ISSN 1027-5495. Functional Materials, 33, No.3 (2026), p. 544-552
doi:http://dx.doi.org/10.15407/fm33.03.544 © 2026 — SSI "Institute for Single Crystals"
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Machine learning–driven design of π-conjugated systems in isomer space with target absorption spectra: a case study of Effenberger dye

M. O. Kyrpa, I. V. Khristenko, V. V. Ivanov*

School of Chemistry, V. N. Karazin Kharkiv National University,
Svobody sq., 4, Kharkiv, 61022, Ukraine
Corresponding author e-mail: *vivanov@karazin.ua*

Received April 9, 2026 , approved June 12, 2026

This work presents a proof-of-concept study exploring the isomer space of the Effenberger dye to evaluate approaches for predicting optical properties aimed at the design of new compounds. A diverse library of isomers was generated, differing from the reference dye, and quantum chemical parameters were calculated using the Tight Binding Density Functional Theory (DFTB) method, a computationally efficient approximation to TD-DFT. This allowed the evaluation of a large number of molecules with reasonable accuracy. For each isomer, a broad set of molecular descriptors was calculated, including geometric, electronic, and topological parameters. The Boruta algorithm was applied to select the most relevant descriptors, reducing dimensionality and avoiding overfitting. Regression analysis was used as a simple baseline to correlate descriptors with long-wavelength transition energies. An artificial neural network (ANN) was then constructed using TensorFlow/Keras, with three hidden layers, ReLU activations, a linear output, and regularization methods. The model was trained on a representative subset, while the remaining molecules (30%) formed the test set. The ANN achieved high predictive performance, showing good agreement between predicted and reference TD-DFT values. This approach demonstrates that even a relatively small but diverse set of isomers can be used to build effective predictive models. The combination of DFTB, molecular descriptors, Boruta-based feature selection, and ANN modeling provides a versatile framework for designing new π-conjugated materials with tailored optical properties.

Keywords: QSAR, regression model, ANN, Effenberger dye, UV-Vis absorption

Дизайн молекул із заданими спектрами поглинання на основі методів машинного навчання у просторі ізомерів π-спряжених систем: приклад барвника Еффенбергера. *М. О. Кирпа, І. В. Христенко, В. В. Іванов*

Робота представляє собою дослідження підходу до прогнозування оптичних властивостей з метою дизайну нових сполук, проведене на прикладі ізомерного простору, сформованого на основі барвника Еффенбергера. Було створено бібліотеку ізомерів. Квантовохімічні параметри барвників було розраховано за допомогою методу теорії функціоналу густини щільного зв'язку (DFTB) — обчислювально-ефективного наближення до TD-DFT. Для кожного ізомеру було розраховано широкий набір молекулярних дескрипторів, включаючи геометричні, електронні та топологічні параметри. Для вибору найбільш релевантних дескрипторів було застосовано алгоритм Boruta, що дозволило зменшити розмірність дескрипторного набору та уникнути перенавчання. Регресійний аналіз було використано як просту базову модель кореляції дескрипторів з енергіями довгохвильових переходів. Далі, за допомогою пакету TensorFlow/Keras було побудовано штучну нейронну мережу з трьома

прихованими шарами, активаціями ReLU, лінійним виходом та методами регуляризації. Модель було навчено на репрезентативній підмножині, тоді як решта молекул (30%) утворила тестову множину. Нейронна мережа досягла високої прогностичної ефективності, що характеризується доброю збіжністю між прогнозованими та еталонними значеннями TD-DFT. Цей підхід демонструє, що навіть відносно невеликий, але різноманітний набір ізомерів може бути використаний для побудови прогностичних моделей. Поєднання DFTB, молекулярних дескрипторів, вибору ознак на основі Boruta та нейронно-мережеве моделювання забезпечує працездатну основу для проектування нових π-спряжених матеріалів із заданими оптичними властивостями.

1. Introduction

The design of new materials with predefined properties remains a central challenge in modern materials science [1-4]. Numerous studies have addressed this problem; however, the majority focuses on the prediction and design of molecular biological activity rather than physicochemical or optical properties. These studies span a broad spectrum of methodologies, ranging from conventional statistical approaches based on regression analysis to modern machine learning techniques, including artificial neural networks (ANNs).

In contrast, the design of molecular systems with targeted optical properties remains significantly less explored, largely due to the complexity of structure–property relationships and the vastness of chemical space. In this work, we introduce a strategy that restricts the search domain to the isomer space of a given molecular framework. In this approach, molecules share a common structural scaffold while differing in the arrangement of substituents or connectivity, allowing for systematic exploration of structure–property relationships.

One of the key challenges in constructing reliable QSAR (quantitative structure–activity relationship) models is the selection of appropriate training data. In problems involving the establishment of structure–spectrum relationships, experimental datasets would ideally serve this purpose. However, there is often a lack of sufficiently systematic and consistent experimental data suitable for building predictive models. Therefore, as a practical alternative, we propose to use the results of time-dependent density functional theory (TD-DFT) calculations in vacuum as the target property. This machine learning–based approach enables high-throughput screening of large molecular libraries to identify candidates with target properties.

In our approach, a comprehensive set of molecular descriptors is initially generated, followed by feature selection using the Boruta algorithm [5] to identify the most relevant variables for training a machine learning model.

