ACS OMEGA

Open Access

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

Article

Utilizing Machine Learning Models for Predicting Diamagnetic Susceptibility of Organic Compounds

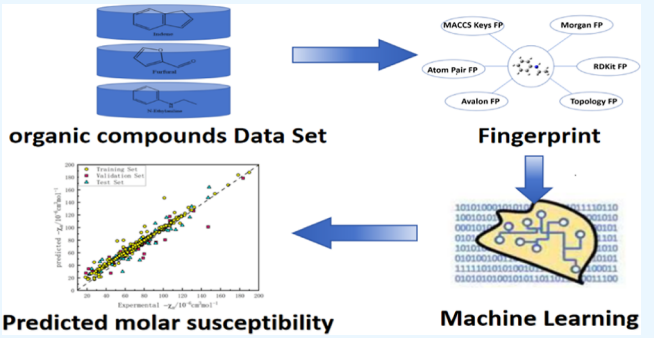
Yining Zhang, Sijie Xing, Lai Wei,* and Tongfei Shi*

Cite This: *ACS Omega* 2024, 9, 14368−14374

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: This research is centered on examining the magnetic characteristics of organic molecules, with a particular emphasis on magnetic susceptibility, an essential physical property that provides insights into molecular microstructures and reaction processes. Traditional approaches for determining and calculating magnetic susceptibility are generally inefficient and demanding. To overcome these challenges, we have introduced a novel approach using quantitative structure−property relationships, which efficiently elucidates the relationship between the structural properties of molecules and their molar magnetic susceptibility. In our study, we utilized a comprehensive database comprising molar magnetic susceptibility data for 382 organic molecules. We applied six distinct molecular fingerprinting methods—RDKit Fingerprint, Morgan Fingerprint, MACCS Keys, atom pair fingerprint, Avalon Fingerprint, and topology fingerprint—as feature inputs for training seven different machine learning models, namely random forest, AdaBoost, gradient boosting, extra trees, elastic net, support vector machine, and multilayer perceptron (MLP). Our findings revealed that the integration of the atom pair fingerprint with the MLP model yielded R^2 values of 0.88 and 0.90 in the validation and test sets, respectively, showcasing exceptional predictive accuracy. This advancement significantly expedites research and development processes related to the magnetic properties of organic molecules. Moreover, by employing this effective predictive method, it is expected to considerably reduce both experimental and computational expenses while maintaining high accuracy. This development represents a breakthrough in the rapid screening and prediction of properties for various compounds, offering a new and efficient pathway in this field of study.



INTRODUCTION

Molar magnetic susceptibility, χ_m, is a critical physical quantity that offers substantial insights into the magnetic characteristics of organic molecules and various materials. It serves as an essential physicochemical property, defining a compound's reaction to an external magnetic field. When a material is subjected to a magnetic field, denoted as H, it experiences an induced magnetization, M, which represents the magnetic moment per unit volume. This induction process is described by the relationship $M = \kappa H$,[1−4] where κ signifies the volumetric magnetic susceptibility. Given that H and M share the same units, κ is a dimensionless quantity.

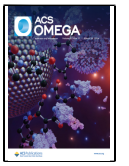
In the realms of chemistry and materials science, the molar magnetic susceptibility χ_m is a more applicable parameter. It is calculated by multiplying the volumetric magnetic susceptibility (κ) by the molar volume of the compound. Alternatively, it can be expressed as the product of κ and the molar mass of the compound divided by its density. This approach offers a practical and insightful way to understand and quantify the magnetic properties of substances in response to magnetic fields, playing a pivotal role in the study of their magnetic behavior. The formula for molar magnetic susceptibility can be expressed as

$$\chi_m = \kappa V_m = \kappa \frac{M}{\rho} \quad (1)$$

where V_m is the molar volume of the substance, M is the molar mass, and ρ is the mass density.

The definition of molar magnetic susceptibility is particularly significant because it accounts for the quantity of a substance, enabling standardized comparisons of magnetic properties across various materials.[5−9] This parameter is instrumental in identifying the diamagnetic or paramagnetic nature of compounds and finds widespread application in coordination chemistry, molecular magnetism, and material science.

Received: December 28, 2023
Revised: February 3, 2024
Accepted: March 8, 2024
Published: March 15, 2024



Historically, the measurement of the molar magnetic susceptibility has depended on intricate experiments and procedures that are often time-consuming.

