

# Accurate Density Prediction of Sesquiterpenoid HEDFs and the Multiproperty Computing Server SesquiterPre

Hang Yang,<sup>#</sup> Zhi-Jiang Yang,<sup>#</sup> Teng-Xin Huang, Li Pan, Xin-Miao Wei, Yan-Fei Hu, Yu-Quan Yuan,<sup>\*</sup> Liang-Liang Wang,<sup>\*</sup> and Jun-Jie Ding<sup>\*</sup>



Cite This: *ACS Omega* 2024, 9, 26213–26221



Read Online

ACCESS |



Metrics & More

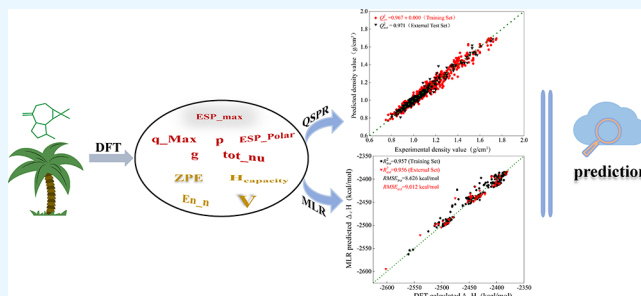


Article Recommendations



Supporting Information

**ABSTRACT:** Accurate and rapid evaluation of density is crucial for evaluating the packing and combustion characteristics of high-energy-density fuels (HEDFs). This parameter is pivotal in the selection of high-performance HEDFs. Our study leveraged a polycyclic compound density data set and quantum chemical (QC) descriptors to establish a correlation with the target properties using the XGBoost algorithm. We utilized a recursive feature elimination method to simplify the model and developed a concise and interpretable density prediction model incorporating only six QC descriptors. The model demonstrated robust performance, achieving coefficients of determination ( $R^2$ ) of 0.967 and 0.971 for internal and external test sets, respectively, and root-mean-square errors (RMSE) of 0.031 and 0.027 g/cm<sup>3</sup>, respectively. Compared to the other two mainstream methods, the marginal discrepancy between the predicted and actual molecular densities underscores the model's superior predictive ability and more usefulness for energy density calculation. Furthermore, we developed a web server (SesquiterPre, <https://sespre.cmdrg.com/#/>) that can simultaneously calculate the density, enthalpy of combustion, and energy density of sesquiterpenoid HEDFs, which greatly facilitates the use of researchers and is of great significance for accelerating the design and screening of novel sesquiterpenoid HEDFs.



## 1. INTRODUCTION

High-energy-density fuels (HEDFs) garner significant attention owing to their substantial density and energy density,<sup>1–5</sup> which enhance the payload and flight range of aircraft. Prior research indicates that employed HEDFs possess high-strain polycyclic hydrocarbon structures,<sup>6–12</sup> as seen in fuels such as RJ-4, RJ-5, JP-10, Syntin, and caged hydrocarbon fuels. However, synthesizing these HEDFs typically entails hazardous and toxic intermediates derived from petroleum,<sup>13,14</sup> making the process costly, perilous, and complex. Research has demonstrated that these molecules can be synthesized through simpler chemical and microbial methods,<sup>15–19</sup> presenting potential candidates for renewable HEDFs. Notably, these renewable fuels match the density and net heat of combustion of the specialty fuel JP-10. Currently, the inefficient yield and purity of sesquiterpenoids obtained through these methods pose significant challenges for the rapid, accurate, and large-scale experimental characterization of these compounds.<sup>16,19</sup> To overcome these obstacles, dependable theoretical computations for evaluating the density and energy of sesquiterpenoid HEDFs can offer a potent solution.

Density is a crucial parameter in gauging the packing density and combustion characteristics of HEDFs. The rapid and accurate evaluation of density is vital for logically selecting new high-performance HEDFs and for calculation of their energy

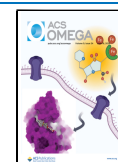
density. Over recent decades, numerous researchers have strived to predict the density of various compounds.<sup>20–27</sup> Previously, cubic equations of state (EoS), such as SRK EoS,<sup>28</sup> PR EoS,<sup>29</sup> and simple RM EoS,<sup>30</sup> have been commonly used for hydrocarbon density prediction. Although these EoS methods predict hydrocarbon density with commendable accuracy and a relative error of less than 5%, their calculation process requires the resolution of multiple parameters, leading to low efficiency and complicated utilization. Moreover, it is assumed that the molecule is a three-dimensional structure surrounded by a 0.001 au electron density.<sup>31</sup> The electrostatic potential parameter at a 0.001 au electron density can quantitatively depict the physical and chemical properties of the compound.<sup>32</sup> Qiu et al.<sup>33</sup> and Rice et al.<sup>34</sup> suggested a direct density prediction method within the density functional theory (DFT) framework, where molecular density is calculated as mass divided by volume. However, this method often produces density values higher than the experimental values, leading to significant errors in

Received: February 27, 2024

Revised: May 24, 2024

Accepted: May 24, 2024

Published: June 5, 2024



subsequent tests. Politzer et al.<sup>35</sup> later introduced a parameter,  $\nu_{\text{tov}}^2$ , which provides a better explanation of the intermolecular interactions for predicting the density of neutral compounds. The results highlight the high reliability of this method, which has an average absolute error of approximately 3% and a strong coefficient of determination  $R^2$  of 0.924. Additionally, molecular dynamics<sup>36,37</sup> and group contribution methods<sup>38,39</sup> have also proven successful in more accurately predicting molecular density.

In addition to established density prediction methods, a newer strategy rapidly determines the molecular density of compounds by constructing a quantitative mathematical function that correlates the density with molecular structural characteristics. This approach leverages machine learning (ML) techniques. With advancements in computing and intelligent algorithms, researchers have developed several quantitative structure–property relationship (QSPR) models.<sup>40–43</sup> These models combine intelligent algorithms with various molecular descriptors, including those predicting the density of energy-containing materials.<sup>43–45</sup> The excellent predictive performance of these models confirms the feasibility of this approach. However, the correlation between molecular properties and their geometric or electronic structures remains unclear and undefined, rendering theoretical models less effective for guiding the design and development of novel HEDFs. Considering the sensitivity of HEDFs' properties to their molecular structures,<sup>46,47</sup> it is essential to investigate how the microstructures of sesquiterpenoid HEDFs impact their density properties. Moreover, the energy density of HEDFs can be directly derived from the mass net heat of combustion, which is obtained by converting the enthalpy of combustion combined with the corresponding density.

At present, no reports on density predictions or energy density calculations for sesquiterpenoid HEDFs are available, and experimental data on sesquiterpene molecules are scarce. In light of this, we explore density prediction by utilizing a transfer learning approach. Initially, we constructed a QSPR model for predicting the density of sesquiterpenoid HEDFs. This was based on a collected data set of 1567 molecules, bearing structural similarity to sesquiterpenes with experimental density values, and a quantum chemical (QC) descriptor derived within the framework of DFT. Subsequently, using the Shapley additive explanation (SHAP) method, we identified key QC features influencing the density of sesquiterpenoid HEDFs. This not only provides a theoretical foundation for the development and design of novel HEDFs but also addresses the interpretability issue of the QSPR model. Finally, for the benefit of users, we incorporated these models into an automated web server that predicts the density, enthalpy of combustion, and energy density of sesquiterpenoid HEDFs.

