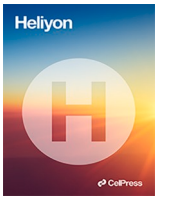
Heliyon 9 (2023) e21332

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Data mining and library generation to search electron-rich and electron-deficient building blocks for the designing of polymers for photoacoustic imaging

Muhammad Ishfaq [a], Tayyaba Mubashir [b], Safaa N. Abdou [c], Mudassir Hussain Tahir [d], Mohamed Ibrahim Halawa [e,f], Mohamed M. Ibrahim [g], Yulin Xie [a,*]

[a] Huanggang Normal University, Huanggang, 438000, China
[b] Institute of Chemistry, University of Sargodha, Sargodha, 40100, Pakistan
[c] Department of Chemistry, Khurmah University College, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia
[d] Research Faculty of Agriculture, Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Hokkaido, 060-8589, 060-0811, Japan
[e] Department of Pharmaceutical Analytical Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura, 35516, Mansoura, Egypt
[f] Guangdong Laboratory of Artificial Intelligence & Digital Economy (SZ), Shenzhen University, Shenzhen, 518060, China
[g] Department of Chemistry, College of Science, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia

ARTICLE INFO

Keywords:
Photoacoustic imaging
Near infrared
Polymers
Machine learning

ABSTRACT

Photoacoustic imaging is a good method for biological imaging, for this purpose, materials with strong near infrared (NIR) absorbance are required. In the present study, machine learning models are used to predict the light absorption behavior of polymers. Molecular descriptors are utilized to train a variety of machine learning models. Building blocks are searched from chemical databases, as well as new building blocks are designed using chemical library enumeration method. The Breaking Retrosynthetically Interesting Chemical Substructures (BRICS) method is employed for the creation of 10,000 novel polymers. These polymers are designed based on the input of searched and selected building blocks. To enhance the process, the optimal machine learning model is utilized to predict the UV/visible absorption maxima of the newly designed polymers. Concurrently, chemical similarity analysis is also performed on the selected polymers, and synthetic accessibility of selected polymers is calculated. In summary, the polymers are all easy to synthesize, increasing their potential for practical applications.

1. Introduction

The nanotechnology has rapidly developed over the past few decades, leading to the development of nano-diagnostic devices. Integrating multi-diagnostics and therapeutics is the beauty of nanomedicine. Several imaging techniques have been developed in recent years as important diagnostic tools in clinics, including computed tomography (CT), ultrasound, positron emission tomography (PET) and magnetic resonance imaging (MRI). As a hybrid imaging technique that combines optical excitation with ultrasonic recognition, photoacoustic imaging (PAI) is a noninvasive hybrid imaging technique [1]. The photoacoustic contrast agent is injected

* Corresponding author.
E-mail address: xieyulin@hgnu.edu.cn (Y. Xie).

https://doi.org/10.1016/j.heliyon.2023.e21332
Received 21 June 2023; Received in revised form 8 October 2023; Accepted 19 October 2023
Available online 26 October 2023

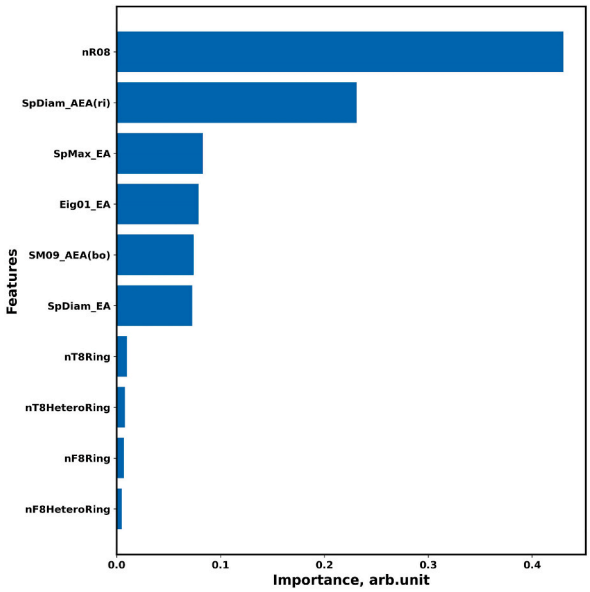


Fig. 1. The importance of features (descriptors).

Table 1
Mean absolute error (MAE) and R^2 values for different machine learning models (test set).

Model	MAE (nm)	R^2
Random forest regression	40	0.91
Bagging regression	45	0.86
Gradient boosting regression	49	0.85
Linear regression	61	0.62

into the body and irradiated by a laser to degenerate the temperature. After the sound wave is diffused by thermoelasticity, an image is constructed at the detector with a high spatial resolution and tissue penetration depth [2,3]. Various types of near-infrared organic fluorescent contrast enhancement agents have been used for PA imaging, such as Indocyanine green (ICG), Poly-L-lysine (PLL), and Prussian blue nanoparticles (PBNP) [4,5]. The low light stability, short emission wavelength, toxicity, aggregation and photothermal conversion efficiency of organic fluorescent materials are major challenges for PA imaging [6]. However, these drawbacks can be overcome by designing a hybrid combination of organic and inorganic semiconducting polymeric nanoparticles (SPNPs). SPNPs are a promising type of organic photonic agent in which electrons can move along the polymer backbone and mainly depend on the chemical structures of semiconducting polymers delocalized by an electron [7,8]. These conjugated polymers generally have excellent photostability and a wide absorption range suitable for multifunctional applications. In addition, donor–acceptor (D–A) polymers are getting significant interest due to their many advantages, such as enhanced light absorption, good biocompatibility, tunability and stability of the acceptor energy levels, over conventional polymer/fullerene [9–11].

In this modern era, polymers are an essential part of many products, and researchers have been paying much attention to enhancing their functional properties through careful structural manipulation [12–14]. In recent years, theoretical studies on polymers have been considered significant for their successful commercialization. Moreover, theoretical modeling-based prediction develops more efficient processes in less time and resources, allowing scientists to understand the nature of polymers efficiently [15,16]. Therefore, theoretical principles play a proficient role in developing novel and specified polymers that tailor their properties according to the applications [17,18].

Machine learning, the subcategory of artificial intelligence and computer science, is an exciting and powerful technique for extracting meaningful information from large-scale or solving complex patterns in recent years [19–21]. The learning and building process of a machine learning algorithm can be divided into three components: an error function, a decision process, and a modal optimization process [22,23]. Machine learning has been applied in various directions outside materials science and engineering, such as navigation, product recommendation, language translation, and work sectors. The prediction from data can provide way to accelerate the work.

In addition, machine learning provides a vital tool for the advancement of polymer studies [24,25]. The researchers can uncover patterns and relationships of polymers by leveraging machine learning algorithms, which is impossible for traditional methods. This allows a more accurate understanding of complex processes such as material properties, performance, and longevity in short time. Moreover, machine learning can be used to improve the design of polymers, developing better structures for specific applications [26,