In addition to direct measurements, methods like Monte Carlo simulations,[10] first-principle calculations,[11] multivariate linear regression,[12−14] and neural networks[13] have been employed to simulate and study the molar magnetic susceptibility of substances. While these techniques can yield precise data, their speed and efficiency are becoming increasingly insufficient to satisfy the demands of rapidly progressing research and industrial applications. The field of materials science is continuously evolving with the introduction of new materials, intensifying the need for quick and accurate prediction of molar magnetic susceptibility. This urgency underscores the necessity for more efficient approaches that can keep pace with the advancements in this domain, ensuring timely and accurate analysis of magnetic properties in a broad range of materials.

In recent years, machine learning, as a powerful tool for data analysis, has shown tremendous potential in cheminformatics and materials science, particularly in predicting the physical and chemical properties of substances. Estrada et al. utilized the TOSS-MODE method and semiempirical models to predict diamagnetic properties.[15] Zhokhova et al. employed descriptors based on the fragment composition of explanatory molecules and a parameter V_x reflecting the molecular volume to predict the magnetic susceptibility of compounds.[16] Mu et al. have utilized molecular graph adjacency matrices ${}^{m}\chi'$ and variable atomic valency connectivity ${}^{m}\chi''$ indices to propose variable molecular connectivity indices and their inverse indices as feature values for predicting the molar magnetic susceptibility of substances.[12] Afantitis et al. have established a QSPR model, selecting three physical topological descriptors from 30 to predict the magnetic susceptibility of organic molecules.[17] Previous research has made strides in expediting the prediction of molar magnetic susceptibility in organic molecules through machine learning methods. However, the features used in these methods, ranging from molecular connectivity indices to physical topological descriptors, still necessitate computations using specific professional software. These features, while not exceedingly complex, do require a level of familiarity with specialized tools, for example, the calculation of physical topological descriptors such as polar surface area (PSAr), principal moment of inertia X (PMIX), and diameter (Diam) typically involves software like AutoDock and Gaussian. Similarly, molecular connectivity indices are often computed using applications such as ChemDraw or Open Babel. As a result, a considerable amount of time is required not only to learn how to use these software tools but also to perform the necessary calculations. This step remains a significant aspect of the process, indicating that while machine learning has streamlined some aspects of predicting molar magnetic susceptibility, the preparatory stages still involve a substantial investment of time and effort in learning and using specialized computational methods.

In this study, we aimed to rapidly predict the molar magnetic susceptibility of organic molecules using machine learning techniques and succeeded in developing a model that can directly predict this property from molecular structures. Our innovative model enables the prediction of an organic molecule's molar magnetic susceptibility simply by inputting its simplified molecular input line entry system (SMILES) string, which represents the molecule's structure. To identify the most effective predictive model, we investigated various combinations of molecular fingerprints and machine learning algorithms. Our machine learning model, trained on an extensive data set, effectively discerned the complex relationship between molecular structure and its physical properties. This capability allows for the rapid prediction of the magnetic properties of unknown substances. Our method not only accelerates the process of examining magnetic properties but also significantly reduces the costs associated with experimental and computational methods, all while maintaining high accuracy.

The implications of this research are substantial. It enhances our understanding of the magnetic properties of organic molecules and accelerates the discovery and development of new materials. Additionally, the application of this model extends to various related fields, marking a significant advancement in the study and application of materials science and molecular magnetism.

METHODS

Method Overview. The data for this study were sourced from the "CRC Handbook of Chemistry and Physics" (International Standard Book Number: 978-1-4987-5429-3, available in PDF format). The data set includes the diamagnetic values of approximately 400 common organic compounds. Table 1 lists the molar magnetic susceptibility χ_m and volumetric magnetic susceptibility κ values for some of these compounds. All data sets used in this work are open source at (https://github.com/ZHANGYNing/Diamagnetic-Susceptibility-of-Organic-Compounds).

Table 1. Values of Molar Magnetic Susceptibility and Volumetric Magnetic Susceptibility for Some Organic Compounds

name	mol. form	$-\chi_m/10^{-6}$ cm^3 mol^{-1}	$-\kappa/10^{-6}$
acenaphthene	$C_{12}H_{10}$	109.9	0.871
acenaphthylene	$C_{12}H_8$	111.6	0.659
acetaldehyde	C_2H_4O	22.2	0.395
acetamide	C_2H_5NO	33.9	0.573
acetic acid	$C_2H_4O_2$	31.8	0.553
acetic anhydride	$C_4H_6O_3$	52.8	0.56
acetone	C_3H_6O	33.8	0.457
acetonitrile	C_2H_3N	27.8	0.532
acetophenone	C_8H_8O	72.5	0.62
acetyl chloride	C_2H_3ClO	39.3	0.553

It is noteworthy that the susceptibility values listed in Table 1 of our study are all negative, signifying that the substances under consideration are diamagnetic. The majority of the volumetric magnetic susceptibility values in our data set represent substances in either a solid or liquid state at room temperature, with a smaller portion accounting for substances in a liquefied gaseous state. This diversity in the states of matter covered by our data prompted our decision to use the molar magnetic susceptibility as the prediction value in this research.