## 2. MATERIALS AND METHODS

**2.1. Data Collection and Preprocessing.** Given the limited experimental density data for sesquiterpenoid HEDFs, the corresponding QSPR prediction models could not be directly constructed. As a result, we curated a data set from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>), consisting of 2100 compounds with molecular structures resembling the target molecule composition and documented experimental density values. To ensure the quality and accuracy of the data, the acquired cyclic compound collection underwent deweighting, desalting, and meticulous manual screening. This process yielded a data set of 1567 molecules suitable for subsequent

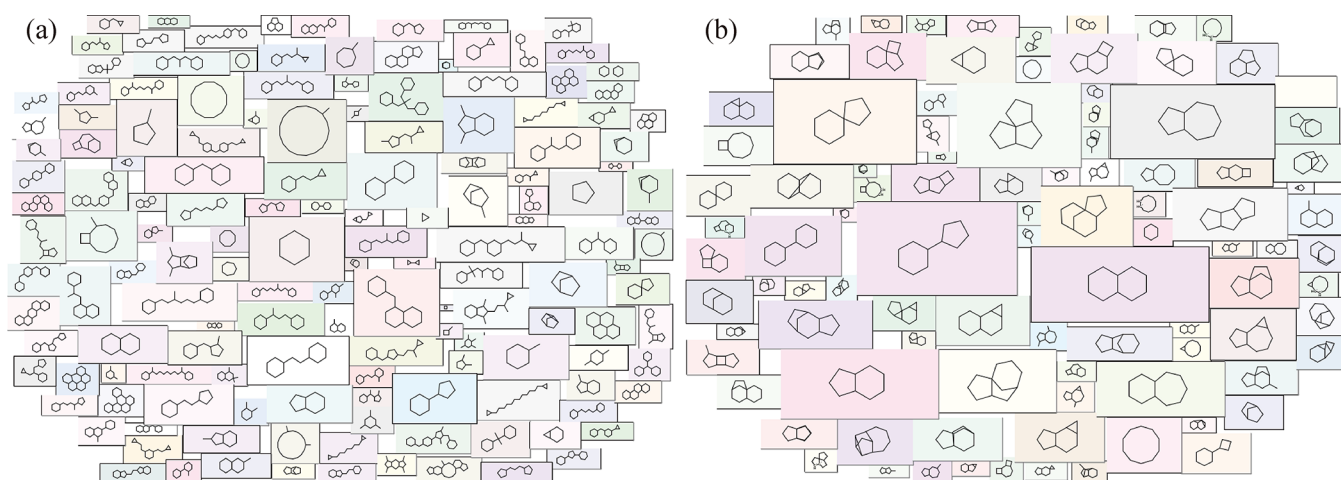
QSPR modeling. Notably, these data sets consist entirely of experimental density values at normal pressure and temperature. This enables the direct calculation of the energy density of sesquiterpenoid HEDFs at normal pressure and temperature based on the existing data sets of their enthalpy of combustion. Additionally, the structures of these cyclic compounds were illustrated using GaussView 6.0 software<sup>48</sup> to aid in calculating and extracting their QC descriptors.

### 2.2. Calculation of Quantum Chemical Descriptors.

The nature of a molecule is intrinsically determined by its structure, and QC descriptors effectively characterize its electronic structure and reactivity.<sup>49–51</sup> These descriptors offered insights into the properties of sesquiterpene molecules at the level of reaction mechanisms enhancing the interpretability of QSPR models. The calculation of QC descriptors involved three primary steps: (1) We used the GMMX 3.0 conformational search module in GaussView 6.0 software to conduct a conformational search of the initially constructed structures based on the MMFF94 force field. (2) Utilizing Gaussian 16<sup>52</sup> software, we selected the top 10 low-energy conformations (from the lowest to the highest energy) from the conformational search results for geometric optimization and frequency calculation at the B3LYP/6-311+G(d)<sup>53,54</sup> theory level. We then chose the conformations with the lowest energy and without imaginary frequency as the initial structures for QC descriptor calculation. Geometric structure optimization and frequency analysis of the low-energy conformations were conducted using the M06-2X/6-31G(d)-D3(BJ)/SMD (solvation model based on density).<sup>37,55,56</sup> (3) At the same theoretical level, single-point energy calculation, and natural bond orbitals,<sup>57</sup> analyses were conducted for molecules in the neutral, one-electron gain, and one-electron loss states of the lowest energy conformation. Next, quantitative molecular surface analysis of the wavefunction file (.chk) computed was performed using the Multiwfn package.<sup>58</sup> Finally, our self-developed software Quantum 1.0, in combination with Multiwfn, was employed for the batch extraction of 54 QC descriptors.

**2.3. Model Development and Performance Evaluation.**  
**2.3.1. Construction of the QSPR Model.** Extreme gradient boosting (XGBoost), a cutting-edge ensemble learning algorithm, has been successfully employed in recent ML competitions and data processing tasks.<sup>59–61</sup> Operating under the gradient boosting framework, XGBoost facilitates more efficient and scalable development, providing insights into cached access patterns, data compression, and fragmentation.<sup>62</sup> In this study, we employed the XGBoost algorithm to construct a QSPR model using QC descriptors and a density data set. This method allowed us to verify the correlation between various molecular features and the density.

For our QSPR model, the density data set was divided into an internal training set and an external test set, based on the type of carbon skeleton in the polycyclic structure compounds, at a 4:1 ratio. To ensure that the model maintained a strong generalization ability in predicting density, we implemented a 5-fold cross-validation approach on the internal training set. Concurrently, to mitigate overfitting and ensure model stability, we repeated the 5-fold cross-validation 1000 times. We then evaluated the quality of the QSPR model using six main statistical parameters: the coefficient of determination for cross-validation ( $Q_{\text{cv}}^2$ ), root-mean-square error for cross-validation ( $\text{RMSE}_{\text{cv}}$ ), mean absolute error for cross-validation ( $\text{MAE}_{\text{cv}}$ ), coefficient of determination for the external test set ( $Q_{\text{ext}}^2$ ), root-



**Figure 1.** Carbon skeletons of 1567 compounds (a) and sesquiterpenoid HEDFs (b).

**Table 1.** Evaluation Parameters of the QSPR Model

model	number of QC descriptors	XGBoost					
		training set			external test set		
		$Q_{cv}^2$	RMSE <sub>cv</sub> (g/cm <sup>3</sup> )	MAE <sub>cv</sub> (g/cm <sup>3</sup> )	$Q_{ext}^2$	RMSE <sub>ext</sub> (g/cm <sup>3</sup> )	MAE <sub>ext</sub> (g/cm <sup>3</sup> )
old model	54	0.973 ± 0.000	0.028 ± 0.000	0.018 ± 0.000	0.975	0.026	0.018
new model	6	0.967 ± 0.000	0.031 ± 0.000	0.020 ± 0.000	0.971	0.027	0.019

mean-square error for the external test set (RMSE<sub>ext</sub>), and mean absolute error for the external test set (MAE<sub>ext</sub>).