The aim of this work is to develop a neural network–based framework for the design of π-conjugated systems with target absorption spectra within a restricted isomer space. The proposed approach combines quantum chemical data obtained at the TD-DFT level with a comprehensive descriptor representation and Boruta-based feature selection to construct predictive models linking molecular structure to optical properties. The novelty of this study lies in the integration of design principles with an isomer-restricted chemical space, enabling systematic exploration of structure–spectrum relationships.

The applicability of the proposed methodology is demonstrated for a series of isomeric derivatives of an Effenberger dye (ED) [6], serving as a representative model system. We intentionally restrict the present study to a dye with a minimal thiophene-based framework. Specifically, the structure comprises only two thiophene rings bearing donor and acceptor π-substituents (Fig. 1), which defines a compact and well-controlled isomer space. This restriction is critical for the rigorous validation of the proposed methodology, as it enables an exhaustive or near-exhaustive exploration of all accessible isomeric forms without introducing combinatorial complexity. At the same time, this choice does not call into question the broader relevance of the study, since thiophene-containing dyes remain a central class of functional materials, extensively investigated as solvatochromic probes [7-11], sensitizers for solar cells [12-15], emitters in light-emitting diodes [16-19], and promising candidates with pronounced nonlinear optical (NLO) responses [20-24].

The QSAR predictive model was developed using a range of statistical and machine learning approaches implemented in Python within our in-house QUASAR software suite [25-29]. The workflow included data preprocessing, descriptor calculation and selection, model training, and validation. The results were evalu-

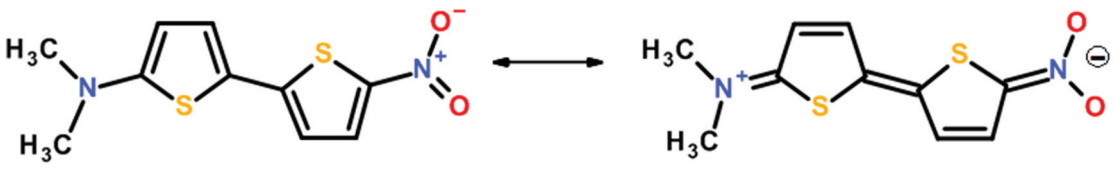

Figure 1. Two limiting mesomeric forms of the Effenberger dye.

ated to ensure the robustness and predictive performance of the models. Model quality was assessed using standard statistical metrics and appropriate validation strategies. All computational procedures were carried out within the QUASAR framework.

2. Computational methods and workflow

Thus, this study focuses on a set of ED (see Fig. 1) isomers, composed of two thiophene rings functionalized with dimethylamino and nitro substituents. The algorithm for constructing the corresponding isomeric chemical space within the QUASAR package has been described in detail in our previous work [27], where the complete list of generated structures is also provided. In total, 59 distinct isomers were obtained (Fig. 2).

The predictive model is based on initial calculations of system parameters from a training set, using quantum chemical data obtained via Tight-Binding Density Functional Theory (DFTB). This efficient method implemented in the DFTB+ program [30-32] enables rapid evaluation of a large number of molecular structures. The target property for the machine learning model is the long-wavelength transition energy (λ, eV) in vacuum, estimated using TD-DFT calculations performed with the ORCA software package [33-34].

To develop a predictive model for describing TD-DFT excitation energy, a wide range of descriptors was used, divided into three groups: DFTB energy descriptors, structural descriptors calculated in QUASAR package, and structural descriptors calculated using the PaDEL-Descriptor program [35].

The DFTB energy descriptors include the HOMO-LUMO energy gap and excitation energy calculated using the mio-1-1 Slater-Koster parameter set [36].

A standard set of 1D and 2D molecular descriptors, as provided by the PaDEL-Descriptor program [35], was calculated, including up to a thousand descriptors.

Also several additional structural descriptors were calculated for the optimal geometry obtained in the DFTB calculation. The structural molecular descriptors include the non-planarity index NPI, elongation R, width D, and sphericity I. From general considerations, some of the structures from library are expected to exhibit significant steric strain. We deliberately retain such structures in order to construct a representative training set for subsequent neural network modeling. To quantify steric distortion, we introduce a new non-planarity index (NPI) [28], defined as the mean deviation of heavy (non-hydrogen) atoms from the best-fit molecular plane

$$\beta_0 + \beta_x x + \beta_y y + \beta_z z = 0\,. \qquad (1)$$

The NPI is computed using the orthogonal distance least squares (ODLS) method, where the parameters of the plane $\beta = (\beta_0, \beta_x, \beta_y, \beta_z)$ obtained by minimization

$$\beta = \arg\min_{\beta} \left(\beta^{+} A^{+} A\beta \,/\, \beta^{+} E\beta\right). \qquad (2)$$

Where matrix A and E are:

$$A = \begin{pmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_N & y_N & z_N \end{pmatrix}, \; E = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (3)$$

x_i, y_i, z_i are coordinates of heavy atoms, while N is total number of atoms. Correspondently NPI (Å) describes the average deviation of heavy atoms from the best-fit plane, equal to

$$NPI = \frac{1}{N} \sum_i \frac{\left|\beta_x x_i + \beta_y y_i + \beta_z z_i + \beta_0\right|}{\sqrt{\beta_x^2 + \beta_y^2 + \beta_z^2}}\,. \qquad (4)$$

The NPI values within the studied library range from 0 Å for planar molecules to 0.70 Å for the most non-planar ones.

