PROSPECTS
IN PHARMACEUTICAL SCIENCES

Prospects in Pharmaceutical Sciences, 24(2), 98-103
https://prospects.wum.edu.pl/

Original Article

PREDICTION OF COMPLEXATION CONSTANTS AND BINDING FREE ENERGIES OF B-CYCLODEXTRIN INCLUSION COMPLEXES USING MACHINE LEARNING METHODS

Jakub Jarząbek[1], Szymon Kamil Araj*[2], Dariusz Maciej Pisklak[2], Aleksandra Kowalska[2] , Łukasz Szeleszczuk[2]

[1]*Institute of Radioelectronics and Multimedia Technology, Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland.*

[2]*Department of Organic and Physical Chemistry, Faculty of Pharmacy, Medical University of Warsaw, Banacha 1 Str., 02-093 Warsaw, Poland*

* Correspondence, e-mail: szymon.araj@wum.edu.pl

Received: 14.02.2026 / Revised: 27.02.2026 /Accepted: 27.02.2026 / Published: 30.04.2026

ABSTRACT

Accurate prediction of complexation constants for cyclodextrin inclusion complexes remains a challenging task due to experimental limitations and the high computational cost of theoretical approaches. In this study, machine learning methods were applied as computational tools to predict the complexation constants of ß-cyclodextrin inclusion complexes based on experimentally derived data. A curated dataset of ß-cyclodextrin–guest complexes measured at 273 K and pH 7 was combined with a comprehensive set of classical molecular descriptors and SMILES-derived Iso2vec embeddings. Classical descriptors were calculated using the Materials Studio software and included fragment-based structural counts and surface-related Jurs descriptors, while Iso2vec embeddings provided an additional representation of molecular structure and stereochemistry. Feature selection was performed using Heat Map–Based Feature Ranking, enabling the identification of compact and informative feature subsets. Several regression models were evaluated, including linear models and tree-based ensemble methods. Among them, gradient-boosted decision tree models, particularly LightGBM, demonstrated the best predictive performance. The inclusion of Iso2vec embeddings consistently improved model accuracy across architectures, indicating that these features capture structural information not accessible through conventional descriptors alone. Model interpretability analysis using SHAP values revealed that both classical descriptors and Iso2vec components contribute to the final predictions. The proposed approach offers a practical and interpretable framework for data-driven prediction of cyclodextrin complexation constants and may support early-stage decision-making in cyclodextrin-based pharmaceutical formulation development.

KEYWORDS: ß-cyclodextrin; inclusion complexes; machine learning; molecular descriptors; Iso2vec embeddings

1. Introduction

Cyclodextrins (CDs) (Fig. 1) are a well-established class of cyclic oligosaccharides widely used in pharmaceutical sciences as solubilizing agents, stabilizers, and carriers for poorly water-soluble drugs [1-3]. Their toroidal structure, composed of α-(1→4)-linked D-glucopyranose units, creates a hydrophobic inner cavity and a hydrophilic outer surface, enabling the formation of non-covalent inclusion complexes with a broad range of guest molecules [4]. Among the naturally occurring cyclodextrins, ß-cyclodextrin (ß-CD) is particularly attractive due to its favorable cavity size, availability, and extensive regulatory acceptance in pharmaceutical formulations [5]. As a result, ß-CD inclusion complexes have been intensively studied both experimentally and theoretically, especially in the context of improving drug solubility, stability, and bioavailability.

A key quantitative parameter describing cyclodextrin-guest interactions is the complexation (binding) constant, which reflects the thermodynamic stability of the inclusion complex under given conditions. Reliable prediction of complexation constants remains a challenging task. Experimental determination typically relies on techniques such as phase-solubility analysis or isothermal titration calorimetry (ITC). However, these

approaches are time-consuming, sensitive to experimental conditions (e.g., temperature, pH, ionic strength), and often limited by practical constraints. Consequently, comprehensive experimental screening of large numbers of potential guest molecules is rarely feasible [7].