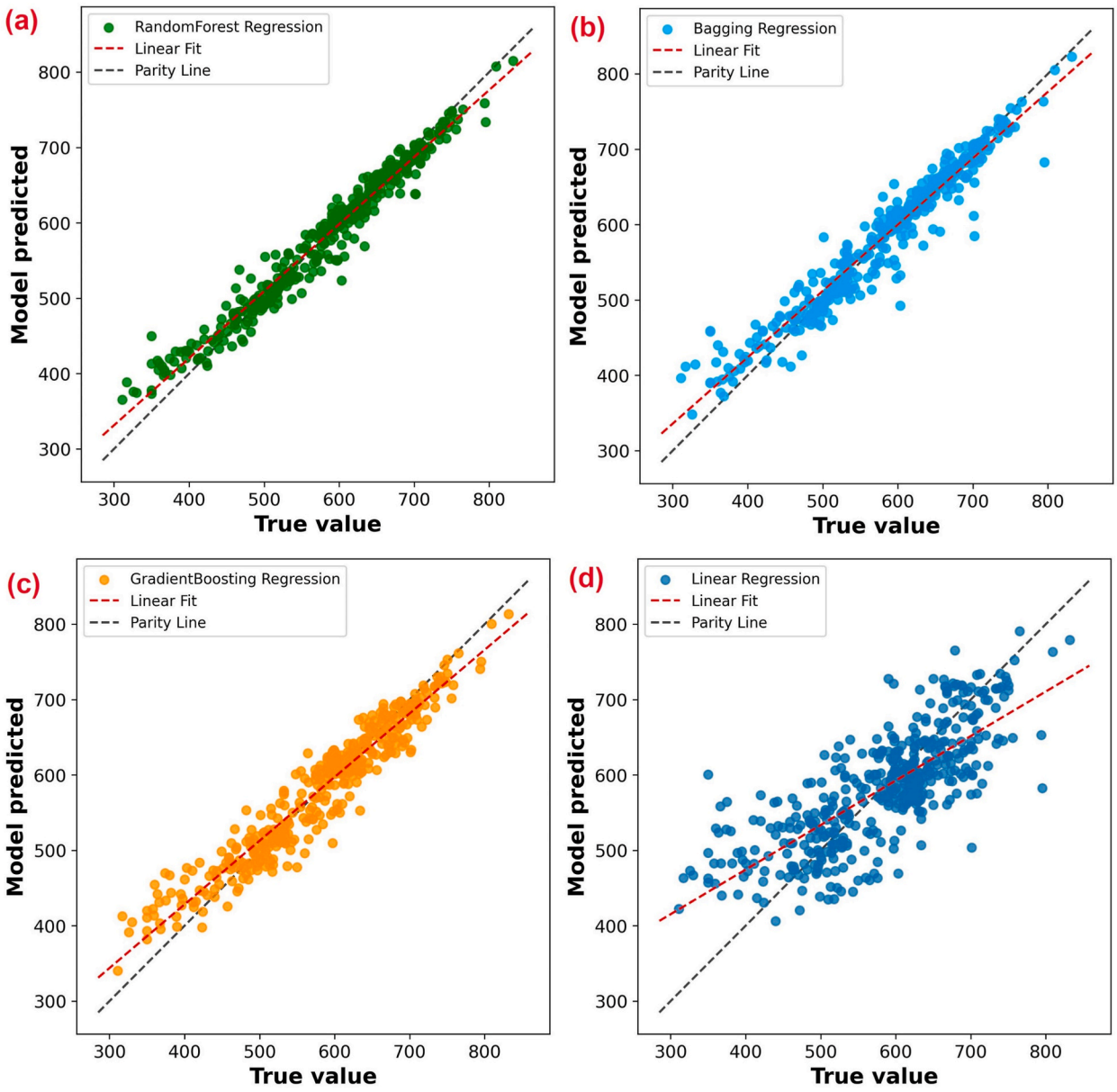


Fig. 2. Analysis of the scatter plot between true and predicted values (a) Random forest regression (b) Bagging regression (c) Gradient boosting regression (d) Linear regression.

27]. Machine learning is very fast as compare with quantum chemical calculations [28]. It can be applied to study any property for which enough data is available or can be generated. The development of more accurate algorithm has increased the potential of machine learning [29].

The purpose of this study is to develop machine learning models that can predict the light absorption behavior of polymers. Various machine learning models undergo training using molecular descriptors. Simultaneously, building blocks are sourced from chemical databases, and novel ones are generated through the chemical library enumeration method. This approach leads to the creation of 10,000 new polymer designs. To anticipate the UV/visible absorption maxima of these polymers, the most proficient machine learning model is employed. Additionally, a chemical similarity analysis is executed on the chosen polymers, followed by the computation of the synthetic accessibility of these selected compounds.

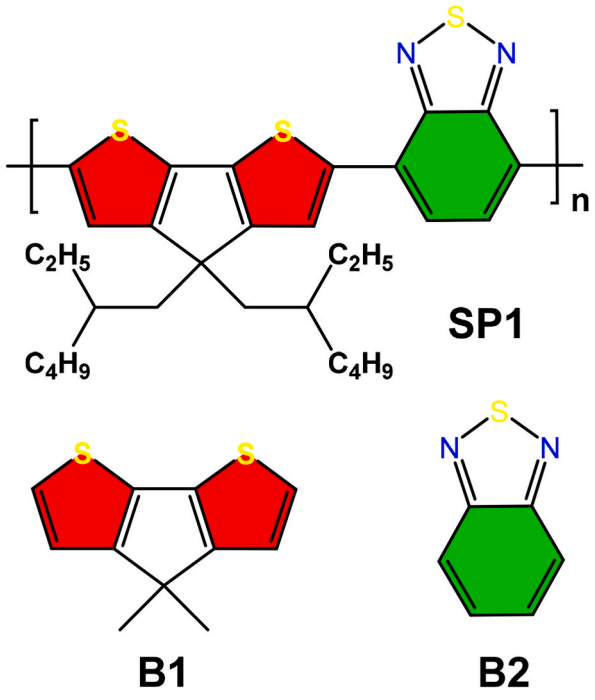


Fig. 3. Structure of reference polymer and reference building blocks.

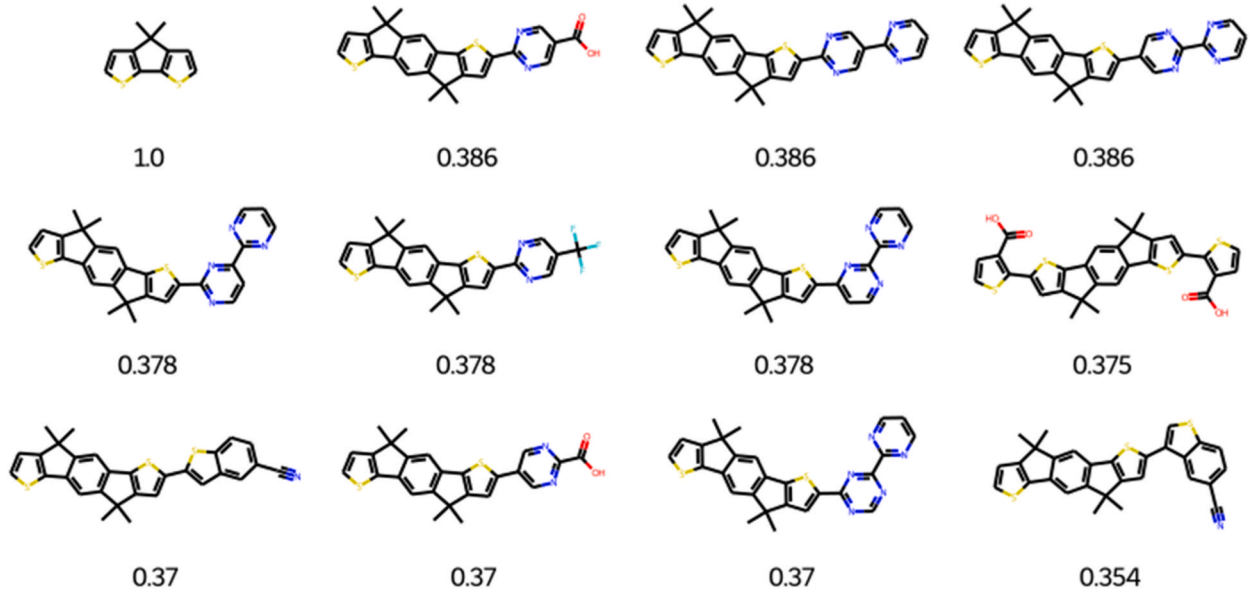


Fig. 4. Top search hits B1 from Harvard Clean Energy Project database.

2. Computational methods

2.1. Statistical analysis (machine learning)

UV/visible absorption data of 600 organic semi-conductors is collected from literature. Molecular descriptors (features) are calculated using Dragon software [30]. About four thousand descriptors were generated. Only those descriptors were further short-listed that exhibited the statistically significant values. The particular descriptors with the constant values, missing values, zero values, and having high self-correlation were excluded from the analysis. After data cleaning, the descriptors are used for machine learning.

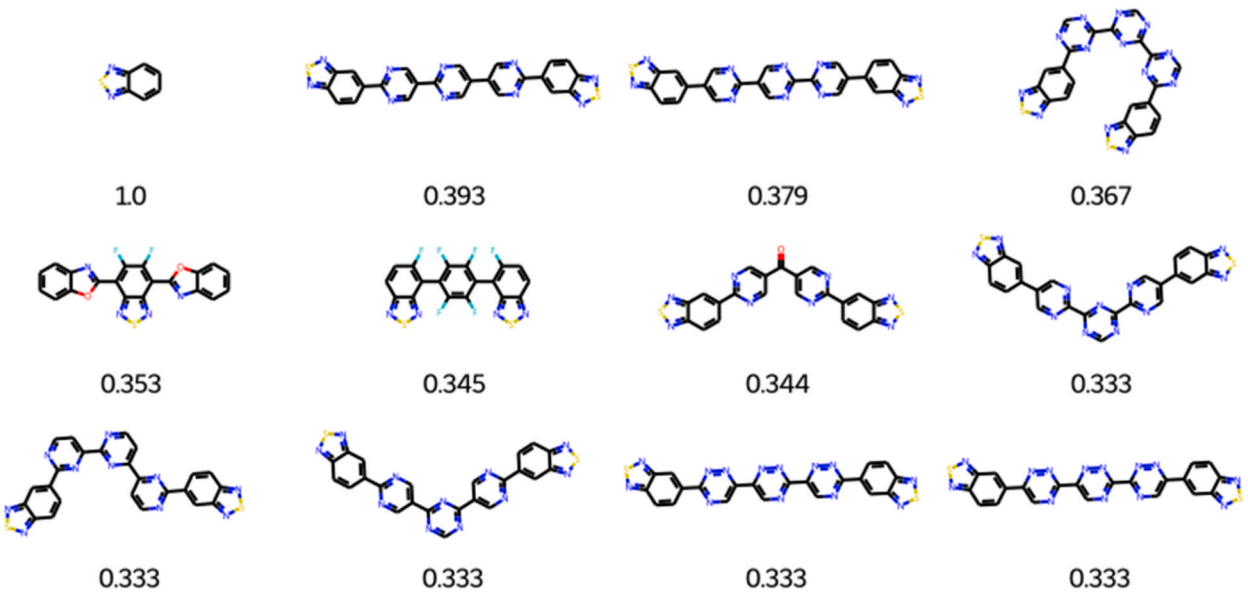


Fig. 5. Top search hits for B2 from Harvard Clean Energy Project database.