Another index that we consider promising in the context of the "structure–spectra" problem is the Petitjean shape index (PI), which characterizes effective molecular dimensions in terms of the molecule diameter (D) and radius (R) [37]. To quantify spherical symmetry,

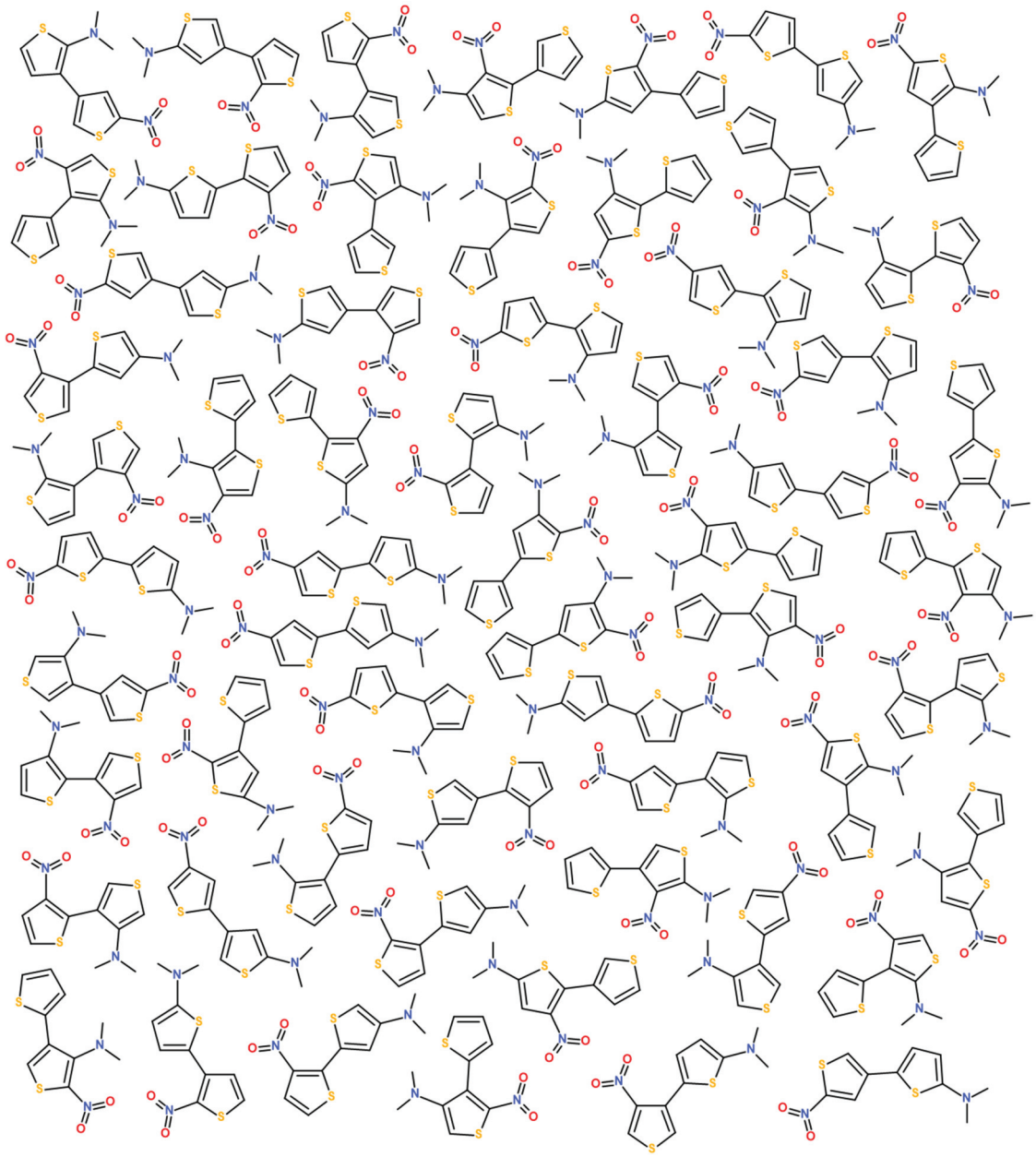

Figure 2. Collection of generated ED isomers used in this study.

the Petitjean index is calculated as:

$$PI = \frac{D - R}{R} \tag{5}$$

This value ranges from 0.53 for molecules with shapes intermediate between spherical and elongated (slightly elongated) to 0.91 for the most elongated molecules.

The total number of descriptors after preliminary selection is several hundred. Such a large number of descriptors may lead to overfitting and a consequent decrease in the predictive performance of the model. Therefore, feature selection was performed using the Boruta algorithm [5] to identify all relevant descriptors contributing to the predictive model. The Boruta method was employed to eliminate irrelevant variables and retain only those features that significantly influence model performance. Boruta-based feature selection ensured robust identification of important descriptors prior to model training, thereby improving the stability and interpretability of the results.