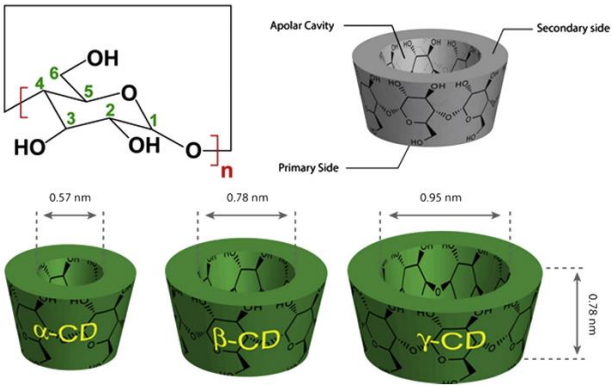

Fig. 1. Toroidal representation of: α-CD (n = 6), β-CD (n = 7), and γ-CD (n = 8) and their structural dimensions. Reprinted with permission from [6]. Copyright 2026 American Chemical Society.

From a theoretical perspective, cyclodextrin inclusion has been investigated using molecular docking, molecular dynamics (MD) simulations, and quantum mechanical calculations [8; 9]. While these approaches can provide valuable mechanistic insight into host–guest recognition, they are computationally demanding and typically require extensive sampling to achieve quantitative accuracy. High-level free-energy methods, in particular, become impractical when applied to large datasets comprising hundreds of chemically diverse guest molecules [10]. This limitation substantially restricts their applicability in high-throughput screening scenarios or data-driven formulation design.

In recent years, machine learning (ML) methods have emerged as a complementary computational strategy for predicting molecular properties and interaction strengths from existing experimental data. In the context of cyclodextrin inclusion complexes, ML models offer the potential to approximate complex, nonlinear relationships between molecular structure and binding affinity at a fraction of the computational cost required by physics-based simulations. Importantly, when trained on experimentally measured complexation constants, such models can serve as fast predictive tools that support early-stage formulation development and hypothesis generation [11-13].

The effectiveness of ML models in this domain strongly depends on how molecular structures are represented numerically. Traditional approaches rely on classical physicochemical descriptors, such as molecular weight, lipophilicity, topological indices, or hydrogen-bonding parameters. While these descriptors capture many relevant aspects of molecular behavior, they often fail to distinguish between stereoisomers or structurally subtle variants that may exhibit markedly different binding affinities toward cyclodextrins. To address this limitation, alternative representations derived directly from molecular line notations, such as SMILES-based embeddings, have been proposed. These representations aim to encode structural and stereochemical information in a form that can be effectively exploited by data-driven models [14].

In this study, we build upon a curated experimental dataset of β-cyclodextrin inclusion complexes and explore the predictive performance of machine learning regression models that combine classical molecular descriptors with Iso2vec embeddings, a SMILES-based numerical representation inspired by natural language processing techniques. Machine learning is employed here strictly as a computational tool to analyze existing experimental data; the objective is not to develop new learning algorithms, but to systematically evaluate how established models and feature representations perform in the specific context of cyclodextrin–guest binding prediction. Particular emphasis is placed on feature selection, model interpretability, and comparative performance analysis across representative regression paradigms.

The aim of the present work is therefore twofold: (i) to assess whether SMILES-derived embeddings provide complementary information to conventional descriptors in predicting β-cyclodextrin complexation constants, and (ii) to identify robust and interpretable machine learning models suitable for practical use on medium-sized experimental datasets. By addressing these points, this study seeks to contribute to the rational, data-driven prediction of cyclodextrin inclusion phenomena and to support the broader application of machine learning methods in pharmaceutical host–guest chemistry

2. Materials and methods

2.1. Calculation of Molecular Descriptors

Classical molecular descriptors for all guest molecules were calculated using Materials Studio software (BIOVIA, Dassault Systèmes), version 2020, employing the Models module. Descriptor calculation was performed on optimized molecular structures. The applied descriptor set was intentionally limited to chemically interpretable features commonly used in quantitative structure–property relationship (QSPR) studies, in order to maintain transparency and facilitate subsequent model interpretation.