In conventional practice, the molar magnetic susceptibility is typically presented in units compatible with the CGS (Centimeter-Gram-Second) system. However, for integration into SI (International System of Units) equations, these values need to be multiplied by 4π to conform to the SI unit system.

https://doi.org/10.1021/acsomega.3c10469
ACS Omega 2024, 9, 14368−14374

In the SI system, the unit of magnetic field strength H is ampere per meter (A/m^{-1}).

Our research aimed to develop a predictive model that directly correlates molecular structure with molar magnetic susceptibility. This approach is intended to facilitate more efficient predictions of the magnetic properties of organic molecules. As illustrated in Figure 1, our initial step involved

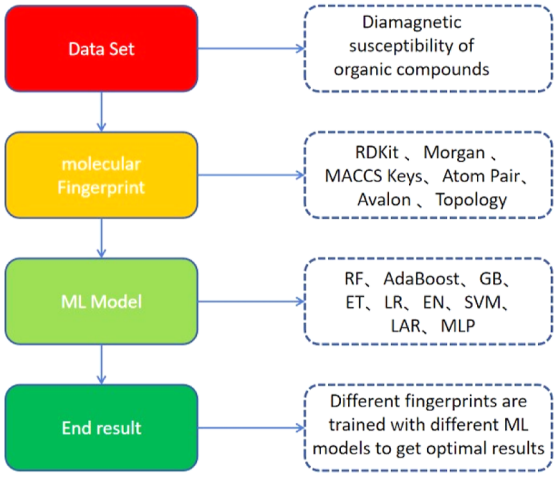


Figure 1. Schematic illustration of the workflow.

representing the molecules in our data set using various molecular fingerprinting techniques, such as RDKit fingerprint and Morgan fingerprint. This was done to thoroughly capture the chemical characteristics of these molecules.

To achieve optimal model development and assessment, we partitioned the data into three distinct sets: training, validation, and test sets. The training set was employed for the model's initial training phase. The validation set played a critical role in fine-tuning the model parameters during training, ensuring the optimization of the model's performance and its ability to generalize. The test set was crucial for evaluating the final performance of the model, confirming its efficacy in accurately predicting the characteristics of previously unseen data. This structured approach ensured the comprehensive and effective development of the predictive model.

After the meticulous processing and division of the data, our study involved the utilization of various machine learning models for the purpose of training. These models included multilayer perceptrons (MLP), support vector machines (SVM), and random forests, among others. Each model was subjected to comprehensive training and fine-tuning processes to ensure that it achieved optimal predictive performance. The effectiveness of these models was assessed by comparing their performance on both the validation and test sets. This comparison was crucial for identifying the model that demonstrated the most superior performance in terms of prediction accuracy.

Molecular Fingerprinting. Molecular fingerprinting is a digital representation method that converts chemical structures into a series of numerical codes. This representation is widely used in cheminformatics, facilitating the rapid comparison and analysis of similarities and differences between different molecules using computer algorithms. Moreover, the fundamental principles and algorithms for generating molecular fingerprints have been extensively studied.[18−23] In this study, we directly utilized RDKit (http://www.rdkit.org/) to compute six different types of molecular fingerprints, which are listed below. These molecular fingerprinting methods each have their unique characteristics, offering a diverse range of perspectives and tools for studying the relationship between molecular structure and their molar magnetic susceptibility.

- RDKit fingerprint.[24] This method generates bit vectors based on the substructures and chemical features of the molecule. It is suitable for rapid screening and comparison of molecular similarity.
- Morgan fingerprint (circular fingerprint).[24] Similar to the widely used extended-connectivity fingerprints (ECFP) in chemistry, this method generates fingerprints based on the circular neighborhood of atoms around a molecule. It is applicable for predicting the similarity and properties of complex molecular structures.
- MACCS keys fingerprint.[24,25] This fingerprint includes 166 predefined chemical structure keys, representing various chemical features that may appear in a molecule. It is commonly used for compound database searches and molecular similarity comparisons.
- Atom pair fingerprint.[24,26] Describes the distance and type between pairs of atoms. It is suitable for capturing features related to the distances between atoms within a molecule.
- Avalon fingerprint.[24] A fingerprint based on hashing algorithms. It can generate larger bit vectors, providing a rich molecular description.
- Topology fingerprint.[24] Based on the molecular topology, such as the arrangement of rings and bonds. It is suitable for describing the overall architecture and topological characteristics of a molecule.

