Journal of Energy Chemistry 102 (2025) 161–171

Contents lists available at ScienceDirect

Journal of Energy Chemistry

journal homepage: www.elsevier.com/locate/jechem

http://www.journals.elsevier.com/journal-of-energy-chemistry/

Machine learning approaches for predicting impact sensitivity and detonation performances of energetic materials

Wei-Hong Liu [a], Qi-Jun Liu [a,*], Fu-Sheng Liu [a], Zheng-Tang Liu [b]

[a] Bond and Band Engineering Group, School of Physical Science and Technology, Southwest Jiaotong University, Chengdu 610031, Sichuan, China
[b] State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China

ARTICLE INFO

Article history:
Received 10 August 2024
Revised 21 October 2024
Accepted 21 October 2024
Available online 05 November 2024

Keywords:
Energetic materials
Machine learning
Impact sensitivity
Detonation performances
Feature descriptors
Balancing strategy

ABSTRACT

Excellent detonation performances and low sensitivity are prerequisites for the deployment of energetic materials. Exploring the underlying factors that affect impact sensitivity and detonation performances as well as exploring how to obtain materials with desired properties remains a long-term challenge. Machine learning with its ability to solve complex tasks and perform robust data processing can reveal the relationship between performance and descriptive indicators, potentially accelerating the development process of energetic materials. In this background, impact sensitivity, detonation performances, and 28 physicochemical parameters for 222 energetic materials from density functional theory calculations and published literature were sorted out. Four machine learning algorithms were employed to predict various properties of energetic materials, including impact sensitivity, detonation velocity, detonation pressure, and Gurney energy. Analysis of Pearson coefficients and feature importance showed that the heat of explosion, oxygen balance, decomposition products, and HOMO energy levels have a strong correlation with the impact sensitivity of energetic materials. Oxygen balance, decomposition products, and density have a strong correlation with detonation performances. Utilizing impact sensitivity of 2,3,4-trinitrotoluene and the detonation performances of 2,4,6-trinitrobenzene-1,3,5-triamine as the benchmark, the analysis of feature importance rankings and statistical data revealed the optimal range of key features balancing impact sensitivity and detonation performances: oxygen balance values should be between −40% and −30%, density should range from 1.66 to 1.72 g/cm^3, HOMO energy levels should be between −6.34 and −6.31 eV, and lipophilicity should be between −1.0 and 0.1, 4.49 and 5.59. These findings not only offer important insights into the impact sensitivity and detonation performances of energetic materials, but also provide a theoretical guidance paradigm for the design and development of new energetic materials with optimal detonation performances and reduced sensitivity.

1. Introduction

Energetic materials, which are substances imbued with substantial chemical energy capable of being released through suitable external stimuli, play an integral role in a variety of industrial and defense applications [1]. In the industrial sector, these materials are widely used for mining and blasting operations; in the realm of defense, they are essential components in the fabrication of various ammunitions, rocket propellants, and explosive devices [2,3]. Detonation performances and sensitivity are the two paramount characteristics of energetic materials. However, a crucial consideration in the design and application of these materials is the trade-off between their high detonation performances and insensitivity to external stimuli, such as heat, impact, and friction. This is because materials with high detonation performances tend to be more sensitive to external stimuli [4,5].

Developing energetic materials that possess both high explosive power and insensitivity is important yet challenging. In the background of increasing global competition in the 21st century, substantial resources are being invested in research efforts, with nations accelerating their pace to discover and deploy high-performance energetic material systems [6,7]. One critical goal of these efforts is the quest for substitutes to replace a handful of predominant energetic materials, namely HMX (1,3,5,7-tetranitro-1,3,5,7-tetrazocan), RDX (cyclotrimethylenetrinitramine), TNT

* Corresponding author.
E-mail address: qijunliu@home.swjtu.edu.cn (Q.-J. Liu).