The calculated descriptors comprised two main categories: fragment-based structural counts and surface-area–related descriptors. Fragment count descriptors quantify the occurrence of predefined functional groups and hydrocarbon fragments within each molecule. These included, among others, acidic and basic functionalities (e.g., carboxylic acid, carboxylate, amide, amine, quaternary amine), heteroatom-containing groups (e.g., nitro, nitrile, nitroso, oxime, hydrazone, thiol, peroxide, phosphate, sulfonic acid), as well as hydrocarbon fragments of varying size and branching (e.g., methyl, ethyl, propyl, butyl, isopropyl, tert-butyl, ethylene, propylene, butadiene). In addition, selected aromatic- and ether-related fragments, such as phenoxy and methoxy groups, were included. This representation captures the presence and frequency of chemically meaningful substructures that are known to influence host–guest interactions with cyclodextrins.

The second group of descriptors consisted of Jurs surface descriptors, which describe molecular surface

area partitioned according to atomic partial charges. These included total solvent-accessible surface area (SASA), total polar surface area (TPSA), total hydrophobic surface area (TASA), and their charge-resolved components, such as PPSA1, PNSA1, DPSA1, RPSA, and RASA. These descriptors provide information on the spatial distribution of polar and nonpolar regions on the molecular surface and are particularly relevant for cyclodextrin inclusion, where hydrophobic effects and polar rim interactions jointly contribute to complex stability.

All descriptors were calculated consistently for the entire dataset using identical computational settings. The resulting descriptor matrix was combined with the Iso2vec embeddings provided in the original dataset to form the complete feature set used in subsequent machine learning analyses. No manual preselection of individual descriptors was performed prior to feature ranking, ensuring an unbiased evaluation of descriptor relevance during the feature selection stage

2.2. Chemical Dataset Construction

The database used for this work was created by Tahil et al. [11]. To adapt its content to effective machine learning processes, it was limited to only complexes with beta cyclodextrin and only for values at 273 K and pH = 7. Additionally, new calculated descriptors were added to the database.

2.3. Feature Encoding and Selection

2.3.1. Molecular Descriptors and Iso2vec Embeddings

The molecular representation used in this study combines classical physicochemical descriptors with Iso2vec embeddings, a SMILES-based representation originally introduced by Tahıl et al. in the context of cyclodextrin–guest binding prediction [15]. Tahıl et al. proposed Iso2vec, a numerical embedding of isomeric SMILES strings inspired by natural language processing techniques, in particular Word2Vec [16]. In Iso2vec, a molecule is treated as a sequence of SMILES tokens (including atoms, bonds, ring indices, stereochemical markers, and special characters), and a prediction-based embedding model is trained to project these sequences into a continuous vector space. In the original work, Iso2vec embeddings consist of 10 numerical components per molecule, learned from the set of guest-molecule isomeric SMILES in the dataset. These embeddings were shown to successfully differentiate stereoisomers such as ephedrine and pseudoephedrine variants, for which classical physicochemical descriptors are identical but experimental binding affinities differ [15]. In the present study, Iso2vec embeddings are used as provided by the original dataset construction and are not retrained or modified. They are treated as complementary input features alongside classical descriptors. A complete ranking of all features, including Iso2vec components, obtained using Heat Map–Based Feature Ranking, is provided in Supplementary Materials.

2.3.2. Heat Map–Based Feature Ranking (HmbFR)

To identify the most informative subset of features, we used the Heat Map–Based Feature Ranker (HmbFR) [17]. HmbFR is a filter-based feature selection method that evaluates features based on their ability to differentiate target values through local, group-wise interactions rather than individual marginal effects. The core idea of HmbFR is to group features into small, ordered subsets and transform them into a compressed, image-like representation. Each group is mapped to a discrete color palette, and the color distributions are compared across target classes or target ranges. Features that consistently induce different color distributions are assigned higher importance scores. In this work, we adapted HmbFR for regression by discretizing the target variable into quantile-based bins before ranking. Our implementation also introduces minor numerical stabilizations, including probability smoothing when estimating color distributions, to improve robustness on finite datasets. The resulting ranked feature list, including both classical descriptors and Iso2vec features, is provided in Supplementary Materials. Based on this ranking, multiple reduced feature subsets were constructed; the best-performing models consistently relied on the 64 top-ranked features.