Model Building. In our work, given that quantitative structure−property relationships (QSPR) predict properties from molecular structures,[27] typically involving complex nonlinear relationships, we experimented with various machine learning models, selecting the one that performed best on the given data set. Throughout the development of these models, we conducted meticulous parameter tuning and cross-validation to ensure optimal performance on our data set. We compared these models across different data sets (training, validation, and test sets) to assess their effectiveness in predicting the molar magnetic susceptibility of organic molecules. This multimodel comparison process not only enhanced our understanding of the data but also provided a solid foundation for ultimately selecting the most suitable model.

- Random forest (RF).[28] An ensemble learning method that builds multiple decision trees and combines their predictions to enhance accuracy and stability.
- AdaBoost.[29] An adaptive boosting algorithm that sequentially connects multiple weak learners, gradually adjusting the weight of each learner to improve model performance.
- Gradient boosting (GB).[30] Adds weak learners sequentially, optimizing each addition based on the residuals of the previous step to enhance predictive capabilities.
- Extra trees (ET).[31] Similar to Random Forest but uses a more random method of splitting in tree construction, aimed at reducing variance.
- Elastic net (EN).[32] Combines features of Lasso and Ridge regression, optimizing model performance through adjustment of regularization parameters.

https://doi.org/10.1021/acsomega.3c10469
ACS Omega 2024, 9, 14368−14374

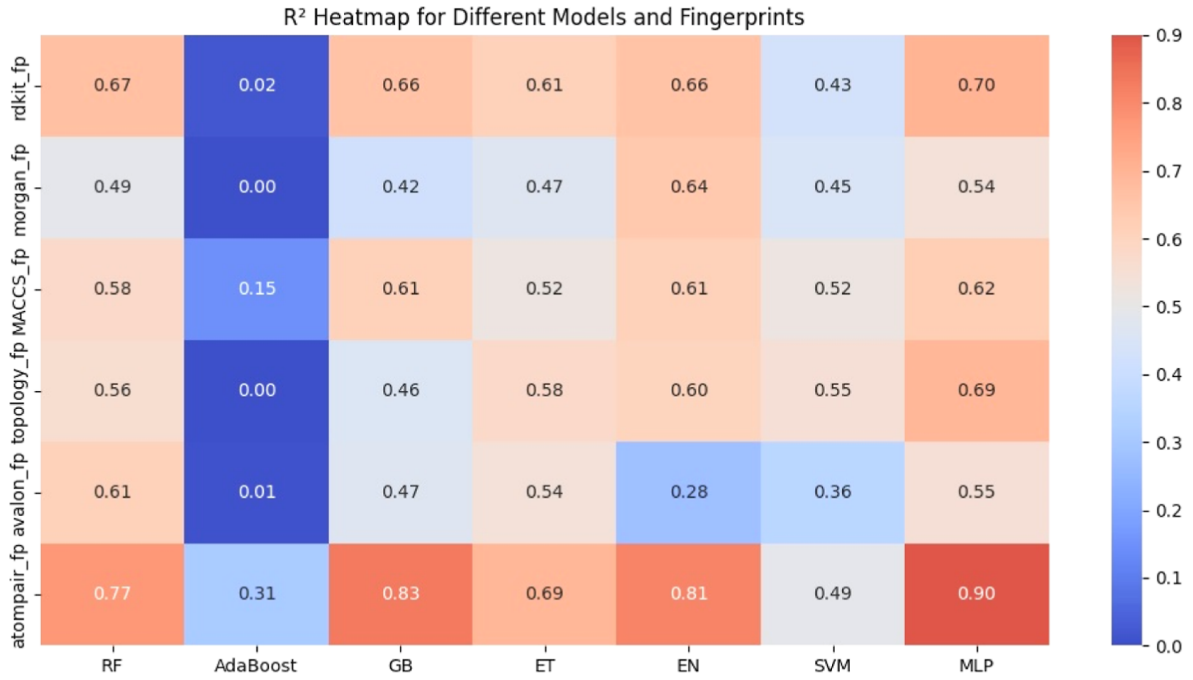


Figure 2. (R^2) heatmap for different models and fingerprints.