Neural network models were implemented using the TensorFlow framework with the Keras API [38-39] within the in-house

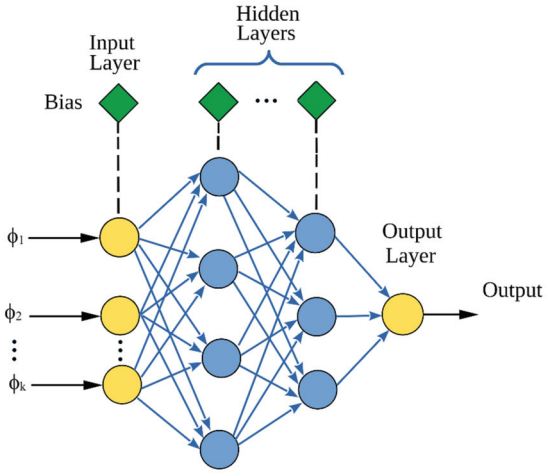

Figure 3. The architecture of ANN

QUASAR software suite. The network architecture comprised an input layer, three hidden layers with Rectified Linear Unit (ReLU) activation, and a linear output layer (Fig. 3).

The model was trained using the ADAM optimizer with the mean squared error (MSE) loss function. A batch size of 32 was employed during training. To improve generalization and mitigate overfitting, regularization techniques including dropout and L_2 regularization were applied. The architecture and hyperparameters were selected to ensure stable and accurate predictive performance.

3. Numerical results and tests

The generated library of isomers, despite its relatively small size, spans a reasonably broad range of structures that differ from the reference structure of the ED (Fig. 1). Analysis based on the NPI parameter (see the histogram in Fig. 4) shows that approximately 40% of the molecules are essentially planar (NPI ≤ 0.3), while about 10% exhibit significant non-planarity. This structural diversity enables the construction of a reasonably representative training set for both regression modeling and neural network training.

Initially, we consider a standard multilinear regression model implemented using the least squares method. Clearly, such an approach requires a minimal set of descriptors. In this case,

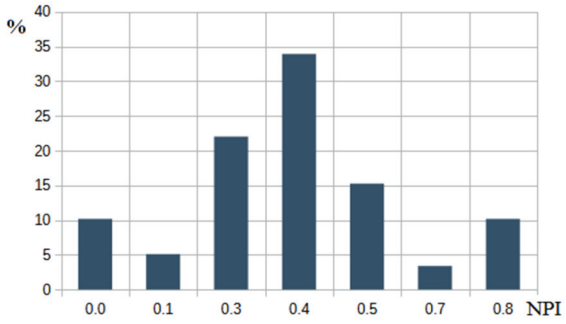

Figure 4. Distribution of NPI for the set of ED isomers

we employ a straightforward method based on analyzing the correlation between descriptors and the target property, namely the long-wavelength transition energy.

The best, although not perfect, correlations with the excitation energy obtained at the reference level of theory were found for the following descriptors: the HOMO–LUMO gap (E_g), VE3_Dzp, VE2_Dzp, and λ_0. The latter two descriptors (VE3_Dzp and VE2_Dzp) correspond to the logarithmic and average coefficient sums of the last eigenvector of the Barysz matrix [40], weighted by atomic polarizabilities, whereas λ_0 represents the excitation energy obtained from TD-DFTB calculations. The corresponding coefficients of determination are 0.611, 0.520, 0.499, and 0.498, respectively. Considering that descriptors can describe excitation energy not only linearly, in addition to training the neural network on the full set of descriptors, we manually calculated a series of nonlinear derivatives, such as $1/d$, $\ln d$, $1/\ln d$, $\sqrt{d}$, $1/\sqrt{d}$, d^2, $1/d^2$, d^3, $1/d^3$, where d is any descriptor from the set. Table 1 lists the descriptors that most strongly correlate with the property being described.

It should be emphasized that the excitation energies obtained using TD-DFTB (λ_0) show rather weak correlation with those computed at the higher level of theory, TD-DFT (λ). Based on the extended set (a total of 89 descriptors), a series of regression equations was constructed, followed by cross-validation using the Leave-One-Out method (CV-LOO) [41]. The best equation for approximation of the TD-DFT

Table 1. Descriptors with best correlations with the excitation energy obtained from TD-DFT.

	$\ln E_g$	$\frac{1}{E_g}$	E_g	$\frac{1}{\ln E_g}$	$\sqrt{VE2_Dzp}$	$\frac{1}{\ln VE2_Dzp}$	$VE3_Dzp$	$VE2_Dzp$	λ_0
R^2	0.634	0.634	0.611	0.569	0.523	0.522	0.520	0.499	0.498

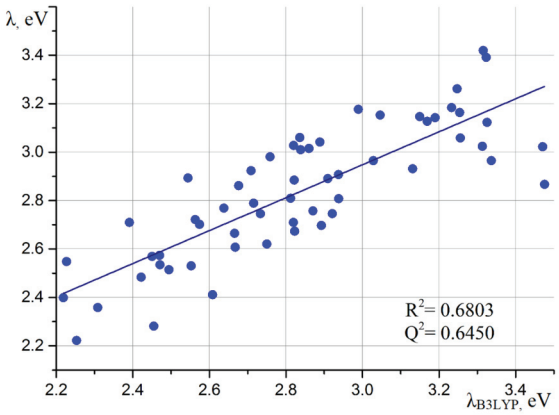

Figure 5. Dependence of excitation energies (λ eV) calculated using Eq. (6) on TD-DFT/B3LYP values ($\lambda_{B3LYP,}$ eV).