2.4. Model Architecture and Selection

Rather than focusing on extensive pipeline engineering, we evaluated a concise set of representative regression models to assess how different learning paradigms interact with the selected features. The evaluated models include linear regression, bagging-based ensembles, and gradient-boosted decision tree methods [18-23]. Model comparison was performed using a unified train–test split and identical evaluation metrics. A summary comparison of model performance is presented in Table 1, based on the held-out test set.

Table 1. Performance comparison of representative models on the test set

Model	MAE	RMSE	R^2
Light GBM	1.969	2.642	0.688
Hist Gradient Boosting	2.031	2.695	0.675
XGBoost (tuned)	2.015	2.735	0.666
Random Forest	2.178	2.895	0.625
Ridge Regression	2.841	3.712	0.384

A comparison across top 8 best performing evaluated configurations is provided in Supplementary Materials. The strongest performance was achieved by LightGBM trained on the feature set including Iso2vec embeddings, using the 64 top-ranked features identified by HmbFR. This result indicates that gradient-boosted decision trees are especially effective at exploiting the nonlinear correlations introduced by structural embeddings.

Deep learning architectures were not considered in this study, as the dataset size and the tabular nature of the feature representation favor tree-based ensemble methods in terms of stability and interpretability.

2.5. Evaluation Metrics

Performance is reported using three complementary regression metrics: mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). RMSE is the primary ranking metric for architecture

selection, aligning with the minimization objective in the Optuna [24] tuning loop and the summary table generation. MAE provides a robust measure of typical absolute deviation, while R^2 summarizes the proportion of variance explained by the model [25]. These metrics are computed on the held-out test set to provide an unbiased comparison across architectures and feature variants.

3. Results and Interpretation

3.1. Impact of Iso2vec Features

Comparing models trained with and without Iso2vec embeddings reveals a consistent performance advantage when embeddings are included. The improvement is seen across multiple architectures, suggesting that Iso2vec features capture information orthogonal to classical descriptors rather than acting as a redundant representation. These findings support the hypothesis that SMILES-derived embeddings provide a meaningful bridge between symbolic molecular representations and numerical learning models.

3.2. Model Interpretability

To interpret the predictions of the final LightGBM model, we utilized SHAP (SHapley Additive exPlanations) values [26]. Fig. 2 presents a SHAP summary plot for the selected model. The analysis shows that classical descriptors such as molecular weight remain influential, but several Iso2vec components similarly contribute to the prediction.

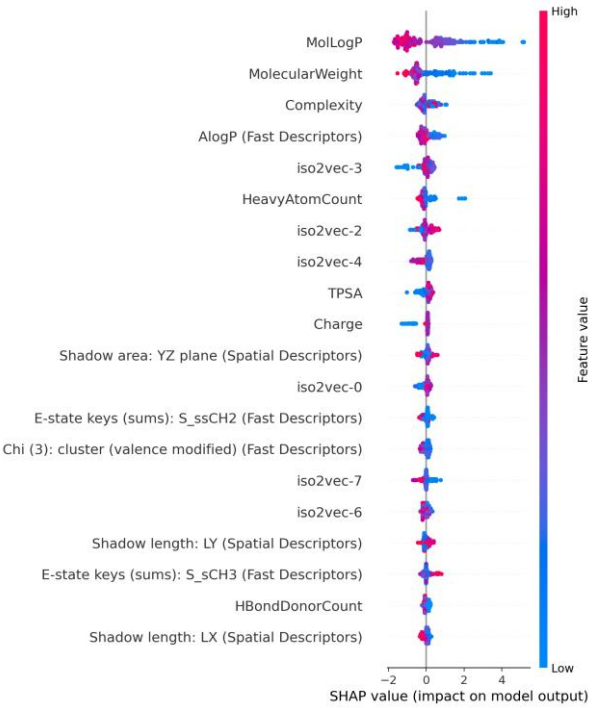

Fig. 2. SHAP summary plot for the final LightGBM model trained on the top-ranked feature subset.