- Support vector machine (SVM).[33] A powerful classifier that finds the optimal hyperplane in high-dimensional space to separate different categories.
- Multi-layer perceptron (MLP).[34] A neural network-based method with strong nonlinear fitting capabilities, suitable for predicting complex data patterns.

Model Evaluation. In this study, we employed three primary statistical metrics to evaluate the performance of the selected machine learning models: the coefficient of determination (R^2), the root-mean-square error (RMSE), the mean squared error (MSE), and the mean absolute error (MAE).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

In these metrics, y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the average of actual values, and n the number of samples. We used these indicators for a comprehensive evaluation of all models. R^2 offers insights into the overall fitting of the model; RMSE, MSE, and MAE measure the model's prediction error from different perspectives. By comparing the performance of different models on these metrics, we were able to objectively determine which models are best at predicting the molar magnetic susceptibility of the organic molecules. This comprehensive evaluation approach ensured that our analysis was both thorough and unbiased, providing a solid basis for the final selection of the model.

RESULTS AND DISCUSSION

In our study, we initially utilized various molecular fingerprints as input features and trained multiple machine learning models. The coefficient of determination (R^2) heatmap analysis, as shown in Figure 2, revealed that the atom pair fingerprint was the most effective feature for training. This finding underscores the impact of factors such as interatomic distances, atom types, their connectivity, and the arrangement of electrons in three-dimensional space on the molar magnetic susceptibility of organic molecules. Molar magnetic susceptibility, a key physical quantity measuring a molecule's magnetic response in an external field, is heavily influenced by the electron distribution within the molecule and the interactions between atoms. The atom pair fingerprint, capturing the distances and relationships between atom pairs within a molecule, effectively encodes these critical aspects of electron arrangement and interatomic interactions, which are crucial for predicting molar magnetic susceptibility.

Among the evaluated machine learning models, the MLP demonstrated the highest performance. As a deep feed-forward neural network, MLP excels at fitting complex functions, enabling it to discern more intricate relationships and patterns than traditional linear models. MLP's hidden layers allow it to represent higher-level and more complex features, which may be challenging for linear models. Additionally, the nonlinear activation functions in MLP, such as ReLU, Sigmoid, or Tanh, enable it to effectively learn and fit nonlinear relationships. Optimized through algorithms such as gradient descent, MLP can adjust weights to better fit training data and make precise predictions. Consequently, the combination of the atom pair fingerprint with MLP achieved an (R^2) of 0.9 on the test set, with MSE and MAE values of 0.14 and 0.0015, respectively.

As shown in Figure 3, the combination of atom pair fingerprint with MLP demonstrated high consistency between predicted and actual values across training, validation, and test sets, further validating the model's accuracy in predicting the molar magnetic susceptibility of organic molecules. The

https://doi.org/10.1021/acsomega.3c10469
ACS Omega 2024, 9, 14368−14374

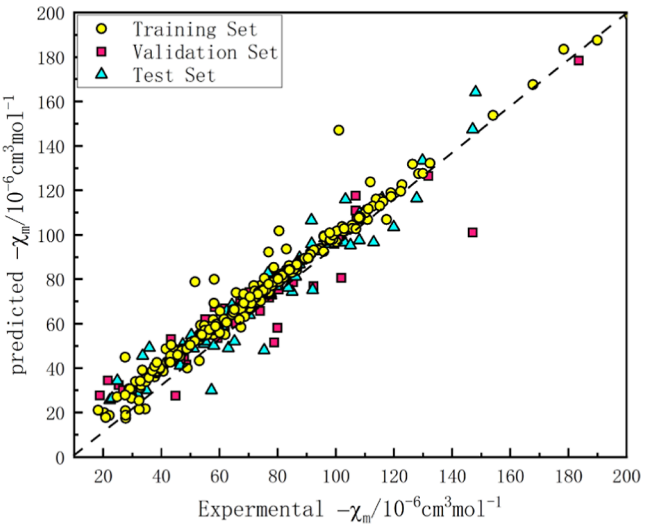


Figure 3. Diagram showing the scatter of the experimental and predicted values of magnetic susceptibility according to the combined models of MLP and atom pair fingerprint.

histogram of prediction errors for the validation and test sets, displayed in Figure 4, shows that the errors are approximately normally distributed and centered around zero, indicating that the model's predictions are neither systematically biased nor overly uncertain. Figure 5 displays the performance of different models in terms of MSE and MAE, objectively proving the superiority of the combination of MLP and atom pair fingerprint in predicting the molar magnetic susceptibility of organic molecules.

