.; Akhtar, M. N. Molecular Engineering of a−D−C−D−a Configured Small Molecular Acceptors (Smas) with Promising Photovoltaic Properties for High-Efficiency Fullerene-Free Organic Solar Cells. *Opt. Quantum Electron.* **2020**, *52*, 364.
(45) Hussain, R.; Mehboob, M. Y.; Khan, M. U.; Khalid, M.; Irshad, Z.; Fatima, R.; Anwar, A.; Nawab, S.; Adnan, M. Efficient Designing of Triphenylamine-Based Hole Transport Materials with Outstanding Photovoltaic Characteristics for Organic Solar Cells. *J. Mater. Sci.* **2021**, *56*, 5113−5131.
(46) Mahmood, A.; Irfan, A.; Wang, J.-L. Machine Learning for Organic Photovoltaic Polymers: A Minireview. *Chin. J. Polym. Sci.* **2022**, *40*, 870−876.
(47) Mahmood, A.; Irfan, A.; Wang, J.-L. Machine Learning and Molecular Dynamics Simulation-Assisted Evolutionary Design and Discovery Pipeline to Screen Efficient Small Molecule Acceptors for Ptb7-Th-Based Organic Solar Cells with over 15% Efficiency. *J. Mater. Chem. A* **2022**, *10*, 4170−4180.
(48) Janjua, M. R. S. A.; Irfan, A.; Hussien, M.; Ali, M.; Saqib, M.; Sulaman, M. Machine-Learning Analysis of Small-Molecule Donors for Fullerene Based Organic Solar Cells. *Energy Technol.* **2022**, *10*, No. 2200019.
(49) Mahmood, A.; Saqib, M.; Ali, M.; Abdullah, M. I.; Khalid, B. Theoretical Investigation for the Designing of Novel Antioxidants. *Can. J. Chem.* **2013**, *91*, 126−130.
(50) Mahmood, A.; Khan, S. U.-D.; Rana, U. A.; Janjua, M. R. S. A.; Tahir, M. H.; Nazar, M. F.; Song, Y. Effect of Thiophene Rings on Uv/Visible Spectra and Non-Linear Optical (Nlo) Properties of Triphenylamine Based Dyes: A Quantum Chemical Perspective. *J. Phys. Org. Chem.* **2015**, *28*, 418−422.
(51) Khan, M. U.; Khalid, M.; Hussain, R.; Umar, A.; Mehboob, M. Y.; Shafiq, Z.; Imran, M.; Irfan, A. Novel W-Shaped Oxygen Heterocycle-Fused Fluorene-Based Non-Fullerene Acceptors: First Theoretical Framework for Designing Environment-Friendly Organic Solar Cells. *Energy Fuels* **2021**, *35*, 12436−12450.
(52) Khan, M. U.; Mehboob, M. Y.; Hussain, R.; Fatima, R.; Tahir, M. S.; Khalid, M.; Braga, A. A. C. Molecular Designing of High-Performance 3d Star-Shaped Electron Acceptors Containing a Truxene Core for Nonfullerene Organic Solar Cells. *J. Phys. Org. Chem.* **2021**, *34*, No. e4119.
(53) Roy, P. P.; Roy, K. On Some Aspects of Variable Selection for Partial Least Squares Regression Models. *QSAR Comb. Sci.* **2008**, *27*, 302−313.
(54) Pratim Roy, P.; Paul, S.; Mitra, I.; Roy, K. On Two Novel Parameters for Validation of Predictive Qsar Models. *Molecules* **2009**, *14*, 1660−1701.
(55) Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N. Some Case Studies on Application of "Rm2" Metrics for Judging Quality of Quantitative Structure−Activity Relationship Predictions: Emphasis on Scaling of Response Data. *J. Comput. Chem.* **2013**, *34*, 1071−1082.
(56) Hähnke, V. D.; Kim, S.; Bolton, E. E. Pubchem Chemical Structure Standardization. *J. Cheminf.* **2018**, *10*, No. 36.
(57) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57*, 3186−3204.
(58) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* **2002**, *45*, 4350−4358.
(59) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; et al. The Chembl Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945−d954.
(60) Souza, S. A.; Held, A.; Lu, W. J.; Drouhard, B.; Avila, B.; Leyva-Montes, R.; Hu, M.; Miller, B. R.; Ng, H. L. Mechanisms of Allosteric and Mixed Mode Aromatase Inhibitors. *RSC Chem. Biol.* **2021**, *2*, 892−905.

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149

(61) Di Nardo, G.; Breitner, M.; Sadeghi, S. J.; Castrignanò, S.; Mei, G.; Di Venere, A.; Nicolai, E.; Allegra, P.; Gilardi, G. Dynamics and Flexibility of Human Aromatase Probed by Ftir and Time Resolved Fluorescence Spectroscopy. *PLoS One* **2013**, *8*, No. e82118.
(62) Mahmood, A.; Irfan, A. Computational Analysis to Understand the Performance Difference between Two Small-Molecule Acceptors Differing in Their Terminal Electron-Deficient Group. *J. Comput. Electron.* **2020**, *19*, 931−939.
(63) Siddiqui, W. A.; Khalid, M.; Ashraf, A.; Shafiq, I.; Parvez, M.; Imran, M.; Irfan, A.; Hanif, M.; Khan, M. U.; Sher, F.; Ali, A. Antibacterial Metal Complexes of O-Sulfamoylbenzoic Acid: Synthesis, Characterization, and Dft Study. *Appl. Organomet. Chem.* **2022**, *36*, No. e6464.

https://doi.org/10.1021/acsomega.2c06174
ACS Omega 2022, 7, 48139−48149