3.3. Summary of Contributions

In summary, the results of this study demonstrate that Iso2vec embeddings lead to a substantial improvement in predictive performance by encoding structural information that is not captured by conventional molecular descriptors. Furthermore, the HmbFR approach proves to be an effective strategy for identifying and prioritizing relevant features within high-dimensional molecular datasets. In addition, gradient-boosted tree-based models, with LightGBM in particular, are shown to be well suited for exploiting the combined feature space derived from both descriptor-based and embedding-based representations. All datasets used in this study, along with detailed feature rankings and model summaries, are provided in the Supplementary Materials.

4. Discussion

The results obtained in this study demonstrate that machine learning models can successfully predict complexation constants of ß-cyclodextrin inclusion complexes when trained on experimentally derived data and appropriately selected molecular features. In particular, the observed performance of gradient-boosted decision tree models highlights their suitability for modeling nonlinear relationships between molecular structure and binding affinity, which are difficult to capture using traditional linear or descriptor-only approaches.

A key finding of this work is the consistent improvement in predictive accuracy achieved by incorporating Iso2vec embeddings alongside classical molecular descriptors. While fragment-based and surface-related descriptors capture chemically intuitive features relevant to cyclodextrin inclusion — such as hydrophobic surface area, functional group composition, and charge distribution — they are inherently limited in their ability to encode subtle structural differences. This limitation becomes especially apparent for stereoisomeric compounds, which may exhibit distinct binding affinities despite sharing identical conventional descriptor values. The inclusion of SMILES-derived Iso2vec embeddings partially addresses this issue by providing an alternative representation that reflects structural and stereochemical patterns embedded in the molecular graph.

Importantly, the benefits of Iso2vec embeddings were observed across multiple model architectures, suggesting that their contribution is not model-specific but rather complementary to classical descriptors. This supports the notion that combining chemically interpretable features with data-driven structural representations provides a more complete description of host–guest interactions than either approach alone. The feature-ranking results obtained using the Heat Map–Based Feature Ranking method further confirm this complementarity, as both descriptor types consistently appear among the most informative features in the optimized models.

From a methodological perspective, the application of HmbFR proved to be an effective strategy for reducing feature dimensionality while preserving predictive performance. By prioritizing features based on their local discriminative power rather than global marginal correlations, HmbFR enables the construction of compact and informative feature subsets suitable for medium-sized experimental datasets. This is particularly relevant in pharmaceutical applications, where available datasets are often limited by experimental cost and heterogeneity rather than data scarcity in the conventional machine learning sense.

The interpretability analysis based on SHAP values provides additional insight into the learned models. The coexistence of classical descriptors and Iso2vec components among the most influential features indicates that the final predictions arise from an interplay between chemically intuitive properties and abstract structural patterns. This balance between performance and interpretability is crucial for the practical adoption of machine learning models in pharmaceutical research, where purely black-box predictions are often met with skepticism.

Several limitations of the present study should be acknowledged. The dataset was restricted to ß-cyclodextrin complexes measured under a single temperature and pH condition, which limits the direct transferability of the models to other experimental settings. In addition, the analysis focused exclusively on ß-cyclodextrin and did not consider chemically modified cyclodextrins or alternative host systems. Future work may extend the present framework to include broader datasets, additional environmental conditions, and transfer-learning strategies to improve generalizability.

5. Conclusions

In this study, we demonstrated that machine learning models combining classical molecular descriptors with SMILES-derived Iso2vec embeddings can effectively predict complexation constants of ß-cyclodextrin inclusion complexes. The results show that gradient-boosted tree-based models, particularly LightGBM, are well suited to exploit the nonlinear relationships present in the combined feature space.

The inclusion of Iso2vec embeddings consistently improved predictive performance, indicating that these representations capture structural information not accessible through conventional descriptors alone. At the same time, feature selection using Heat Map–Based Feature Ranking enabled the construction of compact and interpretable models without sacrificing accuracy. Together, these findings highlight the value of integrating chemically interpretable descriptors with data-driven structural embeddings in host–guest binding prediction.