In conclusion, the atom pair fingerprint provided MLP with a comprehensive set of input features, enhancing the model's understanding and learning of the complex relationship between molar magnetic susceptibility and molecular structure. This combination not only bolstered the model's robustness against noise and outliers but also established a highly effective tool for predicting the molar magnetic susceptibility of organic molecules.

CONCLUSIONS

In this study, we used six types of molecular fingerprints as features to train seven different machine learning models. The comparative analysis revealed that the combination of the atom pair fingerprint and MLP exhibited exceptional performance in predicting the molar magnetic susceptibility of organic molecules, as evidenced by an (R^2) value of 0.9, a MAE of 0.0014, and a MSE of 0.15. This outcome underscores the effectiveness of atom pair fingerprint in capturing critical molecular features that influence magnetic susceptibility, and it also demonstrates the potent capability of MLP in processing complex data and learning nonlinear relationships.

Moreover, the methodology and findings of this study are of substantial importance in advancing the research and development of the magnetic properties of organic molecules. The integration of machine learning models into this research field not only expedites the research process but also significantly reduces the costs associated with experiments and computational analyses, all while maintaining high accuracy. This innovative approach opens up new avenues for the rapid screening and property prediction of compounds, which could have a significant impact in various domains, including material science, pharmaceutical design, and other related fields. The application of these advanced machine learning techniques in predicting molar magnetic susceptibility represents a major step forward in the understanding and development of new materials and compounds. In our future work, we plan to apply our model to a variety of molecular data sets to enhance its applicability and robustness. Additionally, exploring other advanced machine learning techniques, such as deep learning and ensemble methods, will be part of our ongoing efforts. This will not only expand the capabilities of our model but also offer a wider perspective on the application of machine learning in the prediction of chemical properties.

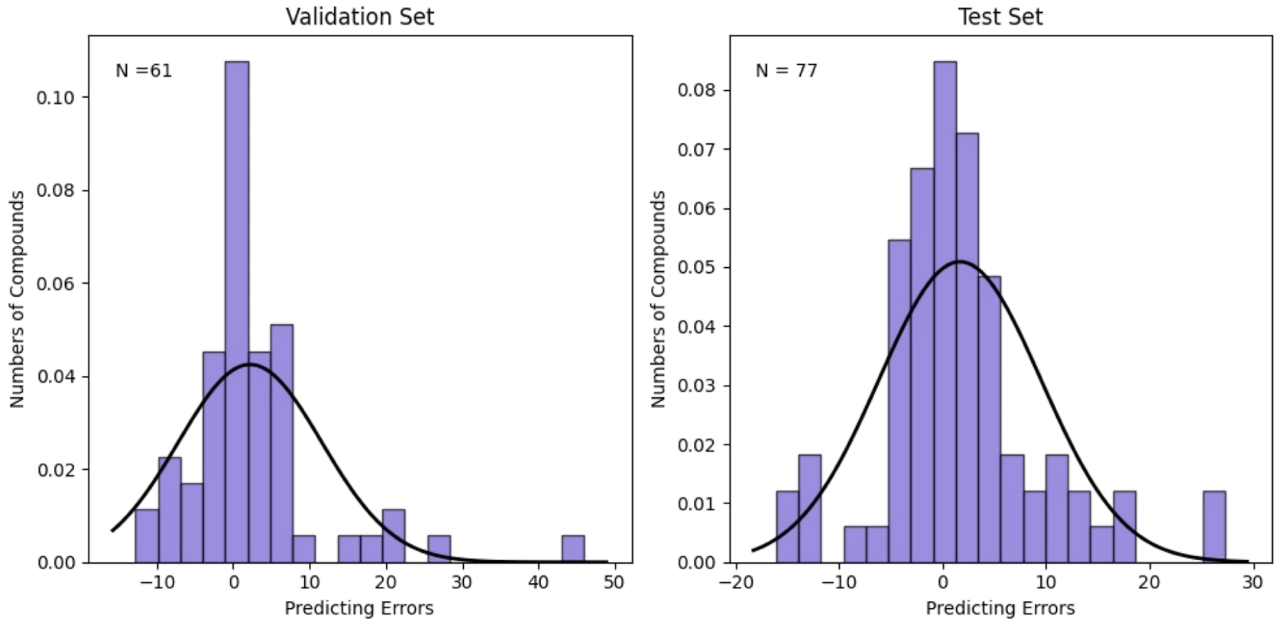


Figure 4. Histogram of prediction errors for the validation set and test set.