https://doi.org/10.1016/j.jechem.2024.10.035

(2,4,6-trinitrotoluene), PETN (pentaerithrityl tetranitrate), and TATB (1,3,5-Triamino-2,4,6-trinitrobenzene), which have been the mainstay in military arsenals worldwide since the Second World War [8–10]. While there has been a wealth of research and synthesis of new energetic materials, resulting in a range of materials with superior performance, the examples that have transitioned into industrial production are few and far between. The development of 2,4,6,8,10,12-hexanitro-2,4,6,8,10,12-hexaazatetracyclo (CL-20) stands as an exception, with its research and development process spanning approximately 15 years from initial synthesis to industrial production, followed by another 15 years to commercialization [11,12]. This transition period from laboratory research to industrial application, typically ranging from 20 to 40 years, is a common phenomenon in the field of material science. The development of most energetic materials is based on a costly trial-and-error approach, involving repetitive synthesis and testing processes. This approach is not only limited in terms of time efficiency, but also consumes extensive resources and poses potential safety risks [13,14].

Researchers have established numerous predictive parameters by correlating H_{50}, which represents the impact sensitivity of a material and is defined as the height at which a 50% probability of detonation occurs under standard conditions, with various properties such as molecular composition [15], voids and free space within crystals [16], oxygen balance [17], nitro group charges [18], molecular electrostatic potential [19], bond dissociation energy [20,21], and vibrational modes [22,23]. However, these parameters often focus on specific types of molecules and struggle to provide a deep understanding of the physical mechanisms influencing the H_{50} of energetic compounds. In recent years, predictive models have emerged based on specific physical mechanisms, linking H_{50} with sensitivity factors obtained from first-principles calculations. This approach is primarily driven by two theories. One is based on the phonon up-pumping model [24,25], which attributes sensitivity differences to varying rates of energy transfer prior to reaction, and the other considers the decomposition rate (k) after the chemical reaction starts as the key factor affecting sensitivity [26,27]. However, both methods entail high computational costs and have limited predictive capabilities for new or modified materials. Predicting the sensitivity of energetic materials through theoretical and simulation approaches has always been challenging, failing to provide consistent guidance for the design process of energetic materials [28]. Therefore, finding a safer and more efficient method for evaluating the performance of energetic materials has become a focal point of research.

In the last ten years, machine learning (ML) has emerged as a critical component in the toolkit for materials research, with extensive applications in microbiology, medical diagnostics, and energy transition [29,30]. The adoption of data-driven methodologies has been proven to be effective in forecasting the stability of new materials [31–33], identifying material characteristics [34–36], improving the precision and efficiency of computations [37,38], and predicting various properties and rational design of materials. The main goal of ML is to learn patterns directly from existing data and apply them to predict new data. This approach differs fundamentally from other methods used to simulate system behavior, like quantum mechanical calculations, in which machine learning does not use predefined rules for making predictions. They are mathematical functions based on generic parameters, which are determined through optimization using the loss function [39]. Within the various branches of materials science, the application of ML in the field of energetic materials is not as widespread as in other areas. This is partly due to the inherent dangers of energetic materials, which impede experimental work and result in the limited data availability. Additionally, the field of energetic materials is often encumbered with strict safety protocols and privacy issues, which constrain the collection and dissemination of data.

In this study, four machine learning models are applied to predict the impact sensitivity and detonation performance of energetic materials. Experimental data for various energetic materials are collected for this purpose. Using the impact sensitivity of 2,3,4-trinitrotoluene and the detonation performance of 2,4,6-trinitrobenzene-1,3,5-triamine as benchmarks, an analysis of feature importance rankings and statistical data reveals the optimal range of key features that balance impact sensitivity and detonation performance. The input variables for the machine learning models include oxygen balance, heat of explosion, decomposition products, density, and various functional groups. Impact sensitivity ($\ln(H_{50})$) and detonation performance (including detonation velocity, detonation pressure, and Gurney energy) are considered the output variables.