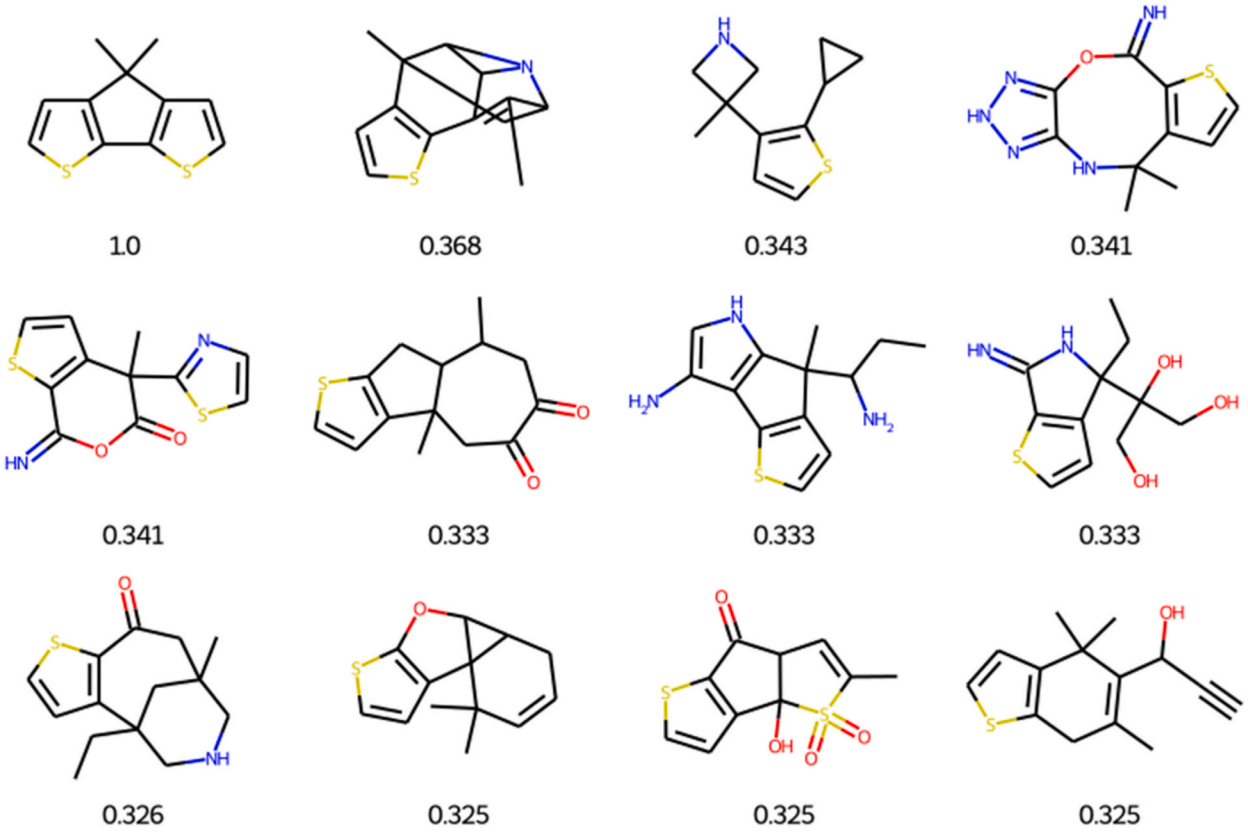


Fig. 6. Top search hits for B1 from GDB17 chemical database.

For the ML modelling and data visualization, a variety of Python-based libraries were used, including Matplotlib, Scikit-Learn, Seaborn, etc. The absorption maxima was elected as the target feature and the selected features (descriptors) were utilized as the independent variables for model training.

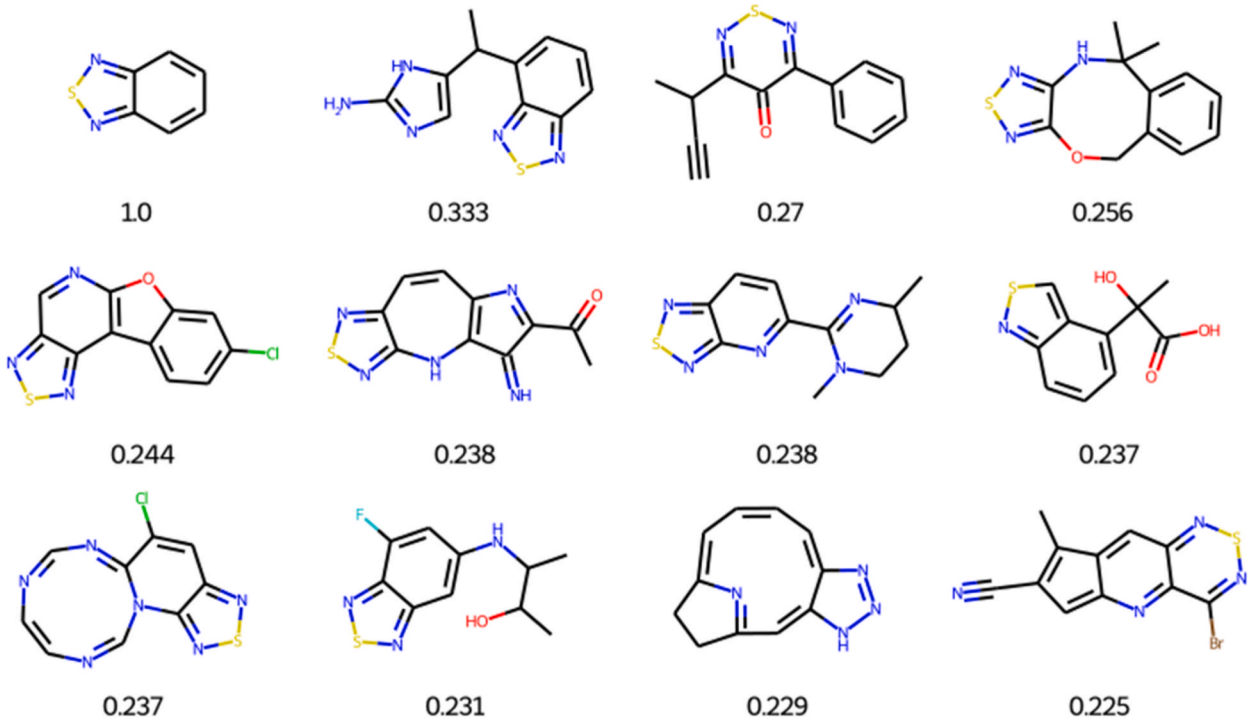


Fig. 7. Top search hits for B2 from GDB17 chemical database.

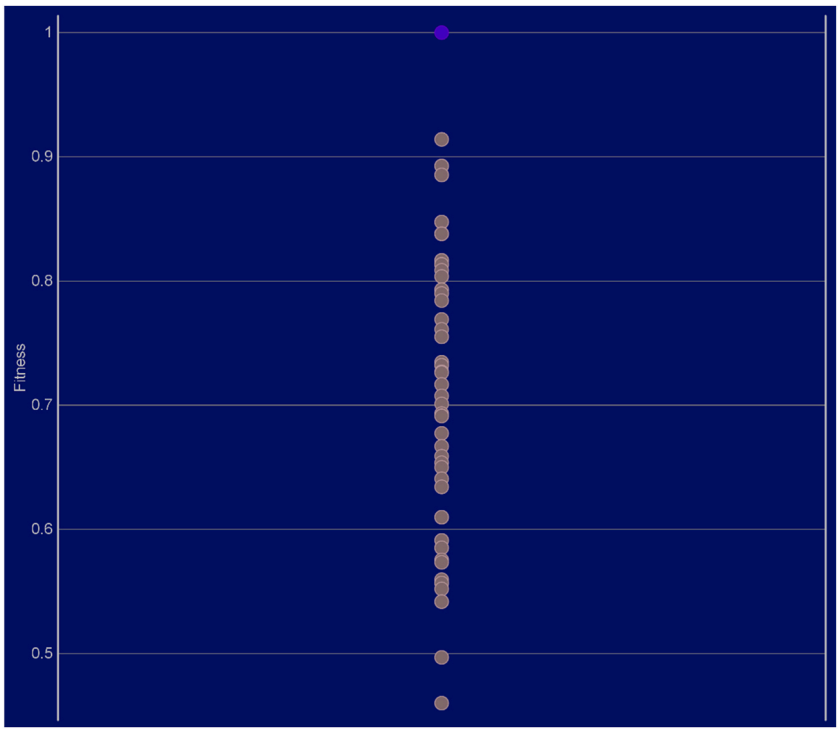


Fig. 8. B1 fitness score for generated building blocks.