From a pharmaceutical perspective, the proposed approach offers a practical and computationally efficient tool for prioritizing guest molecules in cyclodextrin-based formulation development. While not intended to replace experimental measurements or detailed molecular simulations, the presented models can support early-stage decision-making and hypothesis generation. Overall, this work contributes to the growing body of evidence supporting the use of machine learning as a complementary methodology in pharmaceutical host–guest chemistry.

Appendix

All Supplementary Materials have been made available through the repository: https://github.com/Macrume/prediction-of-complexation-constants-and-binding-free-energies-of-bheta-cyclodextrins. The supplementary material includes the following essential files: Primary Dataset, HmbFR Feature Ranking, Model Comparison Summary, Detailed Report for Final Configuration, Train-Ready Dataset Without Ido2vec and Train-Ready Dataset With Iso2vec. Additional files and their descriptions are provided in the repository. For further information or requests regarding the supplementary materials, please contact the corresponding author.

Author Contributions: Conceptualization, S.K.A.; J.J.; Ł.S.; methodology, S.K.A.; J.J.; Ł.S.; validation, S.K.A.; J.J.; Ł.S.; investigation, S.K.A.; resources, Ł.S., A.K.; data curation, Ł.S.; writing—original draft preparation, S.K.A., Ł.S., J.J.; writing—review and editing, S.K.A., D.M.P.; visualization, J.J.; supervision, Ł.S.; project administration, S.K.A.; funding acquisition, Ł.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Higher Education (Poland) under the program "Studenckie koła naukowe tworzą innowacje", project entitled "Wykorzystanie algorytmów sztucznej inteligencji do przewidywania wpływu cyklodekstryn na biodostępność wybranych substancji farmakologicznie aktywnych", grant number SKN/SP/630822/2025. The grant was awarded to Łukasz Szeleszczuk.

Conflicts of Interest: Ł.S. serves as the Editor-in-Chief of Prospects in Pharmaceutical Sciences. To avoid any potential conflict of interest, he was fully excluded from the editorial handling and peer-review process of this manuscript. The editorial assessment and decision were conducted independently by another member of the editorial board. The remaining authors declare no conflicts of interest.

References

1. Pyrak, B.; Gubica, T.; Rogacka-Pyrak, K. Cyclodextrin nanosponges as bioenhancers of phytochemicals. *Prospect. Pharm. Sci.* **2024**, *22(3)*, 170–177. DOI: 10.56782/pps.272
2. Zielińska-Pisklak, M.; Michalik, K.A.; Szeleszczuk, Ł. Complexes of Fat-Soluble Vitamins with Cyclodextrins. *Int. J. Mol. Sci.* **2025**, *26(13)*, Art. No: 6110. DOI: 10.3390/ijms26136110
3. Araj, S.K.; Szeleszczuk, Ł. A Review on Cyclodextrins/Estrogens Inclusion Complexes. *Int. J. Mol. Sci.* **2023**, *24(10)*, Art. No: 8780. DOI: 10.3390/ijms24108780
4. Pyrak, B.; Rogacka-Pyrak, K.; Gubica, T.; Szeleszczuk, Ł. Exploring Cyclodextrin-Based Nanosponges as Drug Delivery Systems: Understanding the Physicochemical Factors Influencing Drug Loading and Release Kinetics. *Int. J. Mol. Sci.* **2024**, *25(6)*, Art. No: 3527. DOI: 10.3390/ijms25063527
5. Napiórkowska, E.; Szeleszczuk, Ł. *Review of* Applications of ß-Cyclodextrin as a Chiral Selector for Effective Enantioseparation. *Int. J. Mol. Sci.* **2024**, *25(18)*, Art. No: 10126. DOI: 10.3390/ijms251810126
6. Crini, G. A History of Cyclodextrins. *Chem. Rev.* **2014**, *114(21)*, 10940–10975. DOI: 10.1021/cr500081p
7. Christoforides, E.; Andreou, A.; Koskina, P.; Bethanis, K. Selective Crystallization of Trans-Nerolidol in ß-Cyclodextrin: Crystal Structure and Molecular Dynamics Analysis. *Crystals* **2025**, *15(9)*, Art. No: 802. DOI: 10.3390/cryst15090802