**2.3.2. Feature Selection.** To further simplify our model and enhance its interpretability, we implemented the recursive feature elimination (RFE)<sup>63</sup> method, which is based on XGBoost, for QC descriptor screening. The operational steps of this feature selection approach are as follows: (i) We built a density prediction model using the QC descriptors from the training set after removing redundant variables. Each descriptor was ranked based on its importance; (ii) descriptors with the lowest importance were eliminated, and the density model was reconstructed with the remaining descriptors. This process yielded a new priority ranking of descriptors in terms of their importance. We repeated this process until all descriptors were removed; (iii) we plotted the RMSE<sub>cv</sub> against the corresponding number of descriptors. From this plot, the concise and high-performing QSPR model as a new density prediction model was selected. Subsequently, we utilized a transfer learning approach<sup>64</sup> to predict density based on the new model parameters combined with the QC descriptors of the target molecules.

**2.3.3. QSPR Model Explanation.** Additionally, using the SHAP method proposed by Lundberg and Lee,<sup>65</sup> we analytically calculated the correlations between these feature vectors and target attributes. This approach addresses the common issue of the lack of interpretability in high-performing models and accelerates the development of novel sesquiterpenoid HEDFs with desired properties.

**2.4. Multiproperty Computing Web Server.** The integration of QC and ML offers a robust solution for the rapid and accurate prediction of properties for both existing and potential new sesquiterpene molecules. This approach can significantly mitigate the extensive and laborious experimental work, decrease financial costs, and provide valuable theoretical and technical support for the early assessment and study of biomass sesquiterpenoid HEDFs. In this study, we incorporated

a density prediction model and a straightforward multiple linear regression (MLR) equation<sup>66</sup> for the enthalpy of combustion into a web server. This server enables direct prediction of the density, enthalpy of combustion, and energy density of sesquiterpenoid HEDFs based on QC descriptors obtained from the DFT framework. The calculation for enthalpy of combustion, mass net heat of combustion, and energy density follows the given methods:

$$y = -1.557V - 5.960ZPE - 3.639H_{\text{capacity}} - 11.395En\_n - 176,862.174 \quad (1)$$

$$NHOC_G = \frac{y}{M} \quad (2)$$

$$NHOC_V = NHOC_G \cdot \rho \quad (3)$$

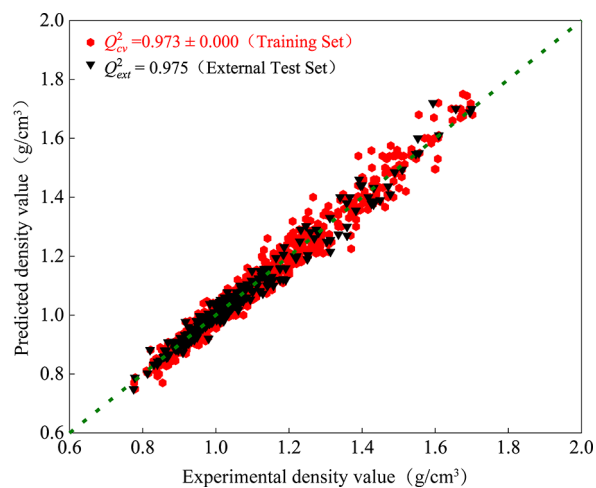
where  $y$  denotes the enthalpy of combustion of the molecule,  $V$  is defined as the volume,  $ZPE$  refers to the zero-point energy of the molecule after structural optimization and frequency calculation,  $H_{\text{capacity}}$  is the constant heat capacity,  $En\_n$  represents the single-point energy of the molecule in the negative state,  $M$  denotes the molecular mass,  $NHOC_G$  is the net mass heat of combustion of the target molecule, and  $\rho$  and  $NHOC_V$  are defined as the density and energy density, respectively.

### 3. RESULTS AND DISCUSSION

**3.1. Skeletal Analysis.** We examined the skeleton structure of 1567 compounds, focusing specifically on the core structure by analyzing the retained ring systems after all substituents were removed from the molecules. Our analysis identified a total of 150 distinct carbon skeletons, which predominantly included macrocycles, simple rings, complex heterocycles, and polycyclic structures. The detailed results are displayed in Figure 1a. The skeletons of the compounds, as revealed by our analysis, cover nearly all of the structural types of sesquiterpenoid HEDFs, as

shown in Figure 1b. This comprehensive range facilitates the accurate prediction of the density of target molecules using the transfer learning method.

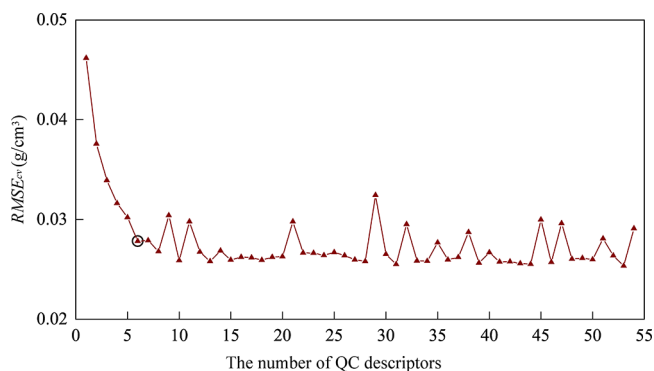
**3.2. Performance Evaluation of the QSPR Model.** In this study, we employed the XGBoost algorithm along with 54 QC descriptors and density data sets to develop a QSPR model for predicting the density of sesquiterpenoid HEDFs. To ensure robust generalization and stability, we incorporated a 5-fold cross-validation technique within the internal training set. The validation results for both the internal and external test sets of the QSPR model are detailed in Table 1, and Figure 2 presents



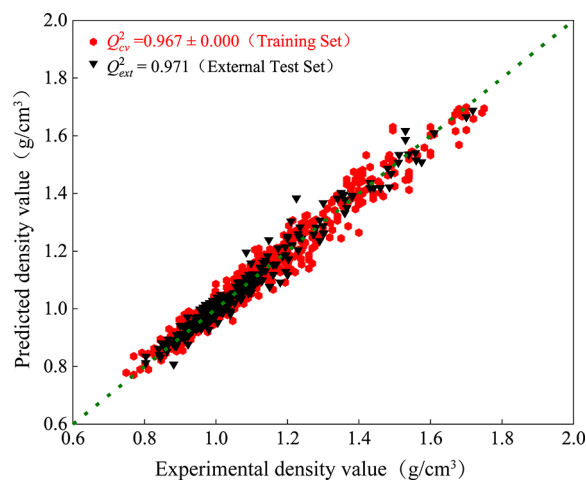
**Figure 2.** Relationship between experimental values of the density and predicted values.

the relationship between the experimental and predicted density values for both the training and test sets. Specifically,  $Q_{cv}^2$  of the 5-fold cross-validation ranged from 0.966 to 0.976, and  $RMSE_{cv}$  ranged from 0.026 to 0.031 g/cm<sup>3</sup>. For the external test set,  $Q_{ext}^2$  was 0.975, and  $RMSE_{ext}$  was 0.026 g/cm<sup>3</sup>. These results suggested that the QSPR model delivers both excellent and stable predictive performance with minimal errors across all statistical results. Moreover, we demonstrated a strong correlation between the QC descriptors and the predicted target. Given the complexity of the original model, it was not suitable for the rapid prediction of sesquiterpenoid HEDFs' density in online integration tools. Therefore, we employed the RFE method to create a series of simplified density prediction models based on the importance rankings of the features. Figure 3 illustrates the relationship between the  $RMSE_{cv}$  of the 5-fold cross-validation and the number of descriptors used. By balancing model complexity and performance, as well as the usability of prediction tools, we determined that a new model containing only six QC descriptors was optimal. As shown in Table 1, this new model, built on the reduced set of six QC descriptors, maintains performance comparable to the original model while simplifying its complexity. The relationship between the predicted density values of the new model and the experimental values is presented in Figure 4.