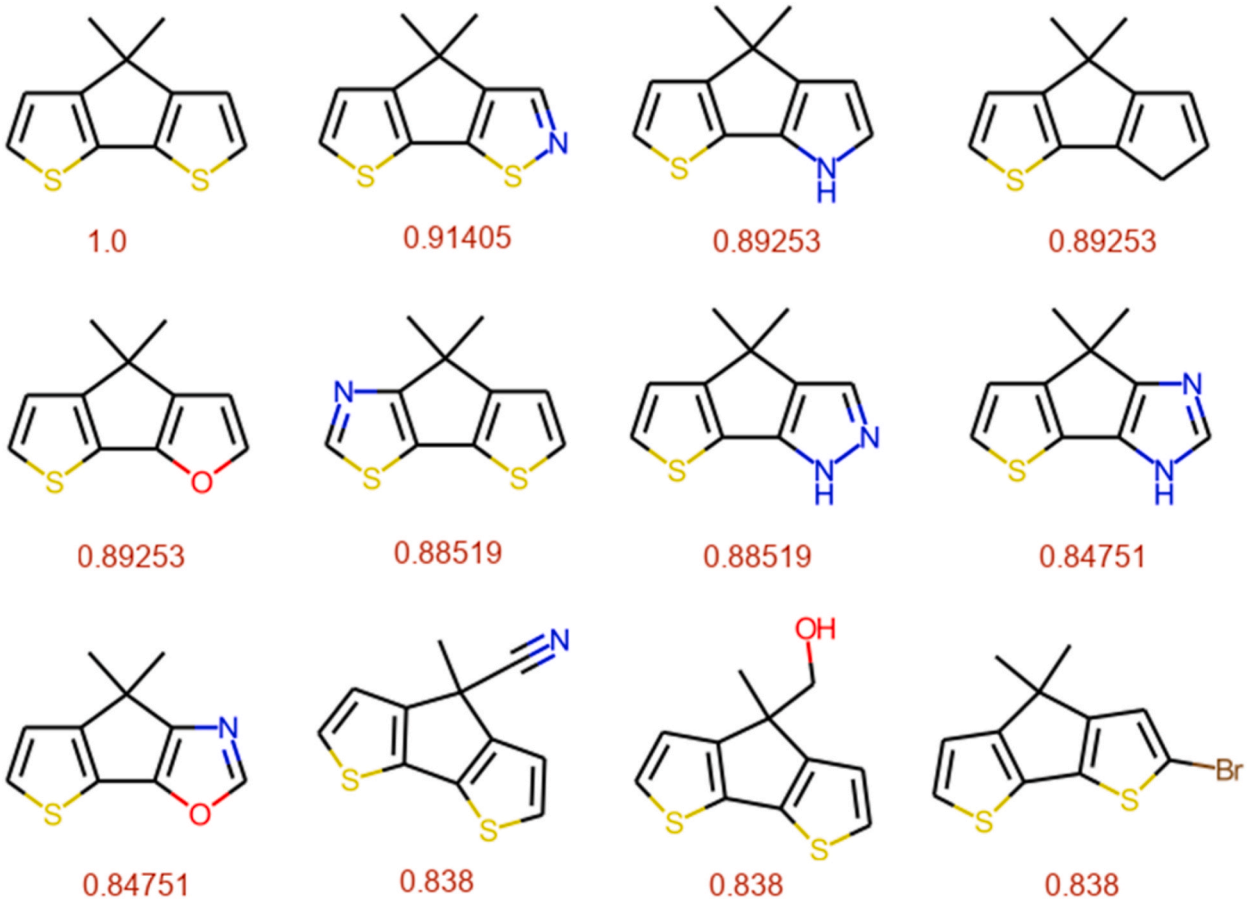


Fig. 9. The top eight building blocks that are similar to B1.

2.2. Chemical similarity analysis

Chemical similarity analysis is implanted on the python based software named as RDKit [31]. Using Chemical similarity analysis similar structures can be searched from databases. Tanimoto similarity index is selected for structural comparison, it is calculated using Extended Connectivity Fingerprints (ECFP4) [32].

2.3. Library generation

The library of electron-deficient and electron-rich building blocks is generated using DataWarrior software [33]. To generate libraries of similar molecules, an evolutionary algorithm is employed. A random structural modification of a reference structure is applied to create new molecules. For each generated molecule, a fitness score is generated. Generated molecules which are structurally more similar to reference molecules will receive a higher score.

2.4. Monomer design

BRICS (Breaking Retrosynthetically Interesting Chemical Substructures) is a method for breaking and joining building blocks together in order to form monomers using a variety of combinations of electron-rich and electron-deficient building blocks [34].

3. Results and discussion

3.1. Results of machine learning

Machine learning analysis need input that can be in several formats. We have used molecular descriptors for machine learning analysis. About 4000 descriptors (features) with the applicability of Dragon software are further screened based on univariate regression. Descriptors are typically integers, matrix, vector, or other data structures notably special characters are also feasible. This is

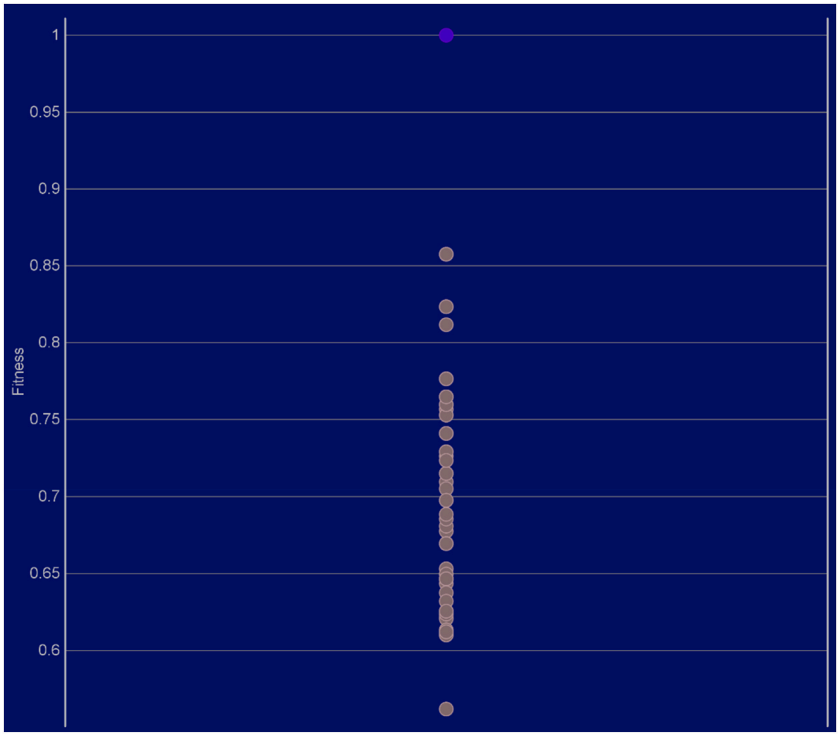


Fig. 10. B2 fitness score of generated building blocks.

a process used for machine learning to anticipate a molecule's property. These can be classified into theoretical and experimental descriptors. The structural formula or empirical formula are two examples of theoretical descriptors generated from the symbolic representation of molecules. Five classes can be made from this group of descriptors. The five kinds of theoretical descriptors are shown in the accompanying graphs along with the correlation between their dimensionality, the data they include, and the convenience of calculation. It is worthy to mention that machine learning model training does not benefit from all features (descriptors) equally. Therefore, it is essential to find the important features. The importance of features is given in Fig. 1 nR08 is most important feature and nF8HeteroRing is less important feature.

The type of machine learning models affects the performance machine learning analysis [35,36]. Therefore, we have tried multiple models to find best model. Additionally, cross-validation (CV) is performed, and results of 10-fold CV have shown higher performance. Table 1 shows the performance parameters for different models, such as mean absolute error (MAE) values and r-square values. It is noted that the performance of random forest regressor and bagging regressor was higher. The hyperparameters of these models are optimized. The dependance on the expensive experimental methods can be decreased with the help of accurate prediction [37–41]. Fig. 2 is showing the scatter plot between true and predicted values for different models. The reliability of machine learning models is checked using external dataset [42]. R^2 values for random forest regression, bagging regression, gradient boosting regression and linear regression are 0.91, 0.86, 0.85 and 0.62 respectively. Random forest regression is best model.

3.2. Similarity analysis

Chemical similarity analysis (CSA) is used to identify the resemblance among molecules [43–45]. The purpose is to check in which aspects the molecules resemble and differ. Those compounds that are recognized by this analysis are produced from the reaction of various compounds with single reagents by comparison of structures [46–48]. A polymer named as SP1 that is successfully used for photoacoustic imaging is selected as reference for the designing of new monomers [49]. The electron-rich (B1) and electron-deficient (B2) parts of SP1 are used reference building blocks for data mining and library generation. The standard structures of reference polymer and reference building blocks are given in Fig. 3. Similarity analysis is performed using RDKit [31].

Harvard Clean Energy Project database is used for similarity analysis [50]. Through high technological designing and screening Harvard clean energy project (CEP) advance organic photovoltaic materials are flourished. CEP is used for the advancement of OPV from the latest compounds. It is a mechanized and highly efficient system for the in-silico investigation and study of molecular materials. Top search hits B1 from Harvard Clean Energy Project database are given in Fig. 4. Highest similarity index is 0.386. This approach can help to select the unique structures. Top search hits for B2 from Harvard Clean Energy Project database are shown in Fig. 5. Here, highest similarity index is 0.393. However, majority of structures are suitable for further designing due to non-availability of connecting sites.