2. Methods and procedure

2.1. Data collection and processing

So far, the literature has reported experimental data on the impact sensitivity and detonation performances using the same procedure for several hundred energetic compounds [40,41]. The validation metrics produced by these data-driven models enhance a sense of confidence in researchers' ability to predict important attributes of energetic materials. The H_{50} data was derived from the ERL Type 12 procedure with a 2.5 kg impact hammer [41,42], and the values for D, P, and E were obtained through the Kamlet-Jacobs (KJ) equation [42].

$$D(\mathrm{km/s}) = 1.01\left[N^{0.5}M_{\mathrm{avg}}^{0.25}Q^{0.25}(1+1.30\rho)\right] \quad (1)$$

$$P(\mathrm{kbar}) = 15.58\left[NM_{\mathrm{avg}}^{0.25}Q^{0.5}\rho^{2}\right] \quad (2)$$

In these equations, N, M, Q, and ρ represent the number of moles of gaseous detonation products per gram of energetic material, the average molecular weight of the gaseous detonation products, the total heat release of the detonation reaction, and the density of the material. The calculation of the KJ equation adopts the H_2O-CO_2 equilibrium assumption. This assumption indicates that in the explosive decomposition reaction, oxygen first reacts with hydrogen atoms to form water (H_2O). Once all the hydrogen atoms have been converted into water, the remaining oxygen will combine with carbon atoms to produce carbon dioxide (CO_2). This assumption determines the priority of oxygen in the chemical reaction, the main types, and quantities of gases produced, thereby affecting the calculated values of N, M, and Q.

In the process of modeling, standardizing the collected dataset is crucial to maintain the accuracy of the machine learning model. The standardization procedure, as depicted in Eq. (1), proportionally scales the data to ensure that the raw data fits within a specified range. Once standardized, the data is adjusted to the interval of [−1, 1], optimizing it for training and analysis in machine learning applications. In every machine learning process used for predicting $\ln(H_{50})$, D, P, and E, the datasets were divided into a training set (80%) and a test set (20%).

$$Z = \frac{X_i - \mu}{\sigma} \quad (3)$$

In this formula, X_i refers to the value of a data point in the dataset, with μ being the average of the data and σ being the standard deviation. The modeling process is showed in Fig. 1.