https://doi.org/10.1021/acsomega.3c10469
ACS Omega 2024, 9, 14368−14374

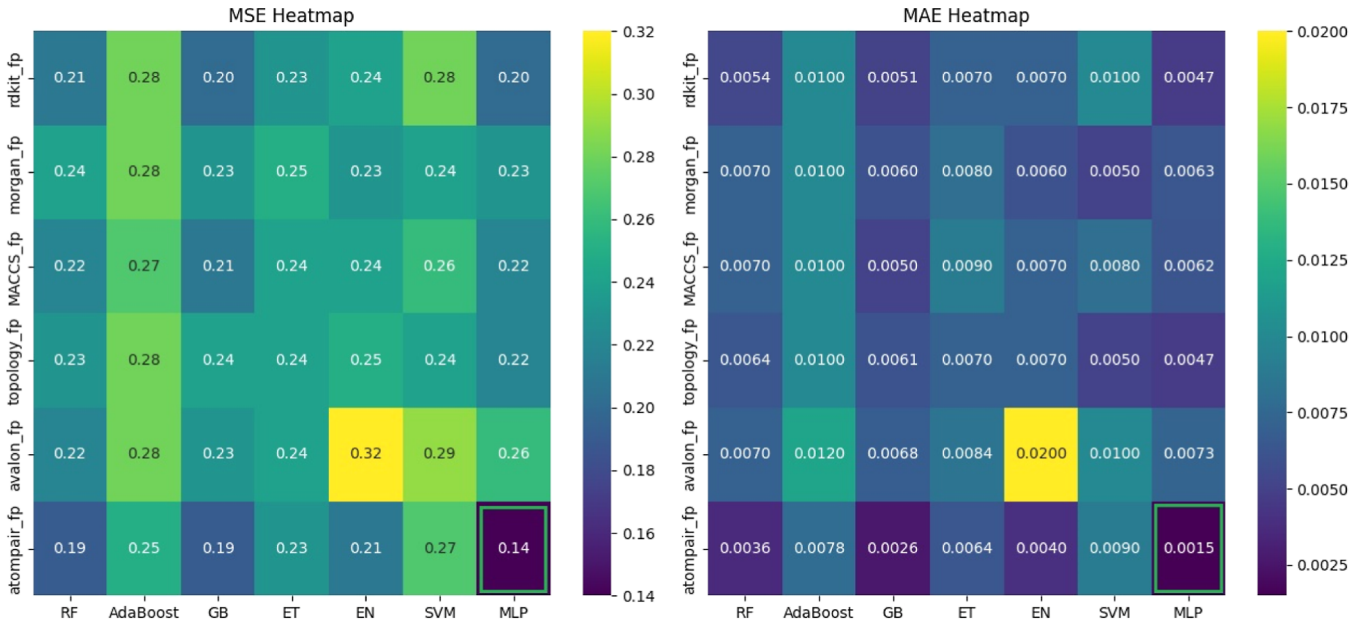


Figure 5. MSE and MAE heatmaps for different models and fingerprints.

AUTHOR INFORMATION

Corresponding Authors

Lai Wei − *Xinjiang Laboratory of Phase Transitions and Microstructures in Condensed Matter Physics, College of Physical Science and Technology, Yili Normal University, Yining 835000, China*; Email: lweiphy@sina.com

Tongfei Shi − *School of Chemical Engineering and Light Industry, Guangdong University of Technology, Guangzhou 510006, People's Republic of China; Xinjiang Laboratory of Phase Transitions and Microstructures in Condensed Matter Physics, College of Physical Science and Technology, Yili Normal University, Yining 835000, China*; orcid.org/0000-0002-6763-2200; Email: tfshi@gdut.edu.cn

Authors

Yining Zhang − *Xinjiang Laboratory of Phase Transitions and Microstructures in Condensed Matter Physics, College of Physical Science and Technology, Yili Normal University, Yining 835000, China*

Sijie Xing − *Alibaba Cloud Big Data Application College, Zhuhai College of Science and Technology, Zhuhai 519041, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c10469

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China (grant no. 22073093).

REFERENCES

(1) Selwood, P. W. *Magnetochemistry*, 2nd ed.; Interscience Publishers New York: New York, 1956.

(2) Vulfson, S. G. *Molecular Magnetochemistry*; Gordon & Breach: Amsterdam, 1998.

(3) Dorfman, Y. G. *Diamagnetism and the Chemical Bond*; American Elsevier Publishing Co.: New York, 1965.

(4) Atkins, A. T. *Physical Chemistry*; Oxford University Press: Oxford, 1986.