excitation energy (λ, eV) was obtained based on the energy gap (E_g) and VE2_Dzp:

$$\lambda = 18.89 + 1.15 \ln E_g + 134.79\sqrt{VE2_Dzp} + \frac{140.94}{\ln VE2_Dzp}, \quad (6)$$

with a coefficient of determination R^2 of 0.680 and a CV-LOO coefficient of determination Q^2 of 0.645. The corresponding values were calculated as follows (Eq. 7):

$$R^2 = 1 - \sum_i^N \frac{(\lambda_i - \hat{\lambda}_i)^2}{(\lambda_i - \bar{\lambda})^2}, \qquad Q^2 = 1 - \sum_i^N \frac{(\lambda_i - \hat{\lambda}_{i/i})^2}{(\lambda_i - \bar{\lambda})^2}, \quad (7)$$

where the values for N molecules in the training set were obtained using the developed model ($\hat{\lambda}_i$) and the LOO procedure ($\hat{\lambda}_{i/i}$). According to the LOO procedure, the $\hat{\lambda}_{i/i}$ value was obtained from a model calibrated on the training set with the *i*-th molecule excluded. The root mean square deviation (RMSD) values for the calculated and LOO data are 0.18 and 0.19, respectively.

The values and close agreement between the two indices (R^2 and Q^2) indicates that the model exhibits not optimal, but satisfactory predictive ability. The corresponding graph is shown in Fig. 5. Since the plots obtained using Eq. 6 and the LOO procedure are visually indistinguishable, only one graph is presented.

The use of a neural network significantly enhanced the predictive capabilities of the model. For model evaluation, 30% of the molecules were randomly assigned to the test set, while the remaining molecules formed the training set. Feature selection using the Boruta algorithm was then performed to identify a minimal set of relevant descriptors, which included the following rather small set of parameters: λ_0, HOMO-LUMO gap, NPI, R, D, AATS8v, MATS8p^2, MATS8p^3. The use of such a relatively small set of descriptors helps to avoid overfitting.

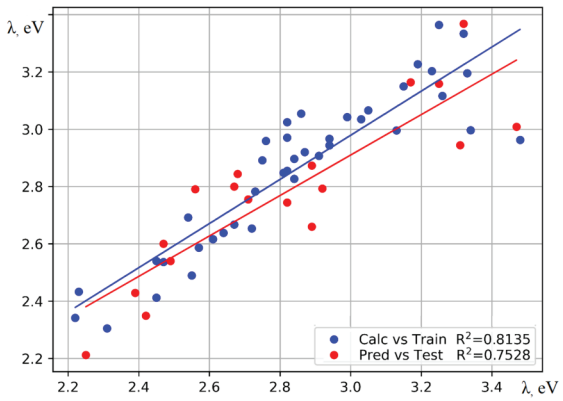

Figure 6. Comparison of ANN-predicted spectral transition energies with training set values (blue circles) and test set values (red circles).

Fig. 6 presents a comparison of spectral transition energies for both the training and test sets. The blue circles correspond to ANN-calculated energies versus TD-DFT/B3LYP values for molecules in the training set, while the red circles indicate predicted energies for molecules in the test set. The close agreement between calculated and predicted values demonstrates the ability of the ANN model to capture subtle structure-property relationships across the entire dataset, even for molecules not included in the training set.

The relatively high coefficients of determination (R^2) and the small difference between training and test sets (~0.06) indicate the efficiency and robustness of the obtained model. The relatively small angle (approximately 8°) of deviation between these two linear trends also reflects the robustness of the obtained model. The RMSD values are 0.14 eV and 0.18 eV for the training and test sets, respectively.

Three representative structures from the test set are shown in Fig. 7, including the parent ED structure (A) and isomers with excitation energies about 2.7 eV (459 nm) and oscillator strengths of f~0.1.

Overall, it can be noted that the obtained ANN model, implemented within the QUASAR software suite in Python, possesses sufficient flexibility to develop predictive models for the design of new materials.