To further validate the predictive performance of our new model, we compared it with two widely accepted density prediction methods: ACD/Labs software (<https://www.acdlabs.com/>) and a linear equation.<sup>35</sup> We applied these methods to predict the density of the 1567 compounds. Table 2 summarizes the evaluation parameters of these methods, while Figure 5 illustrates the relationship between the predicted and



**Figure 3.** Relationship between the  $RMSE_{cv}$  of models and the number of descriptors.



**Figure 4.** Relationship between the predicted values of density of the new model and the experimental values.

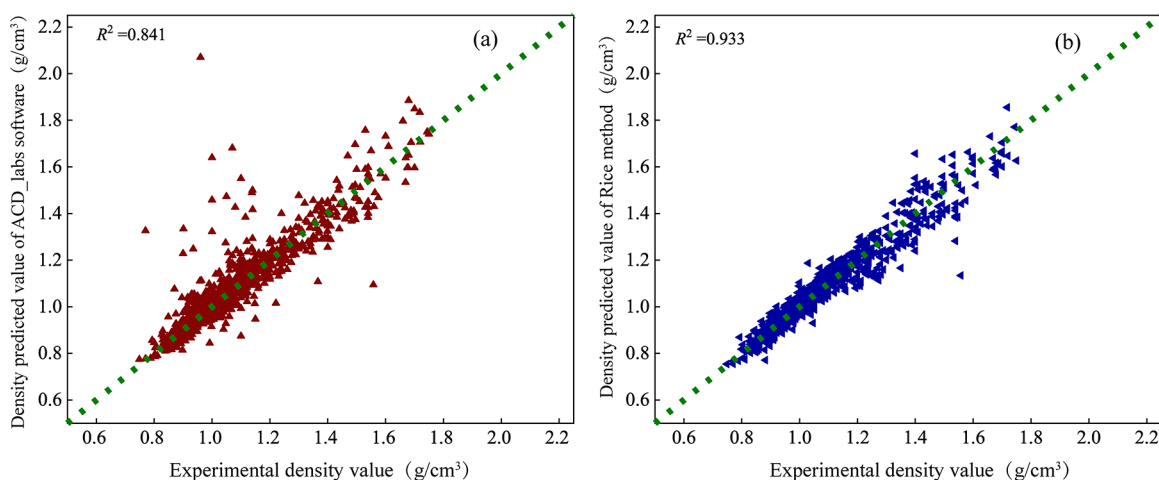
**Table 2. Evaluation Results of Density Prediction by ACD/Labs Software and Linear Equation**

other methods	$R^2$	RMSE (g/cm <sup>3</sup> )	MAE (g/cm <sup>3</sup> )
ACD/Labs software	0.841	0.071	0.036
linear equation	0.933	0.044	0.028

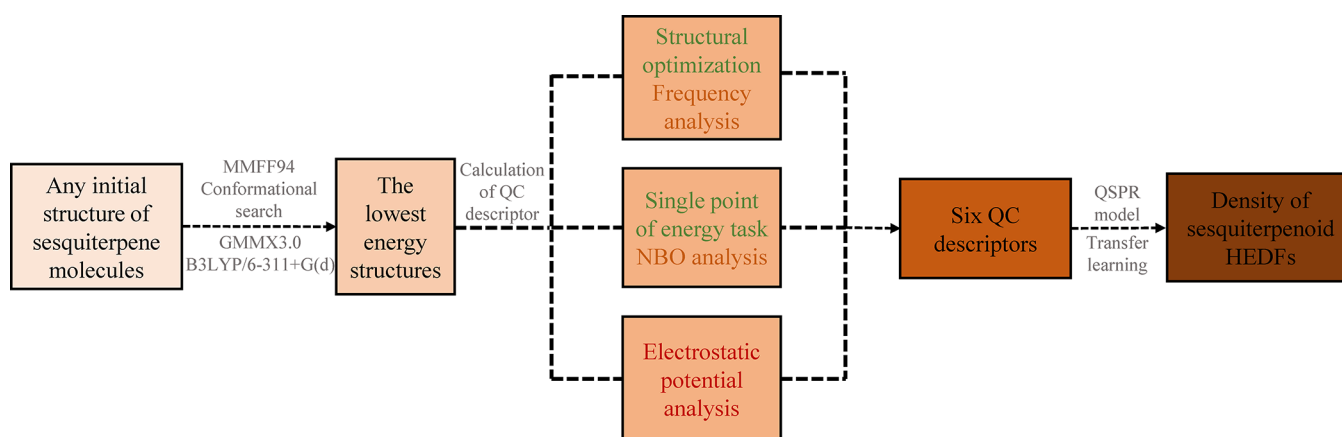
experimental values. The results indicated that the prediction performance of both mainstream methods was inferior to that of our QSPR model, substantiating the reliability of the QSPR model constructed using the XGBoost algorithm. This suggested that it may be more appropriate for predicting the density of sesquiterpenoid HEDFs.

Consequently, we used the new model, which incorporates six QC descriptors of sesquiterpenoid HEDFs, to predict the density of target molecules through a transfer learning approach, as demonstrated in Figure 6. To verify the practical applicability of the QSPR model, we applied it to predict sesquiterpenoid HEDFs with known experimental density values.<sup>16</sup> The results are listed in Table 3. The discrepancy between the experimental density values and the predicted values is small, and its performance is better than the other two prediction schemes. In conclusion, the QSPR model, with only 6 parameters, exhibited exceptional and consistent predictive performance, enabling swift and accurate density predictions for sesquiterpenoid HEDFs.

**3.3. Feature Analysis.** The interpretability challenge of the QSPR model was addressed by using the SHAP method,

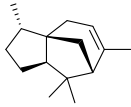
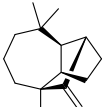
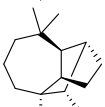


**Figure 5.** Comparison of the predicted density values of ACD/Labs software (a) and the linear method (b) with experimental values.



**Figure 6.** Density prediction flow of sesquiterpenoid HEDFs.

**Table 3.** Experimental and Predicted Values of the Density of Sesquiterpene Molecules such as  $\alpha$ -Cedrene, Longifolene, and Longifolane

Molecular structure	Name	Experimental density value (g/cm <sup>3</sup> )	Model (g/cm <sup>3</sup> )	ACD/labs software (g/cm <sup>3</sup> )	Linear method (g/cm <sup>3</sup> )
	$\alpha$ -Cedrene	0.935	<b>0.925</b>	0.950	0.920
	Longifolene	0.933	<b>0.928</b>	0.940	0.922
	Longifolane	0.927	<b>0.901</b>	0.900	0.897

elucidating the correlation between features and target properties. This approach lays a theoretical foundation for the development and design of sesquiterpenoid HEDFs. The input data consisted of the internal training set of the new model, ensuring accuracy in determining the contribution of each feature. Additionally, the same 1000 iterations of 5-fold

cross-validation were utilized for hyperparameter optimization, enhancing the model's robustness.