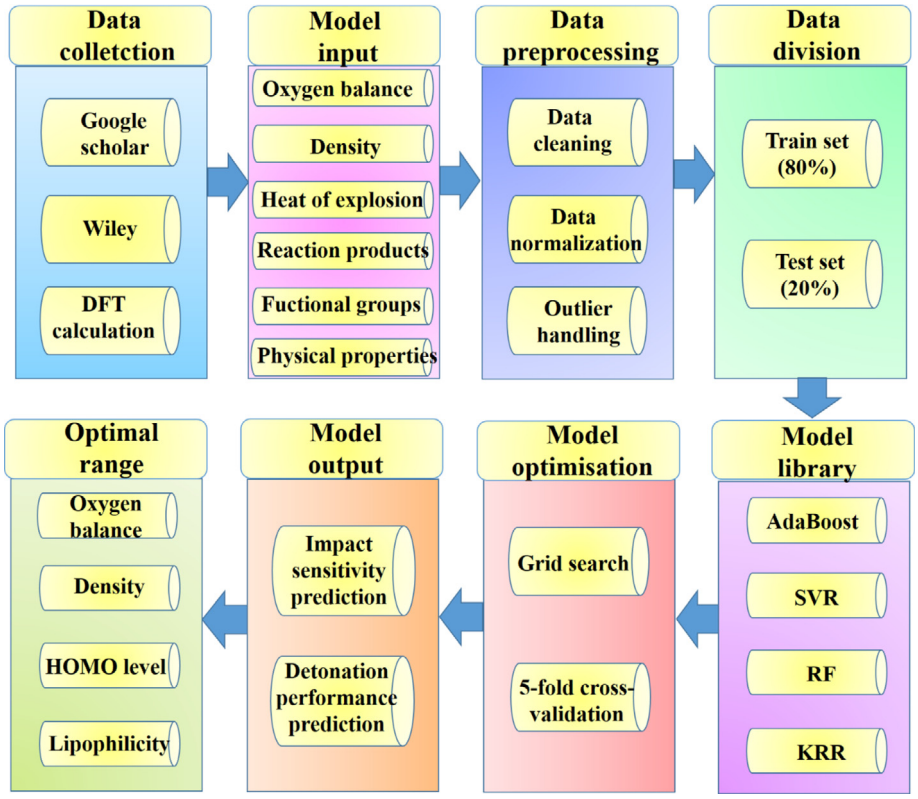


Fig. 1. Machine learning workflow diagram.

2.2. Feature composition

It is significant to acknowledge that appropriate feature descriptors are the keys to successfully using machine learning tools for the prediction of impact sensitivity and detonation performances. Through searches in databases like Wiley and Google Scholar, parameters affecting the impact sensitivity and detonation performance of energetic materials have been identified, such as oxygen balance, density, heat of explosion, reaction products, various functional groups, as well as electronic and vibrational properties. These parameters are treated as input variables and included in the modeling process.

Oxygen balance (OB) measures whether there is enough oxygen in a chemical compound to turn all carbon into CO_2 and all hydrogen into water when burning. The formula for determining the oxygen balance for a compound, like $C_aH_bN_cO_d$, is given by the following equation [17].

$$\mathrm{OB} = \frac{\left[d - (2a) - \left(\frac{b}{2}\right)\right] \times 1600}{M_W} \tag{4}$$

where a, b, d, and M_W is the number of carbon atoms, the number of hydrogen atoms, the number of oxygen atoms in the molecular formula and the molecular weight of the compound.

The crystal density (ρ) is estimated using a physically based additive model, with only 30 parameters predicting the densities of 42,880 crystals composed of elements including carbon (C), hydrogen (H), nitrogen (N), oxygen (O), fluorine (F), and chlorine (Cl), achieving an average relative error of merely 2% [43]. In this work, the values of density are from this calculation method [42].

Heat of explosion is chemical energy of the detonation reaction ($-\Delta H_0$ per gram), which can be estimated from the H_2O-CO_2 arbitrary decomposition assumption. The only necessary data are the explosive elemental composition, its formation heat, and its loading density [44]. In this work, the data for the heat of explosion is derived from the results calculated by Mathieu [42].

Based on the established oxidation hierarchy in explosive reactions as cited in the literature, the reaction products were calculated [45]. Initially, oxygen converts all hydrogen into water vapor (H_2O). Next, it turns carbon into carbon monoxide (CO), further into carbon dioxide (CO_2). Any leftover oxygen is released as oxygen gas (O_2). If there is not enough oxygen, any extra hydrogen and carbon come out as hydrogen gas (H_2) and solid carbon (C). All nitrogen atoms are converted to nitrogen gas (N_2). In this work, the ratio of detonation products, including gaseous CO_2, H_2O, N_2, and solid C, has been selected as feature descriptors.

RDKit is a widely-used open-source toolkit in the field of cheminformatics, offering capabilities for molecular manipulation and the computation of various molecular descriptors from given molecules. By employing RDKit and analyzing SMILES strings, the number of nitro groups ($-NO_2$), the number of amines (including primary amines, secondary amines, tertiary amines, and total amines), the number of hydroxyl groups (–OH), the number of methyl groups ($-CH_3$), aromaticity, lipophilicity, the percentage of nitrogen, oxygen, carbon and hydrogen atoms, and the number of hydrogen-bond donors and acceptors are obtained. These descriptors are then utilized as feature descriptors in machine learning. For each of the 222 energetic materials studied, the values of ln(H_{50}), D, P, and E, the compound name, molecular formula, SMILES string, calculated parameters, and other descriptors are provided in Table S1 (Figs. S1–S7 show the relationship between the ln(H_{50}), D, P, E and 28 feature descriptors of the energetic materials.).