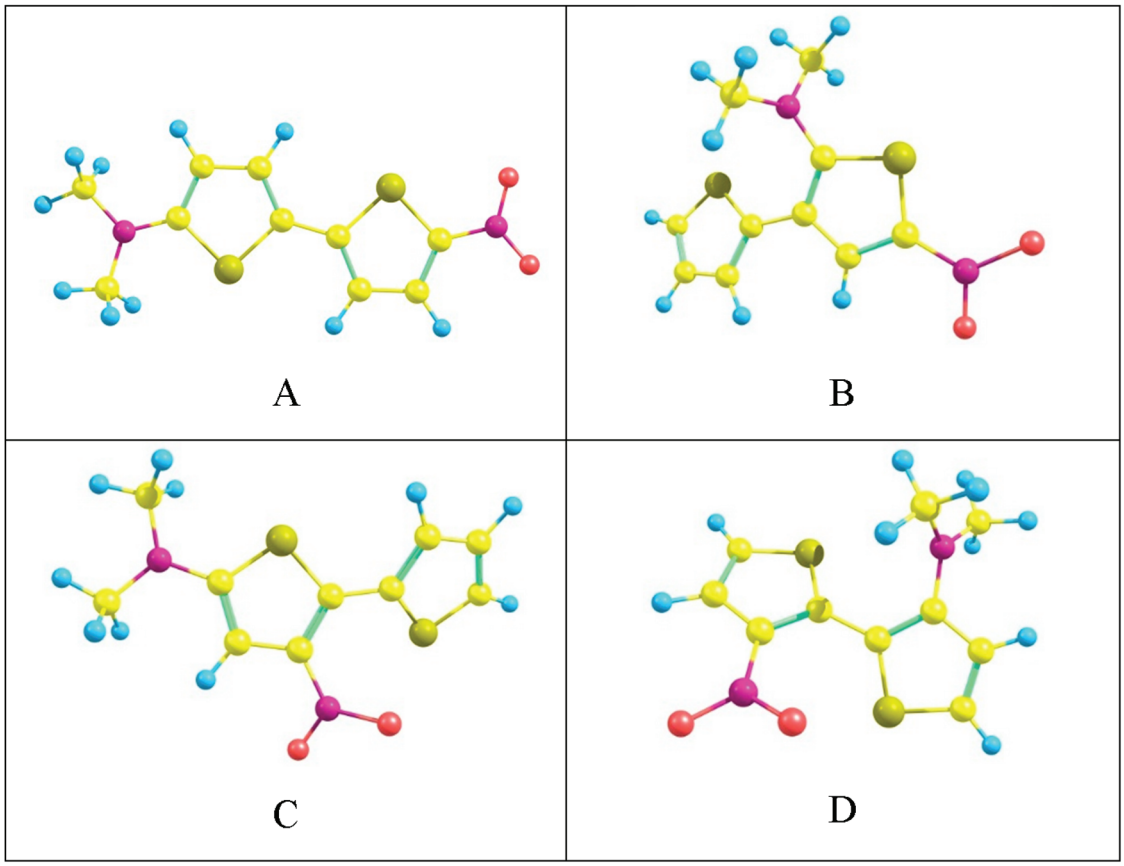

Figure 7. Representative structures from the test set; structure A corresponds to the parent ED molecule.

4. Conclusion

The proposed framework is inherently scalable and transferable, and is therefore expected to retain its predictive capability upon extension to larger chemical spaces and more structurally complex dye architectures, where exhaustive enumeration becomes infeasible.

It should be noted that the use of a relatively small training set of isomers represents a particularly challenging task for predictive modeling. In contrast to chemically diverse datasets, the considered systems exhibit closely related molecular structures, leading to limited structural variability. As a consequence, the descriptor space is weakly dispersed, which complicates the identification of robust structure–property relationships. This intrinsic similarity increases the risk of overfitting and reduces the generalizability of the model. In π-conjugated systems, the situation is further complicated by the fact that the target properties are governed by subtle variations in electronic structure, such as the degree of π-electron delocalization, orbital interactions, and charge-transfer effects. Therefore, the successful training and performance of the neural network under these conditions demonstrate the effectiveness of the proposed algorithm in capturing subtle structure–property relationships, highlighting its ability to operate reliably within a highly constrained and weakly diverse chemical space.

Of course, to describe the optical properties of a wide and diverse molecular library, the ANN must be properly trained.

Acknowledgements

This work is supported by the Ministry of Education and Science of Ukraine. Grant № 0126U001145 “Factors Affecting the Stability of Nanocarbon Structures in Liquids: The Basis for Developing High-Efficiency Current Sources and New Sorbents”.

References

1. K. Roy, A. Banerjee, Materials Informatics I: Methods, first ed., Springer Cham, 2025. https://doi.org/10.1007/978-3-031-78736-2
2. J. Neugebauer, T. Hickel, Density functional theory in materials science, Wiley Interdiscip. Rev. Comput. Mol. Sci. 3(5) (2013) 438-448. https://doi.org/10.1002/wcms.1125