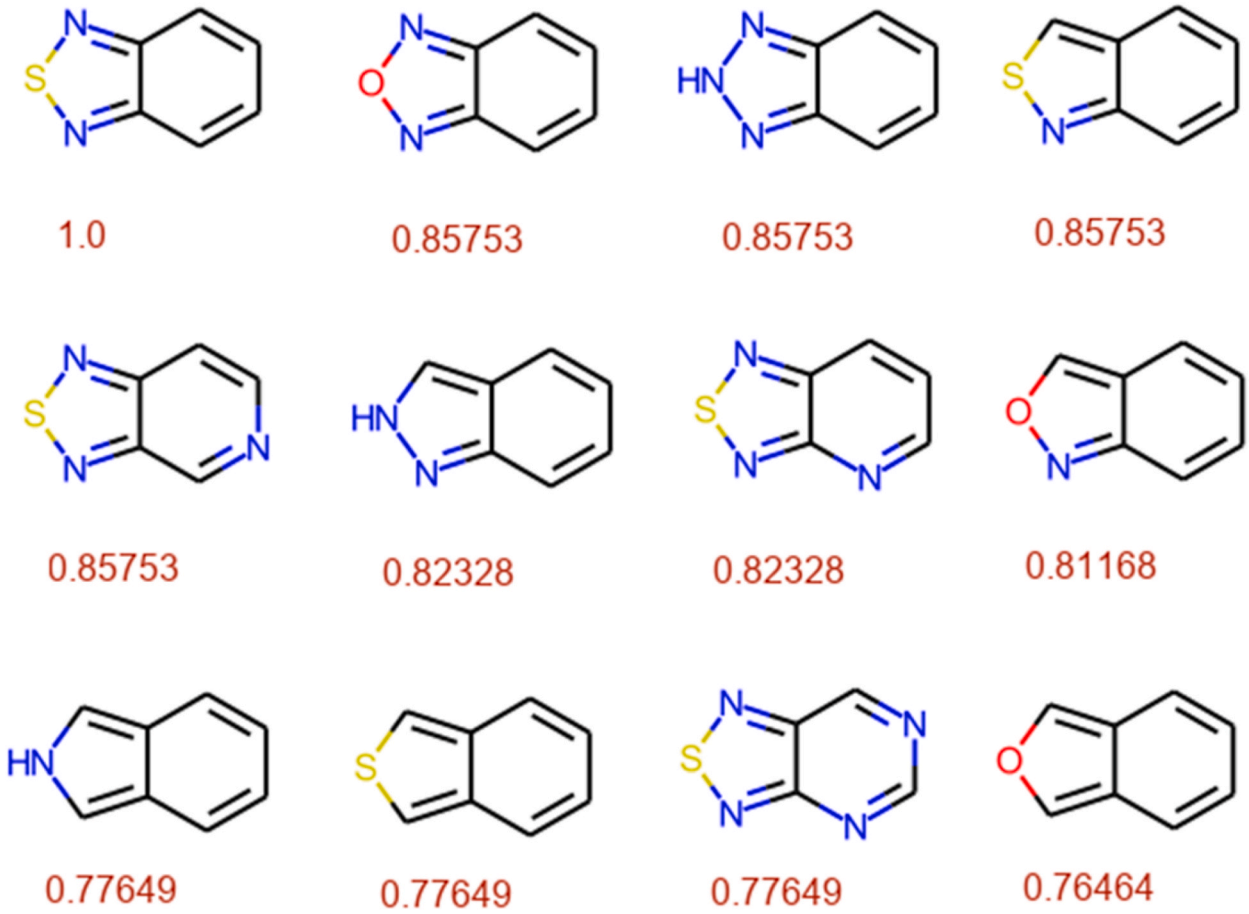


Fig. 11. A list of the top eight building blocks that are similar to B2.

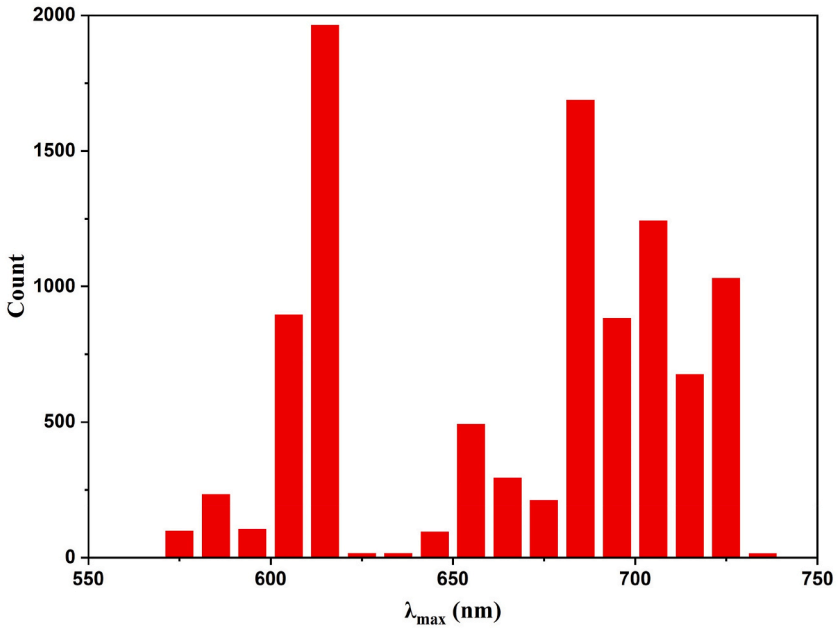


Fig. 12. Distribution of predicted λ_{max} values.

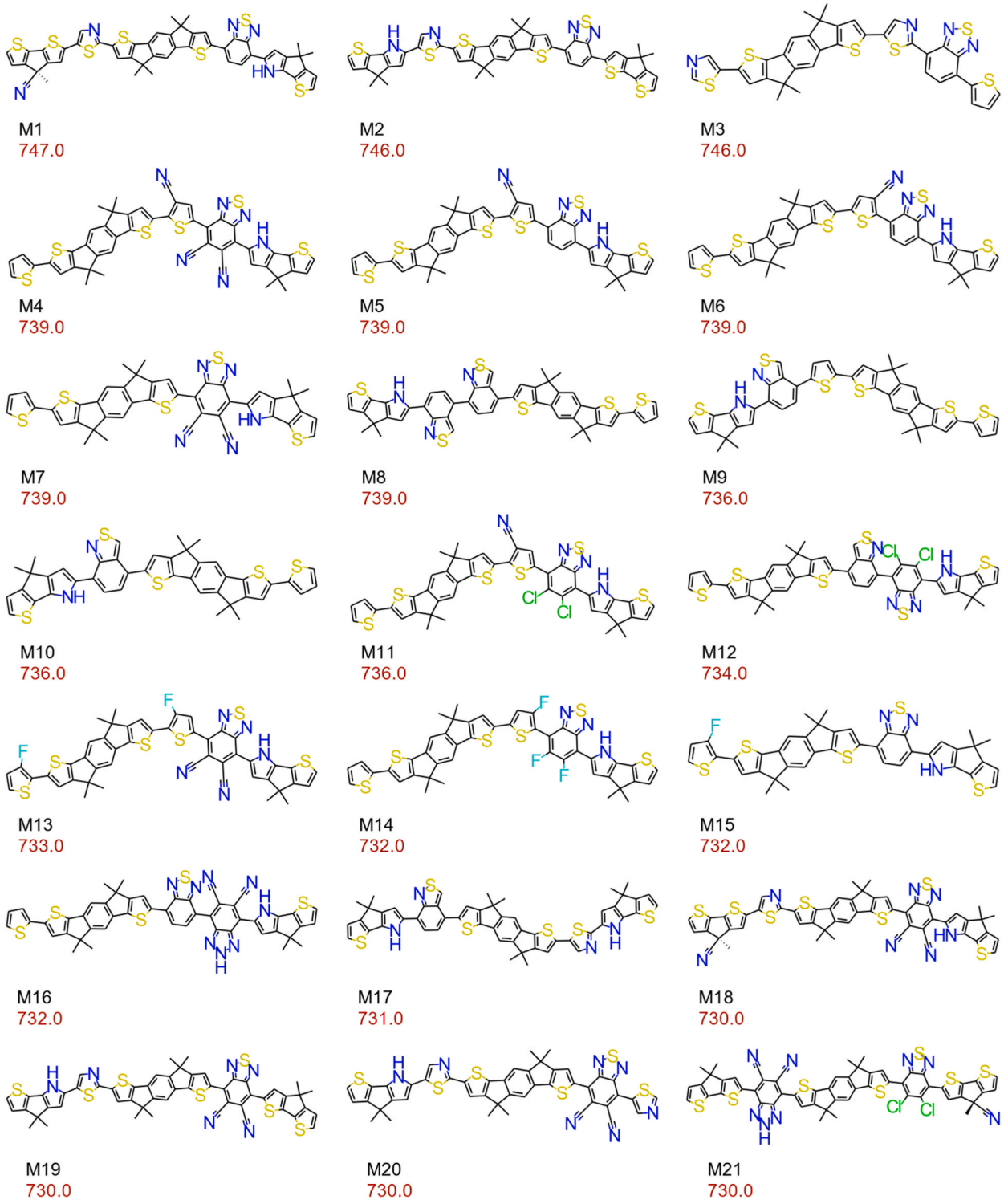


Fig. 13. Structures and λ_{max} values (nm) of best monomers.