2.3. Machine learning model

Multiple linear regression (MLR), artificial neural networks (ANN), kernel ridge regression (KRR), support vector regression

(SVR), random forest (RF), *k*-nearest neighbors (KNN), decision trees (DT), and Gaussian process regression (GPR) were developed and utilized in the field of ML for energetic materials [46,47]. Selecting an appropriate model along with its optimized hyperparameters is a critical prerequisite for the successful application of ML in the study of molecules and materials. In this work, four typical ML methods including adaptive boosting (AdaBoost), SVR, RF, and KRR were explored. The features of the four models are summarized in Table 1 [29,34]. The grid search techniques for hyperparameter optimization of these models, coupled with 5-fold cross-validation, are utilized to assess the error metrics of each. Throughout the training process, the grid search and cross-validation loops were executed over 4000 times in total, resulting in the identification of 16 sets of optimal hyperparameters for the four models in predicting ln(H_{50}), *D*, *P*, and *E*. All training procedures were conducted in an Anaconda virtual environment using Scikit-learn code. To ascertain the robustness of the identified hyperparameters and to evaluate the performance of the four trained models, three distinct metrics were employed for estimating the prediction error. These are the root mean square error (RMSE), Pearson correlation coefficient (*r*), and the coefficient of determination (R^2), each defined by the following expressions.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{5}$$

$$r = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{7}$$

In this analysis, y_i denotes the actual values within the dataset, and $\hat{y}_i$ represents the values predicted by our machine learning models. $\bar{y}$ and $\bar{\hat{y}}$ stand for the mean value of the dataset and prediction, and *n* indicates the number of samples. The current study aims to forecast the impact sensitivity and detonation characteristics of energetic materials, striking a balance between high detonation performances and low impact sensitivity. The predictive errors were independently computed for the training and test sets, as detailed in Table 2. Fig. 2 and Figs. S8–S10 illustrate the performance of four models in predicting ln(H_{50}), *D*, *P*, and *E* on both training and test sets, as well as the distribution of predictive residuals across the entire dataset (Figs. S11–S14).

Table 1
ML algorithms and characteristics [29,34].

Algorithm	Characteristic
Adaboost	Adaboost is an iterative algorithm with a core principle of adjusting sample weights based on the error rate from the previous iteration, thereby enhancing the accuracy of the model. It is well-suited for binary classification tasks.
SVR	SVR is a regression algorithm based on support vector machines. Its aim is to construct a function on the input data such that most data points fall within a specified margin of error. SVR is well-suited for regression tasks involving high-dimensional data and can utilize kernel functions to handle nonlinear relationships.
RF	RF is a unified learning method built upon decision trees. It constructs multiple decision trees by selecting features and samples randomly. The final result is determined by combining the outputs of these trees, either by voting for classification or by averaging for regression. This method lowers the risk of overfitting and enhances the accuracy.
KRR	KRR combines ridge regression with kernel methods. It employs L2 regularization to control the complexity of the model, while using a kernel function to map data into a higher-dimensional feature space to address nonlinear problems.

2.4. High-throughput DFT calculations

All the optimized structures of molecules were achieved through density functional theory (DFT) calculations at the BP86-D3 level [48,49], utilizing Ahlrichs' def2-SVP basis set [50] within the Gaussian 09 software package [51]. Electronic and vibrational properties were also conducted at this theoretical level. The frequency analysis of the 222 molecules showed no imaginary frequencies, confirming that these structures represent energy minima. The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), HOMO level ($E_{HOMO} = -I$, *I* is ionization potential), frequency (the lowest frequency of

Table 2
The performance metrics of four machine learning models for predicting impact sensitivity and detonation properties (ln(H_{50}), *D*, *P*, and *E*) of energetic materials, detailed separately for both the training and testing datasets.