3. F. Giustino, Materials modelling using density functional theory: properties and predictions, first ed., Oxford University Press, 2014.
4. M. Cheng, C. L. Fu, R. Okabe, A. Chotrattanapituk, A. Boonkird, N. T. Hung, M. Li, Artificial intelligence-driven approaches for materials design and discovery, Nat. Mater. 25 (2026) 1-17. https://doi.org/10.1038/s41563-025-02403-7
5. M. B. Kursa, W. R. Rudnicki, Feature selection with the Boruta package, J. Stat. Softw. 36 (2010) 1-13. https://doi.org/10.18637/jss.v036.i11
6. F. Effenberger, F. Würthner, 5-Dimethylamino-5′-nitro-2, 2′-bithiophene - a New Dye with Pronounced Positive Solvatochromism, Angew. Chem. Int. Ed. Engl. 32 (1993) 719. https://doi.org/10.1002/anie.199307191
7. C. E. de Melo, M. Dominguez, M. C. Rezende, V. G. Machado, Solvatochromism of dyes inspired in Effenberger's probe, Dyes Pigm. 184 (2021) 108757. https://doi.org/10.1016/j.dyepig.2020.108757
8. M. C. dos Santos, F. T. Souto, C. R. Nicoleti, M. C. Colaco, G. F. Caramori, V. G. Machado, Synthesis and solvatochromism of dipodal and tripodal dyes derived from Brooker's merocyanine, J. Mol. Liq. 386 (2023) 122476. https://doi.org/10.1016/j.molliq.2023.122476
9. R. Mera-Adasme, M. C. Rezende, M. Domínguez, On the physical-chemical nature of solvent polarizability and dipolarity, Spectrochim. Acta A Mol. Biomol. Spectrosc. 229 (2020) 118008. https://doi.org/10.1016/j.saa.2019.118008
10. B. Pigulski, J. Cichos, S. Szafert, Polyynes as Precursors of Photoluminescent Solvent Polarity Probes, ACS Sustain. Chem. Eng. 5(8) (2017) 7077-7085. https://doi.org/10.1021/acssuschemeng.7b01274
11. I. V. Khristenko, A. V. Panteleimonov, R. Yu. Iliashenko, A.O. Doroshenko, V.V. Ivanov, O.S. Tkachenko, E.V. Benvenutti, Yu. V. Kholin, Heterogeneous polarity and surface acidity of silica-organic materials with fixed 1-n-propyl-3-methylimidazolium chloride as probed by solvatochromic and fluorescent dyes, Colloids Surf. A: Physicochem. Eng. Asp. 538 (2018) 280-286. https://doi.org/10.1016/j.colsurfa.2017.11.018
12. J. Liu, R. Li, X. Si, D. Zhou, Y. Shi, Y. Wang, X. Jing, P. Wang, Oligothiophene dye-sensitized solar cells, Energy Environ. Sci. 3(12) (2010) 1924-1928. https://doi.org/10.1039/c0ee00304b
13. M. R. Elmorsy, F. H. Abdelhamed, A. sultan Alqahtani, E. Abdel-Latif, A. A. Abdel-Shafi, M. A. Ismail, Computational design and experimental investigation of novel thiophene-based organic dyes with N, N-dimethylaminophenyl and morpholinophenyl donors for dye-sensitized solar cells, Opt. Mater. 151 (2024) 115386. https://doi.org/10.1016/j.optmat.2024.115386
14. S. K. Panja, S. S. Sharma, K. Sharma, K. Patel, A.L. Prajapat, S. Saha, S. Bera, N. Srivastava, J. Ray, Thiophene oligomer based NIR dyes: Photophysical properties and application in dye sensitized solar cells, Chem. Phys. 595 (2025) 112721. https://doi.org/10.1016/j.chemphys.2025.112721
15. D. Demeter, J. Roncali, S. Jungsuttiwong, F. Melchiorre, P. Biagini, R. Po, Linearly π-conjugated oligothiophenes as simple metal-free sensitizers for dye-sensitized solar cells, J. Mater. Chem. C 3(29) (2015) 7756-7761. https://doi.org/10.1039/C5TC01183C
16. A. Sreekumar, A. R. Nair, R. Chandramani, P. S. Iyer, G. Sadasivan, P. T. Sylajadevi, R. Pillai, A. Nizam, A. Sivan, Thiophene-centered small molecule emitters, in: AIP Conference Proceedings 3171(1), AIP Publishing LLC, 2024, p. 020001. https://doi.org/10.1063/5.0222858
17. R. Isci, T. Ozturk, Thienothiophene-based organic light-emitting diode: synthesis, photophysical properties and application, Beilstein J. Org. Chem. 19(1) (2023) 1849-1857. https://doi.org/10.3762/bjoc.19.137
18. M. Zavelani-Rossi, G. Lanzani, S. De Silvestri, M. Anni, G. Gigli, R. Cingolani, Tunable efficient laser emission from electroluminescent functionalized oligothiophene, in: Summaries of Papers Presented at the Lasers and Electro-Optics CLEO'02, IEEE, 2002, pp. 459-460. https://doi.org/10.1109/CLEO.2002.1034201
19. W. Zhang, F. Ren, Y. Wen, Z. Xu, J. Sun, Z. Yan, L. Ding, X. Hai, Q. Zhang, Enhancing Hole Transport of Quantum-Dot Light-Emitting Diodes by a Cruciform Oligothiophene for Effective p-Type Doping, Macromol. Rapid Commun. 43(15), (2022) 2200187. https://doi.org/10.1002/marc.202200187
20. J. N. Pedroza-Montero, F. A. Delesma, R. I. Delgado-Venegas, P. Calaminici, A. M. Köster, Static and dynamic polarizabilities of oligothiophenes, Theor. Chem. Acc. 135(9) (2016) 230. https://doi.org/10.1007/s00214-016-1984-8
21. T. Kosa, P. Palffy-Muhoray, H. Zhang, T. Ikeda, Large optical torque enhancement by oligothiophene dye in a nematic liquid crystal host, Mol. Cryst. Liq. Cryst. 421(1) (2010) 107-115. https://doi.org/10.1080/15421400490501581
22. S. Muhammad, A. Bibi, S. Bibi, A. G. Al-Sehemi, H. Algarni, F. Sarwar, Exploring the quinoidal oligothiophenes to their robust limit for efficient linear and nonlinear optical response properties,