The overall importance of the features in our QSPR model was determined by calculating the absolute Shapley values for each feature with the results depicted in Figure 7. This analysis suggested that a small subset of features significantly impacted

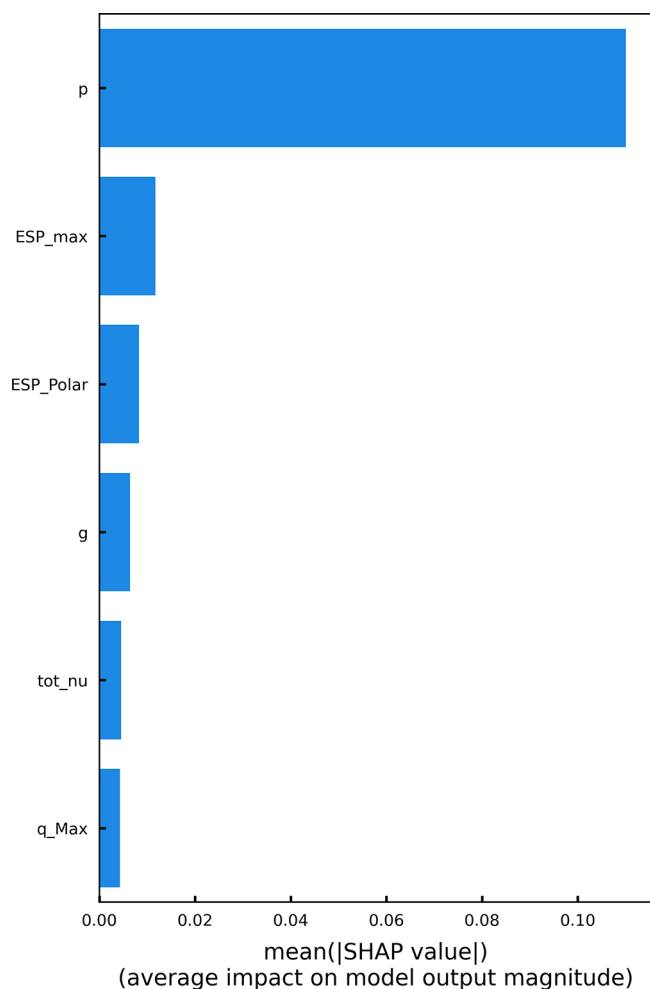


Figure 7. Overall importance of six QC descriptors.

the density of sesquiterpenoid HEDFs. We further analyzed the correlation between these 6 descriptors and density, as illustrated in Figure 8. In this figure, points with larger

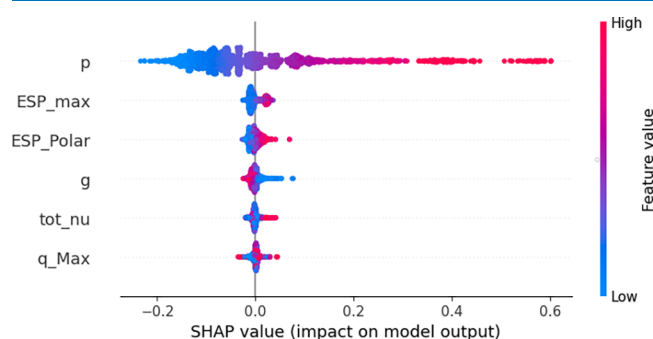


Figure 8. SHAP summary plot.

eigenvalues are colored red, indicating a strong impact, whereas blue points correspond to smaller eigenvalues, indicating a weaker impact. Among these, *p* (microscopic density) had the most profound impact on the QSPR model. This attribute reflects the degree of molecular packing within the cyclic compounds in space. The higher this value, the denser the spatial arrangement of the compound, thus increasing its macroscopic density. Other influential parameters derived from the electrostatic potential analysis at a 0.001 au electron density include

ESP\_max (maximum electrostatic potential), ESP\_Polar (polar surface area (IESP) > 10 kcal/mol), and tot\_nu (product of the overall variance of the electrostatic potential and the constant of charge balance). These parameters signify the strength of intermolecular interaction forces, which, as seen in Figure 8, positively impact density for higher values. Thus, stronger intermolecular forces lead to closer molecular packing, which, in turn, increases the density of the compound. Another more important feature is *g* (chemical hardness) of the molecule, which measures its chemical reactivity and stability. The analysis indicates that higher *g* values, which signify lower reactivity and greater molecular stability, are negatively correlated with the density. For the polycyclic molecules within our study, we postulate that compounds with lower *g* values, which contain more strained rings, lead to tighter molecular packing and thus larger density values. Finally, the SHAP values of q\_Max (the largest positive charge in the molecule) were smaller, indicating that this descriptor had a smaller impact on the density of the compound.

From the analysis, we have drawn several key conclusions: First, only a small number of QC descriptors are crucial for determining the density of compounds. Second, the combination of the XGBoost algorithm with the SHAP method provides a transparent understanding of how features interact with compound properties. Third, sesquiterpenoids that exhibit tight stacking and strong intermolecular interactions and contain a high number of strained rings are more likely to possess higher density.

**3.4. Web Server.** To enhance the practical utility of our density prediction model, we integrated it with a straightforward and efficient MLR equation for calculating the enthalpy of combustion. We developed a web server, SesquiterPre (<https://sespre.cmdrg.com/#/>), designed to predict the density, enthalpy of combustion, and energy density of sesquiterpenoid HEDFs. Notably, this server requires only 10 QC descriptors to compute these properties, as shown in the primary modules depicted in Figure 9.

The density prediction module of this server was created using a high-quality data set, mainly composed of the 1567 multiring structure density data sets and six QC descriptors highlighted earlier. The practical applications and evaluation parameters of the QSPR model demonstrate that the densities of sesquiterpenoid HEDFs can be accurately predicted. The enthalpy of

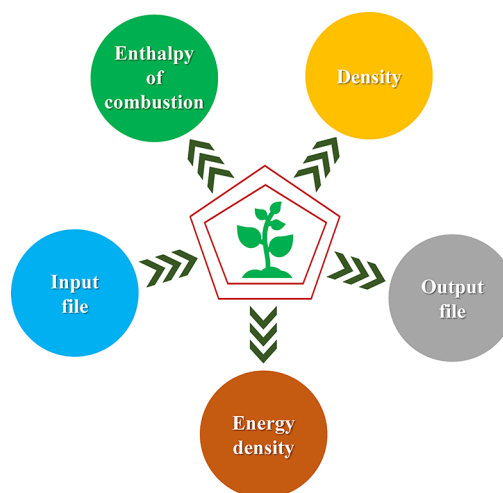


Figure 9. Main function modules of SesquiterPre.