Properties			Training dataset			Test dataset		
		Model	RMSE	*r*	R^2	RMSE	*r*	R^2
Drop height	ln(H_{50})	AdaBoost	0.409	0.913	0.819	0.574	0.828	0.681
		KRR	0.472	0.759	0.871	0.576	0.830	0.680
		RF	0.212	0.978	0.949	0.550	0.870	0.740
		SVR	0.405	0.909	0.822	0.585	0.819	0.670
Detonation velocity	*D*	AdaBoost	0.103	0.983	0.965	0.162	0.973	0.936
		KRR	0.110	0.980	0.960	0.140	0.980	0.952
		RF	0.055	0.995	0.990	0.154	0.976	0.942
		SVR	0.110	0.980	0.960	0.127	0.981	0.960
Detonation pressure	*P*	AdaBoost	7.771	0.984	0.966	13.374	0.967	0.925
		KRR	8.351	0.980	0.961	9.529	0.983	0.962
		RF	4.084	0.996	0.990	12.287	0.971	0.937
		SVR	7.589	0.984	0.967	8.072	0.987	0.973
Gurney energy	*E*	AdaBoost	265.087	0.967	0.929	437.361	0.933	0.854
		KRR	377.806	0.925	0.855	413.469	0.945	0.869
		RF	136.144	0.991	0.981	405.125	0.940	0.875
		SVR	324.440	0.945	0.893	363.897	0.951	0.899