GDB17 chemical database is also used to search building blocks [51]. This database contains large number of drugs that contain hydrogen, carbon, nitrogen, oxygen, sulfur and halogen atoms. This database has large number of nonaromatic heterocycles and quaternary centers. There are much chances to find unique compounds. Top search hits for B1 from GDB17 chemical database are given in Fig. 6. Highest similarity index is 0.368. However, most of structures cannot be used for further designing because these

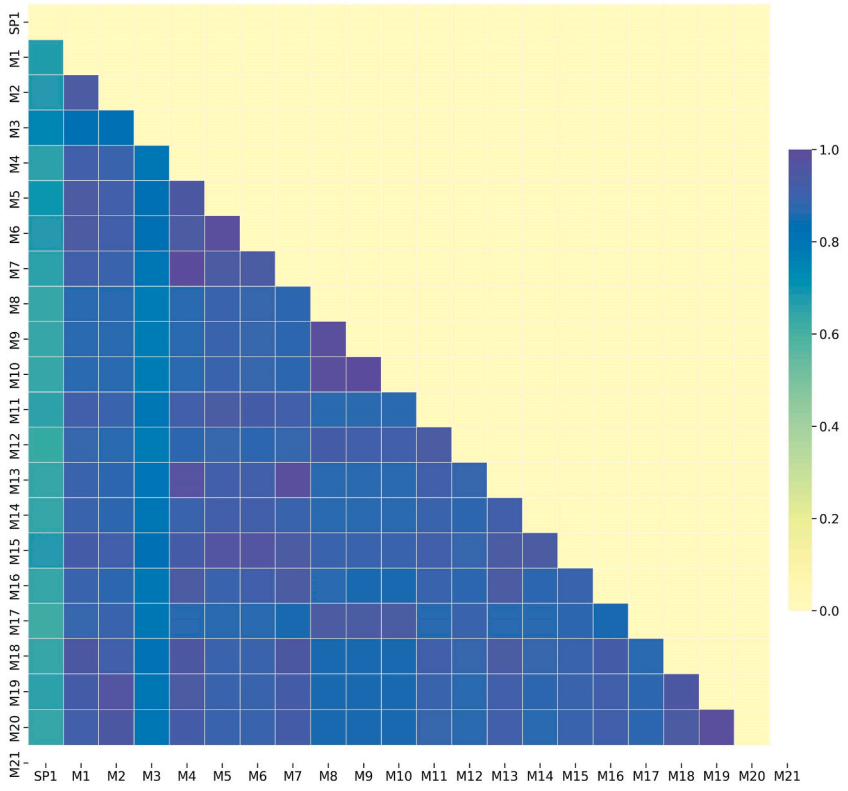


Fig. 14. Heatmap of structural similarity between selected monomers.

structures cannot be connected in a proper way to get suitable polymers. Top search hits for B2 from GDB17 chemical database are given in Fig. 7. These structures are also not very good building blocks for the designing polymers.

3.3. Library generation

Library enumeration is the process used for generating the building blocks through automatic method [52–54]. The goal is to assemble a group of molecules that reflect a certain property or performance metric. Without investigation, the model prediction of molecules for specific application is not possible. The most significant features of the system may, however, be identified and used to guide the selection process. Library generation is done using DataWarrior software [33]. As shown in Fig. 8, the plot of fitness scores for libraries generated using B1 can be seen. The fitness score of only one group exceeds 0.9. The fitness score for the majority of building blocks falls in wide range (0.5–0.9). In Fig. 9, structures of top 8 building blocks are given are given.

Fig. 10 illustrates the fitness score plot for building blocks designed from B2. It is noteworthy that no group has a fitness score higher than 0.9 and only a few groups have a score higher than 0.8. The majority of the groups have a fitness score between 0.6 and 0.8. Fig. 11 is showing the structures of top 8 budling blocks and B2.

3.4. Designing of new polymers

However, designing the polymers with desired functionality is challenging due to their vast chemical space [55–57]. Their rapid and rational designing with optimum functionality demands the practical knowledge of molecular chemistry and material processing [58–61]. This is again labor-intensive and expensive further limiting their on-demand optimum synthesis with specific features. Due to versatile nature of organic building blocks, it is possible to design large number of monomers. Breaking Retrosynthetically Interesting Chemical Substructures (BRICS) method was employed to design new monomers using previously searched and generated buildings blocks. We have generated 10,000 monomers. The λ_{max} of generated monomers are predicted random forest regression model. The distribution of predicted λ_{max} values is given in Fig. 12. 21 monomers are selected on the basis of predicted λ_{max}. Monomers is red-shifted absorption are considered. Their structures are given in Fig. 13.

Additionally, the chemical similarity between selected monomers is computed. The heatmap of chemical similarity is shown in Fig. 14. It is indicating high similarity between selected monomers and relatively less similarity between SP1 and selected monomers. The similarity is also presented in the form of structures (Fig. 15). The green portion of monomers is similar to reference SP1.

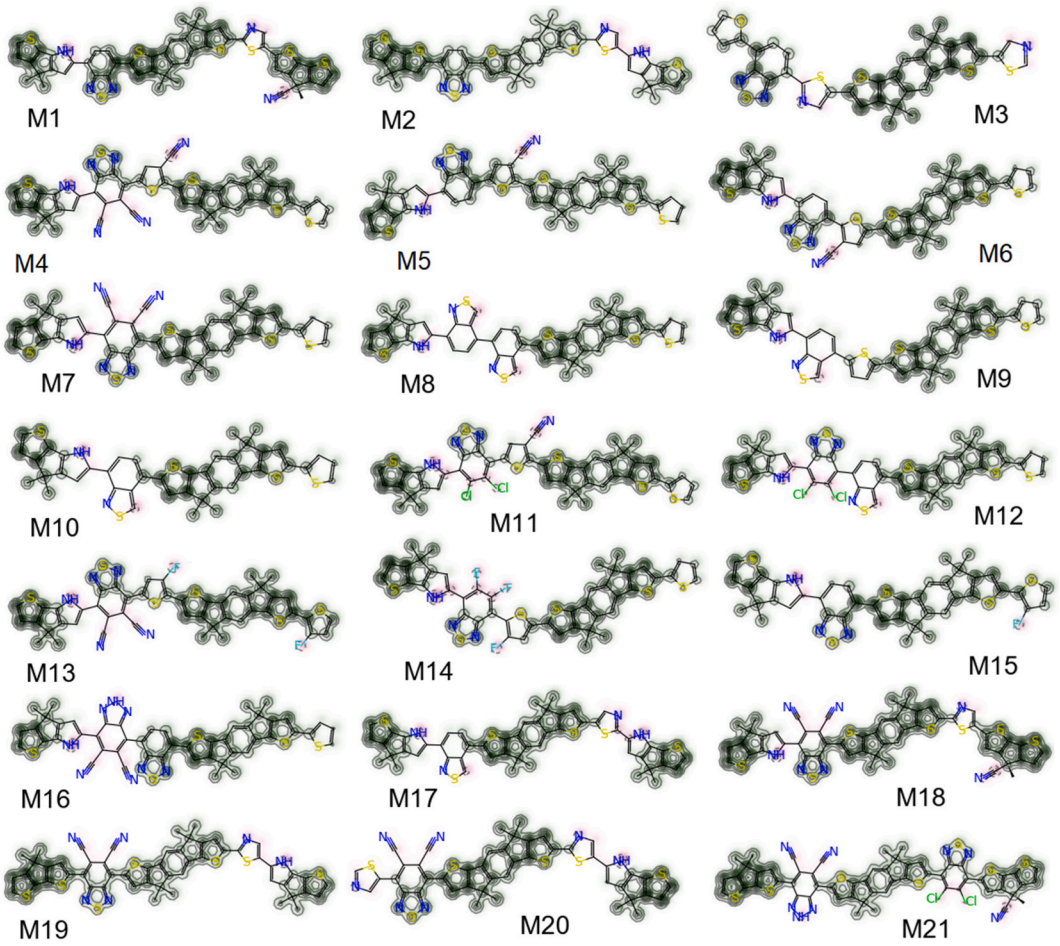


Fig. 15. Graphical presentation of structural similarity between selected monomers.

Table 2
Synthetic accessibility score of selected monomers.

Name	Synthetic accessibility score
M1	5.17
M2	4.57
M3	3.89
M4	4.62
M5	4.35
M6	4.36
M7	4.33
M8	4.46
M9	4.24
M10	4.12
M11	4.51
M12	4.57
M13	4.71
M14	4.47
M15	4.2
M16	4.63
M17	4.73
M18	5.42
M19	4.87
M20	4.62
M21	5.52

3.5. Synthetic accessibility

The synthetic accessibility score (SAS) is a measure of the ease with which a molecule can be synthesized. It considers all the possible factors including availability and cost of starting materials, number of synthetic steps involved, and the possibility of side reactions taking place during synthesis. We have calculated synthetic accessibility score using RDkit. Results are given Table 2. Synthetic accessibility score values fall between 1 (easy to synthesize) and 10 (difficult synthesize). 6 is considered as a threshold to distinguish between easy to synthesize and difficult synthesize [62]. All the selected monomers have synthetic accessibility score values lower 6. It is indicating that these compounds are easy to synthesize.