(5) Himanen, L.; Geurts, A.; Foster, A. S.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv. Sci.* **2019**, *6*, 1900808.

(6) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122*, 17575−17585.

(7) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **2017**, *3*, 54.

(8) Schleder, G. R.; Padilha, A. C. M.; Acosta, C. M.; Costa, M.; Fazzio, A. From DFT to machine learning: recent approaches to materials science-a review. *J. Phys. Mater.* **2019**, *2*, 032001.

(9) Zhang, Y.; Wen, C.; Wang, C.; Antonov, S.; Xue, D.; Bai, Y.; Su, Y. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* **2020**, *185*, 528−539.

(10) Khveshchenko, D. V.; Meshkov, S. V. Particle in a random magnetic field on a plane. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 12051−12058.

(11) Putz, M. V.; Russo, N.; Sicilia, E. Atomic radii scale and related size properties from density functional electronegativity formulation. *J. Phys. Chem. A* **2003**, *107*, 5461−5465.

(12) Mu, L. L.; He, H. M.; Yang, W. H. Improved QSPR Study of Diamagnetic Susceptibilities for Organic Compounds Using Two Novel Molecular Connectivity Indexes. *Chin. J. Chem.* **2009**, *27*, 1045−1054.

(13) Mu, L. L.; Feng, C. G. Topological research on molar diamagnetic susceptibilities for inorganic compounds. *MATCH Commun. Math. Comput. Chem.* **2007**, *58*, 591−607.

(14) Estrada, E. Modelling the diamagnetic susceptibility of organic compounds by a sub-structural graph-theoretical approach. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1407−1410.

(15) Estrada, E.; Gutierrez, Y.; González, H. Modeling diamagnetic and magnetooptic properties of organic compounds with the TOSS-MODE approach. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1386−1399.

(16) Zhokhova, N. I.; Baskin, I.; Palyulin, V. A.; Zefirov, A. N.; Zefirov, N. S. Fragment descriptors in QSPR: Application to magnetic susceptibility calculations. *J. Struct. Chem.* **2004**, *45*, 626−635.

(17) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O. Development and

evaluation of a QSPR model for the prediction of diamagnetic susceptibility. *QSAR Comb. Sci.* **2008**, *27*, 432−436.
(18) Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58−63.
(19) Dong, J.; Cao, D. S.; Miao, H. Y.; Liu, S.; Deng, B. C.; Yun, Y. H.; Wang, N. N.; Lu, A. P.; Zeng, W. B.; Chen, A. F. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminf.* **2015**, *7*, 60.
(20) Rahaman, O.; Gagliardi, A. Deep Learning Total Energies and Orbital Energies of Large Organic Molecules Using Hybridization of Molecular Fingerprints. *J. Chem. Inf. Model.* **2020**, *60*, 5971−5983.
(21) Botu, V.; Ramprasad, R. Adaptive machine learning framework to accelerate *ab initio* molecular dynamics. *Int. J. Quantum Chem.* **2015**, *115*, 1074−1083.
(22) Que, Y.; Ren, S.; Hu, Z.; Ren, J. Machine Learning Prediction of Critical Temperature of Organic Refrigerants by Molecular Topology. *Processes* **2022**, *10*, 577.
(23) Townsend, J.; Micucci, C. P.; Hymel, J. H.; Maroulas, V.; Vogiatzis, K. D. Representation of molecular structures with persistent homology for machine learning applications in chemistry. *Nat. Commun.* **2020**, *11*, 3230.
(24) James, C.; Weininger, D.; Delaney, J. D. *Daylight Theory Manual Version 4.9*; Daylight Chemical Information Systems, Inc.: Laguna Niguel, CA, USA, 2011.
(25) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−1280.
(26) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.
(27) Gantzer, P.; Creton, B.; Nieto-Draghi, C. Inverse-QSPR for *de novo* Design: A Review. *Mol. Inf.* **2020**, *39*, 21.
(28) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(29) Bauer, E.; Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach. Learn.* **1999**, *36*, 105−139.
(30) Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorob.* **2013**, *7*, 21.
(31) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3−42.
(32) Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc. B Stat. Methodol.* **2005**, *67*, 301−320.
(33) Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273−297.
(34) Mat Isa, N. A.; Mamat, W. Clustered-Hybrid Multilayer Perceptron network for pattern recognition application. *Appl. Soft Comput.* **2011**, *11*, 1457−1466.

https://doi.org/10.1021/acsomega.3c10469
ACS Omega 2024, 9, 14368−14374