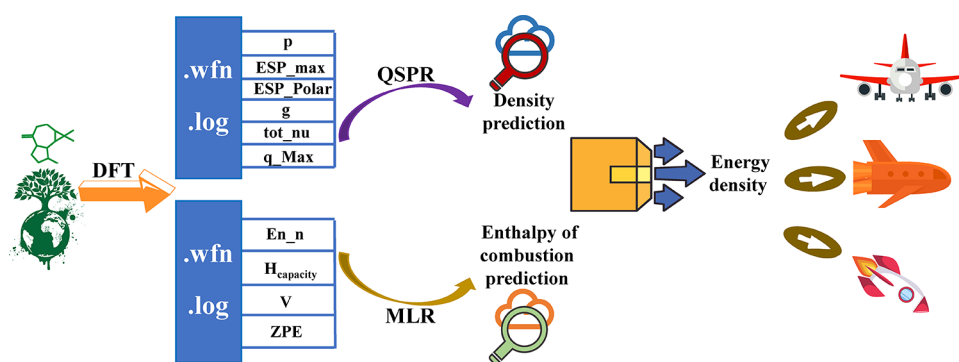


Figure 10. Three-property calculation process of SesquiterPre.

combustion for input molecules is directly predicted by using the established MLR equation. In this module,  $V$ ,  $ZPE$ ,  $H_{\text{capacity}}$ , and  $En\_n$  need to be provided to swiftly calculate the enthalpy of combustion. Energy density calculations are carried out using eq 3.

SesquiterPre was developed using Python and can handle multiple simultaneous accesses. To initiate a prediction, users are required to prepare five types of computational output files. These include the following: (i) the result files obtained from geometric structure optimization and frequency calculation, (ii) the results of single-point energy calculation for three different states, and (iii) quantitative molecular surface analysis. The necessary 10 QC descriptors are obtained from the above output files. Recognizing that these descriptors require unit conversion and computational extraction, this tool automates these tasks to simplify the process for users. Subsequently, SesquiterPre can predict all three properties after a clicking run. The detailed prediction flow is illustrated in Figure 10. This result table displays the SMILES, 3D structure, molecular mass, and property prediction results for the input molecules. Lastly, SesquiterPre provides a user-friendly layout for the results' table, which can be downloaded as a .CSV file.

Looking ahead, in addition to the existing data set comprising 1567 polycyclic compounds and 292 sesquiterpenoid HEDFs, we plan to continue gathering relevant data. This new piece of data will be incorporated into the model construction data set to continuously enhance the predictive performance of this server.

#### 4. CONCLUSIONS

Owing to the absence of density values for target molecules, we utilized a data set of polycyclic compounds with structures akin to those of sesquiterpenoid molecules and known experimental density values as substitutes. We then constructed a model to predict the density of sesquiterpenoid HEDFs using the XGBoost algorithm, a density data set, and 54 QC descriptors. The high-performance QSPR model confirmed a strong correlation between the predicted target and QC descriptors. To decrease the complexity of the model, we employed the RFE method for feature screening. This led to the development of a new density prediction model containing only six QC features using the same algorithm. The coefficients of determination  $R^2$  for the internal training set and external test set were approximately 0.967 and 0.971, respectively. The RMSE values were as low as 0.031 and 0.027 g/cm<sup>3</sup>, while the mean absolute errors (MAE) were 0.020 and 0.019 g/cm<sup>3</sup>, respectively. Furthermore, compared to the two mainstream methods, the calculation results further illustrated the exceptional predictive performance and practical application of the constructed QSPR

model. Subsequently, SHAP analysis led us to speculate that sesquiterpenoid HEDFs with a high degree of stacking, robust intermolecular interactions, and numerous tensegrity rings would exhibit a higher density. Finally, we developed a web server, SesquiterPre (<https://sespre.cmdrg.com/#/>), using an accurate density prediction model and the MLR equation for the enthalpy of combustion combined with 10 QC descriptors. This server can swiftly and accurately predict the density, enthalpy of combustion, and energy density of sesquiterpenoid HEDFs. This study will aid in the rapid screening of high-performance HEDFs, thereby improving the processes of fuel development and design.

#### ■ ASSOCIATED CONTENT

##### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.4c01898>.

The Supporting Information is available free of charge, contains S1–S4 (.csv files)

Experimental density values and descriptors of 1567 compounds (XLSX)

Six QC descriptors for density prediction of sesquiterpenoids (XLSX)

Four QC descriptors for calculation of enthalpy of combustion of sesquiterpenoids (XLSX)

Enthalpy of combustion and density of sesquiterpenoids (XLSX)

#### ■ AUTHOR INFORMATION

##### Corresponding Authors

Yu-Quan Yuan – School of Physics and Electronics Engineering, Sichuan University of Science & Engineering, Zigong 643000, China; Email: [yuquan\\_yuan@suse.edu.cn](mailto:yuquan_yuan@suse.edu.cn)

Liang-Liang Wang – State Key Laboratory of NBC Protection for Civilian, Beijing 102205, China; [orcid.org/0000-0002-0551-3409](https://orcid.org/0000-0002-0551-3409); Email: [wangliangliang@pku.org.cn](mailto:wangliangliang@pku.org.cn)

Jun-Jie Ding – State Key Laboratory of NBC Protection for Civilian, Beijing 102205, China; Email: [djj224@163.com](mailto:djj224@163.com)

##### Authors

Hang Yang – State Key Laboratory of NBC Protection for Civilian, Beijing 102205, China; School of Physics and Electronics Engineering, Sichuan University of Science & Engineering, Zigong 643000, China

Zhi-Jiang Yang – State Key Laboratory of NBC Protection for Civilian, Beijing 102205, China; [orcid.org/0000-0003-1884-8782](https://orcid.org/0000-0003-1884-8782)

**Teng-Xin Huang** – School of Physics and Electronics Engineering, Sichuan University of Science & Engineering, Zigong 643000, China  
**Li Pan** – State Key Laboratory of NBC Protection for Civilian, Beijing 102205, China  
**Xin-Miao Wei** – State Key Laboratory of NBC Protection for Civilian, Beijing 102205, China  
**Yan-Fei Hu** – Department of Applied Physics, Chengdu University of Technology, Chengdu 610059, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.4c01898>

### Author Contributions

<sup>#</sup>H.Y. and Z.-J.Y. contributed equally to this work. H.Y. performed conceptualization, methodologies, and writing of the original draft; Z.-J.Y. performed data curation and software analysis; T.-X.H. performed resources acquisition and supervision; L.P. performed resources acquisition and investigation; X.-M.W. performed visualization and investigation; Y.-F.H. performed supervision and review and editing of the manuscript; Y.-Q.Y. performed supervision and methodologies; L.-L.W. performed review and editing of the manuscript and software analysis; J.-J.D. performed review and editing of the manuscript, resources acquisition, and supervision.

### Notes

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

The authors are grateful to the Innovation Fund of Postgraduate Sichuan University of Science & Engineering (grant nos. y2020073 and y2021008), the Opening Project of Sichuan Province University Key Laboratory of Bridge Nondestructive Detecting and Engineering Computing (grant no. 2020QYJ02), and the Innovation and Entrepreneurship Training Program of Sichuan Province (grant nos. S202010622080 and S202010622082). This work was supported by the High-Performance Computing Center of Sichuan University of Science & Technology.