In present study, only single property is predicted, and monomers are screened on the basis of single property. The availability of data of other related properties can increase the chances of more successful screening. Extensive research can increase the development in this field.

4. Conclusions

The present study uses machine learning models to predict the light absorption behavior of polymers. Through the utilization of molecular descriptors, these machine learning models undergo training. In parallel, the building blocks are retrieved from chemical databases. Complementing this, new building blocks are generated by enumerating chemical libraries. While, the BRICS method is used to design over 10000 new polymers. Both the searched and designed budling blocks are used as input. Using the best machine learning model, the UV/visible absorption maxima of designed polymers are predicted. Chemical similarity analysis is also performed on the selected polymers. Simultaneously, synthetic accessibility of selected polymers is calculated. It is worth noting that the synthesis of all these polymers is straightforward. The method is an effective tool for creating high-performance UV/visible monomers for OSCs, and it can accelerate the search for novel materials in this field.

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Muhammad Ishfaq: Conceptualization. **Tayyaba Mubashir:** Methodology, Software. **Safaa N. Abdou:** Formal analysis. **Mudassir Hussain Tahir:** Data curation, Formal analysis. **Mohamed Ibrahim Halawa:** Investigation, Methodology, Resources, Software. **Mohamed M. Ibrahim:** Writing – original draft. **Yulin Xie:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project is supported by the Huanggang Normal University Project (2042022005, 30120230102). The researchers would like to acknowledge Deanship of Scientific Research, Taif University for funding this work.

References

[1] X. Zhen, K. Pu, X. Jiang, Photoacoustic imaging and photothermal therapy of semiconducting polymer nanoparticles: signal amplification and second near-infrared construction, Small (2021), e2004723.
[2] Y. Zhu, C. Gu, Y. Miao, B. Yu, Y. Shen, H. Cong, D–A polymers for fluorescence/photoacoustic imaging and characterization of their photothermal properties, J. Mater. Chem. B 7 (2019) 6576–6584.
[3] J. Zhang, X. Zhen, P.K. Upputuri, M. Pramanik, P. Chen, K. Pu, Activatable photoacoustic nanoprobes for in vivo ratiometric imaging of peroxynitrite, Adv. Mater. 29 (2017), 1604764.
[4] Q. Fan, K. Cheng, Z. Yang, R. Zhang, M. Yang, X. Hu, X. Ma, L. Bu, X. Lu, X. Xiong, W. Huang, H. Zhao, Z. Cheng, Perylene-diimide-based nanoparticles as highly efficient photoacoustic agents for deep brain tumor imaging in living mice, Adv. Mater. 27 (2015) 843–847.
[5] K. Pu, A.J. Shuhendler, J. Rao, Semiconducting polymer nanoprobe for in vivo imaging of reactive oxygen and nitrogen species, Angew. Chem., Int. Ed. Engl. 52 (2013) 10325–10329.
[6] G. Hong, Y. Zou, A.L. Antaris, S. Diao, D. Wu, K. Cheng, X. Zhang, C. Chen, B. Liu, Y. He, J.Z. Wu, J. Yuan, B. Zhang, Z. Tao, C. Fukunaga, H. Dai, Ultrafast fluorescence imaging in vivo with conjugated polymer fluorophores in the second near-infrared window, Nat. Commun. 5 (2014) 4206.
[7] S.U.-D. Khan, A. Mahmood, U.A. Rana, S. Haider, Utilization of electron-deficient thiadiazole derivatives as π-spacer for the red shifting of absorption maxima of diarylamine-fluorene based dyes, Theor. Chem. Acc. 134 (2014) 1596.
[8] M.R. Janjua, A. Mahmood, M.F. Nazar, Z. Yang, S. Pan, Electronic absorption spectra and nonlinear optical properties of ruthenium acetylide complexes: a DFT study toward the designing of new high NLO response compounds, Acta Chim. Slov. 61 (2014) 382–390.
[9] C. Dou, X. Long, Z. Ding, Z. Xie, J. Liu, L. Wang, An electron-deficient building block based on the B←N unit: an electron acceptor for all-polymer solar cells, Angew. Chem. Int. Ed. 55 (2016) 1436–1440.
[10] C.M. MacNeill, R.C. Coffin, D.L. Carroll, N.H. Levi-Polyachenko, Low band gap donor-acceptor conjugated polymer nanoparticles and their NIR-mediated thermal ablation of cancer cells, Macromol. Biosci. 13 (2013) 28–34.
[11] Y. Jiang, K. Pu, Advanced photoacoustic imaging applications of near-infrared absorbing organic nanoparticles, Small 13 (2017), 1700710.

[12] R. Hussain, F. Hassan, M.U. Khan, M.Y. Mehboob, R. Fatima, M. Khalid, K. Mahmood, C.J. Tariq, M.N. Akhtar, 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. Quant. Electron. 52 (2020) 364.

[13] R. Hussain, M.Y. Mehboob, M.U. Khan, M. Khalid, Z. Irshad, R. Fatima, A. Anwar, S. Nawab, M. Adnan, Efficient designing of triphenylamine-based hole transport materials with outstanding photovoltaic characteristics for organic solar cells, J. Mater. Sci. 56 (2021) 5113–5131.

[14] L. Wang, Q. An, L. Yan, H.-R. Bai, M. Jiang, A. Mahmood, C. Yang, H. Zhi, J.-L. Wang, Non-fullerene acceptors with hetero-dihalogenated terminals induce significant difference in single crystallography and enable binary organic solar cells with 17.5% efficiency, Energy Environ. Sci. 15 (2022) 320–333.

[15] M.H. Tahir, T. Mubashir, T.-U.-H. Shah, A. Mahmood, Impact of electron-withdrawing and electron-donating substituents on the electrochemical and charge transport properties of indacenodithiophene-based small molecule acceptors for organic solar cells, J. Phys. Org. Chem. 32 (2019), e3909.

[16] H.-R. Bai, Q. An, H.-F. Zhi, M. Jiang, A. Mahmood, L. Yan, M.-Q. Liu, Y.-Q. Liu, Y. Wang, J.-L. Wang, A random terpolymer donor with similar monomers enables 18.28% efficiency binary organic solar cells with well polymer batch reproducibility, ACS Energy Lett. (2022) 3045–3057.

[17] M. Khalid, A. Ali, M.U. Khan, M.N. Tahir, A. Ahmad, M. Ashfaq, R. Hussain, S.F.d.A. Morais, A.A.C. Braga, Non-covalent interactions abetted supramolecular arrangements of N-Substituted benzylidene acetohydrazide to direct its solid-state network, J. Mol. Struct. 1230 (2021), 129827.

[18] M. Khalid, M.U. Khan, S. Ahmed, Z. Shafiq, M.M. Alam, M. Imran, A.A.C. Braga, M.S. Akram, Exploration of promising optical and electronic properties of (non-polymer) small donor molecules for organic solar cells, Sci. Rep. 11 (2021), 21540.

[19] M. Ishfaq, Z. Rahman, M. Aamir, I. Ali, Y. Guan, Z. Hu, Insight into potent TLR2 inhibitors for the treatment of disease caused by Mycoplasma pneumoniae based on machine learning approaches, Mol. Divers. 27 (2023) 371–387.

[20] T. Mubashir, M. Hussain Tahir, Y. Altaf, F. Ahmad, M. Arshad, A. Hakamy, M. Sulaman, Statistical analysis and visualization of data of non-fullerene small molecule acceptors from Harvard organic photovoltaic database. Structural similarity analysis with famous non-fullerene small molecule acceptors to search new building blocks, J. Photochem. Photobiol., A 437 (2023), 114501.

[21] A. Mahmood, A. Irfan, J.-L. Wang, 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 10 (2022) 4170–4180.

[22] M. Ishfaq, M. Aamir, F. Ahmad, M.M. A, S. Elshahat, Machine learning-assisted prediction of the biological activity of aromatase inhibitors and data mining to explore similar compounds, ACS Omega 7 (2022) 48139–48149.

[23] Z. Zhu, Z. Rahman, M. Aamir, S.Z.A. Shah, S. Hamid, A. Bilawal, S. Li, M. Ishfaq, Insight into TLR4 receptor inhibitory activity via QSAR for the treatment of Mycoplasma pneumonia disease, RSC Adv. 13 (2023) 2057–2069.

[24] Y. Zhang, X. Xu, Machine learning glass transition temperature of styrenic random copolymers, J. Mol. Graphics Modell. 103 (2021), 107796.

[25] Y. Zhang, X. Xu, Machine learning glass transition temperature of polyacrylamides using quantum chemical descriptors, Polym. Chem. 12 (2021) 843–851.

[26] Y. Zhang, X. Xu, Machine learning decomposition onset temperature of lubricant additives, J. Mater. Eng. Perform. 29 (2020) 6605–6616.

[27] Y. Zhang, X. Xu, Machine learning glass transition temperature of polymers, Heliyon 6 (2020), e05055.

[28] J.A. Bhutto, T. Mubashir, M.H. Tahir, Hafsa, F. Ahmad, S.R.M. Sayed, H.O. El-ansary, M. Ishfaq, Virtual screening and library enumeration of new hydroxycinnamates based antioxidant compounds: a complete framework, J. Saudi Chem. Soc. 27 (2023), 101670.