8. Napiórkowska, E.; Szeleszczuk, Ł. Conformational landscape of ß-cyclodextrin: a computational resource for host-guest modeling in supramolecular systems. *J. Comput. Aided Mol. Des.* **2025**, *39*, Art. No: 117. DOI: 10.1007/s10822-025-00694-1

9. Gackowski, M.; Madriwala, B.; Studzińska, R.; Koba, M. Novel Isosteviol-Based FXa Inhibitors: Molecular Modeling, In Silico Design and Docking Simulation. *Molecules* **2023**, *28(13)*, Art. No: 4977. DOI: 10.3390/molecules28134977

10. Spirande, E.; Miryashkin, T.; Kolmakov, A.; Shapeev, A. Automated prediction of thermodynamic properties via Bayesian free-energy reconstruction from molecular dynamics. *Comput. Condens. Matter* **2025**, *45*, Art. No: e01163. DOI: 10.1016/j.cocom.2025.e01163

11. Boczar, D.; Michalska, K. A Review of Machine Learning and QSAR/QSPR Predictions for Complexes of Organic Molecules with Cyclodextrins. *Molecules* **2024**, *29(13)*, Art. No: 3159. DOI: 10.3390/molecules29133159

12. Gackowski, M.; Szewczyk-Golec, K.; Pluskota, R.; Koba, M.; Mądra-Gackowska, K.; Woźniak, A. Application of Multivariate Adaptive Regression Splines (MARSplines) for Predicting Antitumor Activity of Anthrapyrazole Derivatives. *Int. J. Mol. Sci.* **2022**, *23(9)*, Art. No: 5132. DOI: 10.3390/ijms23095132

13. Gackowski, M.; Madriwala, B.; Koba, M. In silico design, docking simulation, and ANN-QSAR model for predicting the anticoagulant activity of thiourea isosteviol compounds as FXa inhibitors. *Chem. Pap.* **2023**, *77*, 7027–7044. DOI: 10.1007/s11696-023-02994-y

14. Todkar, R.; Shirote, P.; Mohite, S. In Silico Screening and DFT Analysis of *Nelumbo nucifera* Phytochemicals as Potential BACE-1 Inhibitors for Alzheimer's disease. *Prospect. Pharm.* Sci. **2025**, *23(4)*, 29–36. DOI: 10.56782/pps.379

15. Tahıl, G.; Delorme, F.; Le Berre, D.; Monflier, É.; Sayede, A.; Tilloy, S. Curated dataset of association constants between a cyclodextrin and a guest for machine learning. *Chem. Data Collect.* **2023**, *45*, Art. No: 101022. DOI: 10.1016/j.cdc.2023.101022

16. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* 2013, arXiv:1301.3781. DOI: 10.48550/arXiv.1301.3781

17. Huertas, C.; Juárez-Ramírez, R.; Raymond, C. Heat Map based Feature Ranker: In Depth Comparison with Popular Methods. *Intell. Data Anal.* **2018**, *22(5)*, 1009–1037. DOI: 10.3233/IDA-173481

18. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12(85)*, 2825–2830.

19. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proc. KDD* **2016**, *2016*, 785–794. DOI: 10.1145/2939672.2939785

20. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.

21. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6638–6648.

22. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29(5)*, 1189–1232. DOI: 10.1214/aos/1013203451

23. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. DOI: 10.1023/A:1010933404324

24. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. *Proc. KDD* **2019**, *2019*, 2623–2631. DOI: 10.1145/3292500.3330701

25. Hyndman, R.J.; Koehler, A.B. Another Look at Measures of Forecast Accuracy. *Int. J. Forecast.* **2006**, *22(4)*, 679–688. DOI: 10.1016/j.ijforecast.2006.03.001

26. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.