Chem. Pap. 76(7) (2022) 4273-4288. https://doi.org/10.1007/s11696-022-02167-3
23. A. Bibi, S. Muhammad, S. UrRehman, S. Bibi, S. Bashir, K. Ayub, M. Adnan, M. Khalid, Chemically modified quinoidal oligothiophenes for enhanced linear and third-order nonlinear optical properties, ACS omega, 6(38) (2021) 24602-24613. https://doi.org/10.1021/acsomega.1c03218
24. J. Yokota, A. Shishido, Nonlinear optical molecular orientation of oligothiophene-doped liquid crystals, Mol. Cryst. Liq. Cryst. 768(17) (2024) 975-992. https://doi.org/10.1080/15421406.2024.2360220
25. M. I. Berdnyk, A. B. Zakharov, V. V. Ivanov, Application of L1-regularization approach in QSAR problem. Linear regression and artificial neural networks, Methods Objects Chem. Anal. 14 (2019) 79–90. https://doi.org/10.17721/moca.2019.79-90
26. A. B. Zakharov, D. K. Tsarenko, V. V. Ivanov, Topological characteristics of iterated line graphs in the QSAR problem: a multigraph in the description of properties of unsaturated hydrocarbons, Struct Chem. 32 (2021) 1629–1639. https://doi.org/10.1007/s11224-021-01737-7
27. A. B. Zakharov, M. Kyrpa, A. V. Kyrychenko, S. M. Kovalenko, O. N. Kalugin, V. V. Ivanov, L. Adamowicz, Towards the computational design of organic molecules with specified properties, Struct Chem. 36 (2025) 723–738. https://doi.org/10.1007/s11224-024-02441-y
28. M. Kyrpa, A. Zakharov, V. Ivanov, Fast vs Accurate: Evaluating TD-DFTB for Large-Scale Screening of Organic Dyes, Kharkiv Univ Bull Chem Ser 44 (2025) 72-80. https://doi.org/10.26565/2220-637X-2025-44-07
29. A. B. Zakharov, A. V. Dyachenko, V. V. Ivanov, Topological characteristics of iterated line graphs in QSAR problem: octane numbers of saturated hydrocarbons, J. Chemom. 33(9), (2019) e3169. https://doi.org/10.1002/cem.3169
30. B. Hourahine, M. Berdakin, J. A. Bich, F. P. Bonafé, C. Camacho, Q. Cui, M. Y. Deshaye, G. Díaz Mirón, S. Ehlert, M. Elstner, T. Frauenheim, N. Goldman, R. A. González León, T. van der Heide, S. Irle, T. Kowalczyk, T. Kubař, I. S. Lee, C. R. Lien-Medrano, A. Maryewski, T. Melson, S. K. Min, T. Niehaus, A. M. N. Niklasson, A. Pecchia, K. Reuter, C. G. Sánchez, C. Scheurer, M. A. Sentef, P. V. Stishenko, V. Q. Vuong, B. Aradi, Recent developments in DFTB+, a software package for efficient atomistic quantum mechanical simulations, J. Phys. Chem. A. 129(24) (2025) 5373-5390. https://doi.org/10.1021/acs.jpca.5c01146
31. DFTB+ simulation package — DFTB+ https://dftbplus.org/
32. DFTB+ recipes - Read the Docs https://dftbplus-recipes.readthedocs.io/en/latest/
33. F. Neese, The ORCA program system, Wiley Interdiscip. Rev. Comput. Mol. Sci. 2(1) (2012) 73-78. https://doi.org/10.1002/wcms.81
34. F. Neese, Software update: the ORCA program system - version 6.0, Rev. Comput. Mol. Sci. 15 (2025) e70019. https://doi.org/10.1002/wcms.70019
35. C. W. Yap, PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32(7) (2011) 1466-1474. https://doi.org/10.1002/jcc.21707
36. D. A. Papaconstantopoulos, M. J. Mehl, The Slater–Koster tight-binding method: a computationally efficient and accurate approach, J. Phys. Condens. Matter. 15(10) (2023) R413. https://doi.org/10.1088/0953-8984/15/10/201
37. M. Petitjean, Applications of the radius-diameter diagram to the classification of topological and geometrical shapes of chemical compounds, J. Chem. Inf. Comput. Sci. 32(4) (1992) 331-337. https://doi.org/10.1021/ci00008a012
38. M. Abadi M, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, Tensorflow: A system for large-scale machine learning, in: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283. https://doi.org/10.48550/arXiv.1605.08695
39. C. Mattmann, Machine Learning with TensorFlow, second ed., Manning Publications Co., New York, 2020.
40. M. Barysz, G. Jashari, L. S. Lall, V. K. Srivastava, N. Trinajsti, On the Distance Matrix of Molecules Containing Heteroatoms, in: Chemical applications of Graph Theory and Topology, R. B. King (Ed.), Elsevier Science Publishers, Amsterdam, 1983. http;s://doi.org/10.1016/S0003-2670%2800%2982967-0
41. M. Stone, Cross-validatory choice and assessment of statistical predictions, J. R. Stat. Soc. Series B Stat. 36(2) (1974) 111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x