### REFERENCES

- (1) Kim, J.; Han, J.; Kwon, T. S.; Park, Y.-K.; Jeon, J.-K. Oligomerization and isomerization of dicyclopentadiene over mesoporous materials produced from zeolite beta. *Cataly Today* **2014**, *232*, 69–74.
- (2) Liang, R.; Harvey, B. G.; Quintana, R. L.; Sufflita, J. M. Assessing the Biological Stability of a Terpene-Based Advanced Biofuel and Its Relationship to the Corrosion of Carbon Steel. *Energ Fuel* **2015**, *29*, 5164–70.
- (3) Arias-Ugarte, R.; Wekesa, F. S.; Schunemann, S.; Findlater, M. Iron(III)-Catalyzed Dimerization of Cycloolefins: Synthesis of High-Density Fuel Candidates. *Energ Fuel* **2015**, *29*, 8162–7.
- (4) Sun, D.; Cai, W.; Li, C.; Lu, J. Experimental study on atomization characteristics of high-energy-density fuels using a fuel slinger. *Energy* **2021**, *234*, No. 121222.
- (5) Zarezin, D. P.; Rudakova, M. A.; Shorunov, S. V.; Sultanova, M. U.; Samoilov, V. O.; Maximov, A. L.; et al. Design and preparation of liquid polycyclic norbornanes as potential high performance fuels for aerospace propulsion. *Fuel Process. Technol.* **2022**, *225*, No. 107056.
- (6) Sarría, S.; Wong, B.; Martín, H. G.; Keasling, J. D.; Peralta-Yahya, P. Microbial synthesis of pinene. *ACS Synth. Biol.* **2014**, *3*, 466–75.
- (7) Harvey, B. G.; Wright, M. E.; Quintana, R. L. High-Density Renewable Fuels Based on the Selective Dimerization of Pinenes. *Energ Fuel* **2010**, *24*, 267–73.
- (8) Sutton, G. P. *History of liquid propellant rocket engines*. AIAA; 2006.
- (9) Sutton, G. P.; Biblarz, O. *Rocket propulsion elements*. John Wiley & Sons; 2016.
- (10) Zou, J.-J.; Zhang, X.; Pan, L. *High-energy-density fuels for advanced propulsion: Design and synthesis*. John Wiley & Sons; 2020.
- (11) Morris, D. M.; Quintana, R. L.; Harvey, B. G. High-Performance Jet Fuels Derived from Bio-Based Alkenes by Iron-Catalyzed [2 + 2] Cycloaddition. *ChemSusChem* **2019**, *12*, 1646–52.
- (12) Ma, S.; Chen, Y.; Liu, X.; Pan, L.; Zhang, X.; Zou, J.-J. Synthesis of caged high-energy-density fuel as potential high-performance energetic additive for liquid aerospace fuel. *Fuel Process. Technol.* **2022**, *229*, No. 107179.
- (13) Pietruszka, J. Synthesis and properties of oligocyclopropyl-containing natural products and model compounds. *Chem. Rev.* **2003**, *103*, 1051–70.
- (14) Barton, L. M.; Chen, L.; Blackmond, D. G.; Baran, P. S. Electrochemical borylation of carboxylic acids. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, No. e2109408118.
- (15) Meylemans, H. A.; Quintana, R. L.; Harvey, B. G. Efficient conversion of pure and mixed terpene feedstocks to high density fuels. *Fuel* **2012**, *97*, 560–8.
- (16) Wang, W.; Liu, Y.; Shi, C.; Pan, L.; Zhang, X.; Zou, J.-J. High energy density renewable fuels based on multicyclic sesquiterpene: Synthesis and performance. *Fuel* **2022**, *318*, No. 123665.
- (17) Liu, Y.; Shi, C.; Pan, L.; Zhang, X.; Zou, J.-J. Synthesis and performance of cyclopropanated pinanes with high density and high specific impulse. *Fuel* **2022**, *307*, No. 121906.
- (18) Harvey, B. G.; Meylemans, H. A.; Gough, R. V.; Quintana, R. L.; Garrison, M. D.; Bruno, T. J. High-density biosynthetic fuels: the intersection of heterogeneous catalysis and metabolic engineering. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9448–57.
- (19) Liu, C. L.; Tian, T.; Alonso-Gutierrez, J.; Garabedian, B.; Wang, S.; Baidoo, E. E. K.; et al. Renewable production of high density jet fuel precursor sesquiterpenes from *Escherichia coli*. *Biotechnol Biofuels* **2018**, *11*, 285.
- (20) Keshavarz, M. H. New method for calculating densities of nitroaromatic explosive compounds. *J. Hazard Mater.* **2007**, *145*, 263–9.
- (21) Keshavarz, M. H. Prediction of densities of acyclic and cyclic nitramines, nitrate esters and nitroaliphatic compounds for evaluation of their detonation performance. *J. Hazard Mater.* **2007**, *143*, 437–42.
- (22) Keshavarz, M. H. Novel method for predicting densities of polynitro arene and polynitro heteroarene explosives in order to evaluate their detonation performance. *J. Hazard Mater.* **2009**, *165*, 579–88.
- (23) Keshavarz, M. H.; Pouretedal, H. R. A reliable simple method to estimate density of nitroaliphatics, nitrate esters and nitramines. *J. Hazard Mater.* **2009**, *169*, 158–69.
- (24) Keshavarz, M. H.; Monjezi, K. H.; Esmailpour, K.; Zamani, M. Performance Assessment of Some Isomers of Saturated Polycyclic Hydrocarbons for Use as Jet Fuels. *Propellants, Explosives, Pyrotechnics* **2015**, *40*, 309–14.
- (25) Keshavarz, M. H.; Soury, H.; Motamedoshariati, H.; Dashtizadeh, A. Improved method for prediction of density of energetic compounds using their molecular structure. *Structural Chemistry* **2015**, *26*, 455–66.
- (26) Keshavarz, M. H.; Ebadpour, R.; Jafari, M. Reliable prediction of crystal density of high nitrogen-containing organic compounds as powerful, less sensitive, eco-friendly energetic materials for dependable assessment of their performance. *Fluid Phase Equilib.* **2023**, *565*, No. 113653.
- (27) Keshavarz, M. H.; Ebadpour, R.; Jafari, M. An easy pathway for reliable assessment of crystal density of high nitrogen-containing salts and ionic liquids without using complex descriptors. *Fluid Phase Equilib.* **2023**, *565*, No. 113667.
- (28) Soave, G. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chem. Eng. Sci.* **1972**, *27*, 1197–203.
- (29) Peng, D.-Y.; Robinson, D. B. A New Two-Constant Equation of State. *Ind. Eng. Chem. Fundamen* **1976**, *15*, 59–64.