[29] M. Ishfaq, S.Z.A. Shah, I. Ahmad, Z. Rahman, Multinomial classification of NLRP3 inhibitory compounds based on large scale machine learning approaches, Mol. Divers. (2023). https://doi.org/10.1007/s11030-023-10690-y.

[30] A. Mauri, V. Consonni, M. Pavan, R. Todeschini, DRAGON software: an easy approach to molecular descriptor calculations, MATCH Commun. Math. Comput. Chem. 56 (2006) 237–248.

[31] G. Landrum, RDKit: Open-source cheminformatics. http://www.rdkit.org.

[32] D. Rogers, M. Hahn, Extended-connectivity Fingerprints, J. Chem. Inf. Model. 50 (2010) 742–754.

[33] T. Sander, J. Freyss, M. von Korff, C. Rufener, DataWarrior: an open-source program for chemistry aware data visualization and analysis, J. Chem. Inf. Model. 55 (2015) 460–473.

[34] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, M. Rarey, On the art of compiling and using 'drug-like' chemical fragment spaces, ChemMedChem 3 (2008) 1503–1507.

[35] A. Mahmood, A. Irfan, J.-L. Wang, Machine learning for organic photovoltaic polymers: a minireview, Chin. J. Polym. Sci. 40 (2022) 870–876.

[36] A. Mahmood, Y. Sandali, J.-L. Wang, Easy and fast prediction of green solvents for small molecule donor-based organic solar cells through machine learning, Phys. Chem. Chem. Phys. 25 (2023) 10417–10426.

[37] M.U. Khan, M. Khalid, R. Hussain, A. Umar, M.Y. Mehboob, Z. Shafiq, M. Imran, A. Irfan, Novel W-shaped oxygen heterocycle-fused fluorene-based non-fullerene acceptors: first theoretical framework for designing environment-friendly organic solar cells, Energy Fuel. 35 (2021) 12436–12450.

[38] A. Mahmood, M. Saqib, M. Ali, M.I. Abdullah, B. Khalid, Theoretical investigation for the designing of novel antioxidants, Can. J. Chem. 91 (2013) 126–130.

[39] M.U. Khan, M.Y. Mehboob, R. Hussain, R. Fatima, M.S. Tahir, M. Khalid, A.A.C. Braga, Molecular designing of high-performance 3D star-shaped electron acceptors containing a truxene core for nonfullerene organic solar cells, J. Phys. Org. Chem. 34 (2021) e4119.

[40] A. Mahmood, S.U.-D. Khan, U.A. Rana, M.R.S.A. Janjua, M.H. Tahir, M.F. Nazar, Y. Song, 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. 28 (2015) 418–422.

[41] S. Hussain, S. Ali, S. Shahzadi, M.N. Tahir, M. Shahid, Synthesis, characterization, single crystal XRD and biological screenings of organotin(IV) derivatives with 4-(2-hydroxyethyl)piperazine-1-carbodithioic acid, J. Coord. Chem. 69 (2016) 687–703.

[42] A. Mahmood, A. Irfan, J.-L. Wang, 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. 28 (2022), e202103712.

[43] M.U. Khan, R. Hussain, M.Y. Mehboob, M. Khalid, M.A. Ehsan, A. Rehman, M.R.S.A. Janjua, First theoretical framework of Z-shaped acceptor materials with fused-chrysene core for high performance organic solar cells, Spectrochim. Acta: Mol. Biomol. Spectrosc. 245 (2021), 118938.

[44] M.U. Khan, R. Hussain, M. Yasir Mehboob, M. Khalid, Z. Shafiq, M. Aslam, A.A. Al-Saadi, S. Jamil, M.R.S.A. Janjua, In silico modeling of new "Y-Series"-Based near-infrared sensitive non-fullerene acceptors for efficient organic solar cells, ACS Omega 5 (2020) 24125–24137.

[45] S. Hussain, S. Ali, S. Shahzadi, M.N. Tahir, S. Ramzan, M. Shahid, Synthesis, spectroscopic characterization, X-ray crystal structure and biological activities of homo- and heterobimetallic complexes with potassium-1-dithiocarboxylatopiperidine-4-carboxylate, Polyhedron 119 (2016) 483–493.

[46] M. Khalid, M.U. Khan, E.-t. Razia, Z. Shafiq, M.M. Alam, M. Imran, M.S. Akram, Exploration of efficient electron acceptors for organic solar cells: rational design of indacenodithiophene based non-fullerene compounds, Sci. Rep. 11 (2021), 19931.

[47] M. Khalid, Momina, M. Imran, M.F.u. Rehman, A.A.C. Braga, M.S. Akram, Molecular engineering of indenoindene-3-ethylrodanine acceptors with A2-A1-D-A1-A2 architecture for promising fullerene-free organic solar cells, Sci. Rep. 11 (2021), 20320.

[48] S. Hussain, S. Ali, S. Shahzadi, S.K. Sharma, K. Qanungo, I.H. Bukhari, Homobimetallic complexes containing Sn(IV) with acetylene dicarboxylic acid: their syntheses and structural interpretation by spectroscopic, semi-empirical, and DFT techniques, J. Coord. Chem. 65 (2012) 278–285.

[49] K. Pu, J. Mei, J.V. Jokerst, G. Hong, A.L. Antaris, N. Chattopadhyay, A.J. Shuhendler, T. Kurosawa, Y. Zhou, S.S. Gambhir, Z. Bao, J. Rao, Diketopyrrolopyrrole-based semiconducting polymer nanoparticles for in vivo photoacoustic imaging, Adv. Mater. 27 (2015) 5184–5190.

[50] S.A. Lopez, E.O. Pyzer-Knapp, G.N. Simm, T. Lutzow, K. Li, L.R. Seress, J. Hachmann, A. Aspuru-Guzik, The Harvard organic photovoltaic dataset, Sci. Data 3 (2016), 160086.

[51] L. Ruddigkeit, R. van Deursen, L.C. Blum, J.-L. Reymond, Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17, J. Chem. Inf. Model. 52 (2012) 2864–2875.

[52] M.U. Khan, M. Khalid, M.N. Arshad, M.N. Khan, M. Usman, A. Ali, B. Saifullah, Designing star-shaped subphthalocyanine-based acceptor materials with promising photovoltaic parameters for non-fullerene solar cells, ACS Omega 5 (2020) 23039–23052.

[53] M.I. Abdullah, M.R.S.A. Janjua, M.F. Nazar, A. Mahmood, Quantum chemical designing of efficient TC4-based sensitizers by modification of auxiliary donor and π-spacer, Bull. Chem. Soc. Jpn. 86 (2013) 1272–1281.

[54] S. Hussain, S. Ali, S. Shahzadi, M.N. Tahir, M. Shahid, K.S. Munawar, S.M. Abbas, Synthesis, spectroscopy, single crystal XRD and biological studies of multinuclear organotin dicarboxylates, Polyhedron 117 (2016) 64–72.
[55] A. Mahmood, A. Irfan, J.-L. Wang, Molecular level understanding of the chalcogen atom effect on chalcogen-based polymers through electrostatic potential, non-covalent interactions, excited state behaviour, and radial distribution function, Polym. Chem. 13 (2022) 5993–6001.
[56] M.R.S.A. Janjua, Deciphering the role of invited guest bridges in non-fullerene acceptor materials for high performance organic solar cells, Synth. Met. 279 (2021), 116865.
[57] X. Zhao, Q. An, H. Zhang, C. Yang, A. Mahmood, M. Jiang, M.H. Jee, B. Fu, S. Tian, H.Y. Woo, Y. Wang, J.-L. Wang, Double asymmetric core optimizes crystal packing to enable selenophene-based acceptor with over 18 % efficiency in binary organic solar cells, Angew. Chem. Int. Ed. 62 (2023), e202216340.
[58] M.R.S.A. Janjua, A. Mahmood, F. Ahmad, Solvent effects on nonlinear optical response of certain tetramminеruthenium(II) complexes of modified 1,10-phenanthrolines, Can. J. Chem. 91 (2013) 1303–1309.
[59] M.R.S.A. Janjua, Impact of symmetry breaking on the performance of non-fullerene acceptors (NFAs) for photo and thermally stable organic solar cells (OSCs): a DFT-based interrogation and investigation, J. Photochem. Photobiol., A 444 (2023), 115003.
[60] A. Mahmood, J.-Y. Hu, B. Xiao, A. Tang, X. Wang, E. Zhou, Recent progress in porphyrin-based materials for organic solar cells, J. Mater. Chem. A 6 (2018) 16769–16797.
[61] L. Yan, H. Zhang, Q. An, M. Jiang, A. Mahmood, M.H. Jee, H.-R. Bai, H.-F. Zhi, S. Zhang, H.Y. Woo, J.-L. Wang, Regioisomer-Free difluoro-monochloro terminal-based hexa-halogenated acceptor with optimized crystal packing for efficient binary organic solar cells, Angew. Chem. Int. Ed. 61 (2022), e202209454.
[62] P. Ertl, A. Schuffenhauer, Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions, J. Cheminf. 1 (2009) 8.