- (30) Riazi, M.-R.; Mansoori, G. A. Simple equation of state accurately predicts hydrocarbon densities. *Oil Gas J.* **1993**, *91*, 108–11.
- (31) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. Properties of atoms in molecules: atomic volumes. *J. Am. Chem. Soc.* **1987**, *109*, 7968–79.
- (32) Murray, J. S.; Politzer, P. Statistical analysis of the molecular surface electrostatic potential: an approach to describing noncovalent interactions in condensed phases. *J. Mol. Struct.-THEOCHEM* **1998**, *425*, 107–14.
- (33) Qiu, L.; Xiao, H.; Gong, X.; Ju, X.; Zhu, W. Crystal density predictions for nitramines based on quantum chemistry. *J. Hazard Mater.* **2007**, *141*, 280–8.
- (34) Rice, B. M.; Hare, J. J.; Byrd, E. F. Accurate predictions of crystal densities using quantum mechanical molecular volumes. *J. Phys. Chem. A* **2007**, *111*, 10874–9.
- (35) Politzer, P.; Martinez, J.; Murray, J. S.; Concha, M. C.; Toro-Labbé, A. An electrostatic interaction correction for improved crystal density prediction. *Mol. Phys.* **2009**, *107*, 2095–101.
- (36) Wang, J.; Hou, T. Application of Molecular Dynamics Simulations in Molecular Property Prediction I: Density and Heat of Vaporization. *J. Chem. Theory Comput* **2011**, *7*, 2151–65.
- (37) Wang, S.; Cui, Z.; Yu, C.; Tan, T. Computational Assessment of the Molecular Structure and Properties for High Energy Density Fuel. *J. Phys. Chem. A* **2020**, *124*, 6660–6.
- (38) Ihmels, E. C.; Gmehling, J. Extension and Revision of the Group Contribution Method GCVOL for the Prediction of Pure Compound Liquid Densities. *Ind. Eng. Chem. Res.* **2003**, *42*, 408–12.
- (39) Padaszyński, K.; Domańska, U. A New Group Contribution Method For Prediction of Density of Pure Ionic Liquids over a Wide Range of Temperature and Pressure. *Ind. Eng. Chem. Res.* **2012**, *51*, 591–604.
- (40) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Ferrando, N.; Creton, B. Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods. *Energ Fuel* **2012**, *26*, 2416–26.
- (41) Haghbaksh, R.; Adib, H.; Keshavarz, P.; Koolivand, M.; Keshtkari, S. Development of an artificial neural network model for the prediction of hydrocarbon density at high-pressure, high-temperature conditions. *Thermochim. Acta* **2013**, *551*, 124–30.
- (42) El-Harbawi, M.; Samir, B. B.; Babaa, M.-R.; Mutalib, M. I. A. A New QSPR Model for Predicting the Densities of Ionic Liquids. *Arab J. Sci. Eng.* **2014**, *39*, 6767–75.
- (43) Fathollahi, M.; Sajady, H. Prediction of density of energetic cocrystals based on QSPR modeling using artificial neural network. *Struct Chem.* **2018**, *29*, 1119–28.
- (44) Keshavarz, M. H.; Makvandi, L. Assessment of Recent Researches for Reliable Prediction of Density of Organic Compounds as well as Ionic Liquids and Salts Containing Energetic Groups at Room Temperature. *Propell Explos Pyrot* **2020**, *45*, 1680–90.
- (45) Yang, C.; Chen, J.; Wang, R.; Zhang, M.; Zhang, C.; Liu, J. Density Prediction Models for Energetic Compounds Merely Using Molecular Topology. *J. Chem. Inf Model* **2021**, *61*, 2582–93.
- (46) Harvey, B. G.; Harrison, K. W.; Davis, M. C.; Chafin, A. P.; Baca, J.; Merriman, W. W. Molecular Design and Characterization of High-Cetane Alkyl Diamondoid Fuels. *Energ Fuel* **2016**, *30*, 10171–8.
- (47) Harrison, K. W.; Rosenkoetter, K. E.; Harvey, B. G. High Density Alkyl Diamondoid Fuels Synthesized by Catalytic Cracking of Alkanes in the Presence of Adamantane. *Energ Fuel* **2018**, *32*, 7786–91.
- (48) Dennington, R.; Keith, T. A.; Millam, J. M. *GaussView 6.0*. 16. Semichem Inc.: Shawnee Mission, KS, 2016.
- (49) Liu, S.-B. Conceptual Density Functional Theory and Some Recent Developments. *Acta Phys.-Chim. Sin.* **2009**, *25*, 590–600.
- (50) Chattaraj, P. K.; Giri, S.; Duley, S. Update 2 of: Electrophilicity Index. *Chem. Rev.* **2011**, *111*, PR43–PR75.
- (51) Wang, L.-L.; Ding, J.-J.; Pan, L.; Fu, L.; Tian, J.-H.; Cao, D.-S.; et al. Quantitative structure-toxicity relationship model for acute toxicity of organophosphates via multiple administration routes in rats and mice. *J. Hazard Mater.* **2021**, *401*, No. 123724.
- (52) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R. *Gaussian 16 Rev. C.01*. Gaussian Inc.: Wallingford, CT, 2016.
- (53) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–4.
- (54) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–52.
- (55) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–96.
- (56) Pahima, E.; Hoz, S.; Ben-Tzion, M.; Major, D. T. Computational design of biofuels from terpenes and terpenoids. *Sustain Energy Fuels* **2019**, *3*, 457–66.
- (57) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899–926.
- (58) Lu, T.; Chen, F.-W. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580–92.
- (59) Cao, D.-S.; Xu, Q.-S.; Liang, Y.-Z.; Zhang, L.-X.; Li, H.-D. The boosting: A new idea of building models. *Chemom Intell Lab Syst* **2010**, *100*, 1–11.
- (60) Volkovs, M.; Yu, G. W.; Poutanen, T. Content-based Neighbor Models for Cold Start in Recommender Systems. In *Proceedings of the Recommender Systems Challenge 2017*. Association for Computing Machinery: Como, Italy; 2017, Article 7.
- (61) Adam-Bourdarios, C.; Cowan, G.; Germain, C.; Guyon, I.; Kégl, B.; Rousseau, D. The Higgs boson machine learning challenge. Glen, C.; Germain, C.; Guyon, I.; Kégl, B.; Rousseau, D.; eds. In *Proceedings of the NIPS 2014 Workshop on High-energy Physics and Machine Learning*. Proceedings of Machine Learning Research: PMLR; 42 2015, 19–55.
- (62) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery: San Francisco, California, USA; 2016: 785–794.
- (63) Fu, L.; Liu, L.; Yang, Z.-J.; Li, P.; Ding, J.-J.; Yun, Y.-H.; et al. Systematic Modeling of  $\log D_{7,4}$  Based on Ensemble Machine Learning, Group Contribution, and Matched Molecular Pair Analysis. *J. Chem. Inf Model* **2020**, *60*, 63–76.
- (64) Baxter, J. Theoretical Models of Learning to Learn. Thrun, S.; Pratt, L. editors. In *Learning to Learn*. Springer US: Boston, MA; 1998; p 71–94.
- (65) Lundberg, S. M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 47668.
- (66) Yang, H.; Yang, Z.-J.; Yang, Q.-F.; Wei, X.-M.; Yuan, Y.-Q.; Wang, L.-L.; et al. Simple and high-precision DFT-QSPR prediction of enthalpy of combustion for sesquiterpenoid high-energy-density fuels. *Fuel* **2023**, *332*, No. 126157.