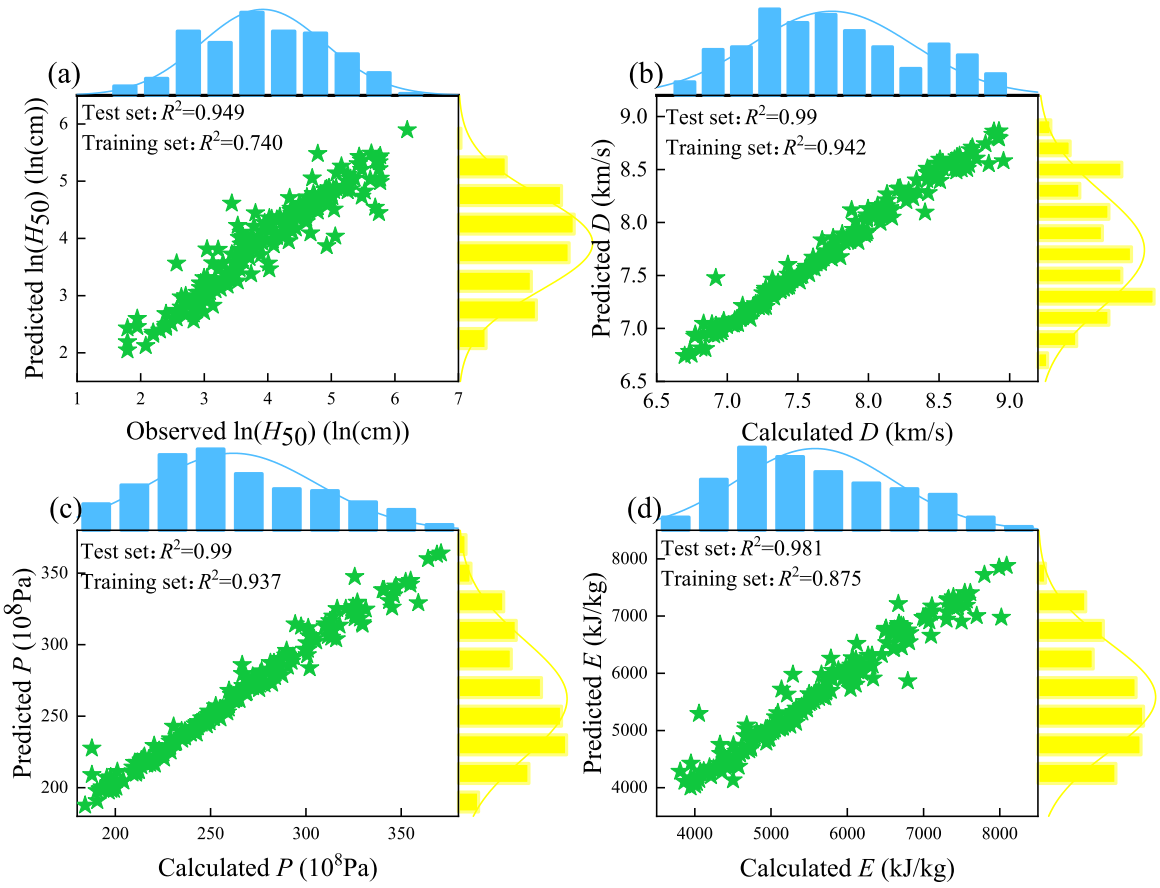


Fig. 2. Histograms and scatter plots comparing the observed values vs. predicted values of ln(H_{50}), detonation velocity D, detonation pressure P, and Gurney energy E using the Random Forest (RF) model. (a) Histogram and scatter plot of observed vs. predicted ln(H_{50}) values; (b) histogram and scatter plot of observed vs. predicted D values; (c) histogram and scatter plot of observed vs. predicted P values; (d) histogram and scatter plot of observed vs. predicted E values.

the vibrational mode), zero-point vibrational energy, polarizability, and dipole moment have been selected as feature descriptors for machine learning.

3. Results and discussion

Correlation analysis: Prior to modeling the logarithmically transformed H_{50} (ln (H_{50})), along with D, P, and E as functions of various descriptors, the assessments of the relationship between each descriptor and these four target variables and the interrelationship among the descriptors themselves were conducted. In Figs. S1–S7, it can be observed that the descriptors with the strongest correlation to the ln(H_{50}) were the heat of explosion followed by oxygen balance. For the prediction of D, P, and E (excluding the heat of explosion due to its role as a performance indicator for explosives), the descriptor with the strongest correlation was the oxygen balance. The Pearson correlation coefficient was employed to quantify the linear relationship between pairs of variables, reflecting how one variable changes to another in a fixed proportion. Selecting descriptors that exhibit a higher correlation with the target variables can enhance the training efficiency of machine learning models. The criteria for choosing input features for our ML models were based on a r absolute value greater than 0.15 with the target variables. Fig. 3 and Tables S2–S5 display the r between the input features and ln(H_{50}), D, P, and E. The number of input features for ln(H_{50}), D, P, and E was identified as 17, 16, 18, and 20, respectively, which are detailed in Tables S6–S9.

Impact sensitivity characterized based on drop height: In the prediction of drop height (ln(H_{50})), the RF model achieved the most accurate predictions of ln(H_{50}), with an RMSE of 0.212 cm, r of 0.978, and R^2 of 0.949. Similar accuracy was observed in the test set (RMSE of 0.55 cm, r of 0.87, and R^2 of 0.74), summarized in Table 2. Histogram, scatter plot of observed vs. predicted ln(H_{50}) values, and residual distribution using four models are depicted in Fig. 2(a), and Figs. S8(a), S9(a) S10(a), S11(a), S12(a), S13(a), S14(a), and S15, respectively. It can be seen that the residual distribution plot of the predicted impact sensitivity closely resembles a normal curve. Furthermore, a simple statistical review highlighted that 56.3% of the RF model predictions of ln(H_{50}) had a relative error of less than 5% compared to the experimental impact sensitivity values, and 84.7% of the predictions exhibited a relative error below 10%.

Detonation performances characterized based on detonation velocity (D), detonation pressure (P), and Gurney energy (E): In the process of predicting detonation velocity (D), all four models demonstrated commendable performance. The RMSE was uniformly below 0.11 km/s, the r exceeded 0.98, and the R^2 was all above 0.96, as shown in Table 2. Among these models, the RF model exhibited the highest congruence with D calculated through the KJ equation for the training set, achieving metrics of RMSE of 0.055 km/s, r of 0.995, and R^2 = 0.990. The SVR model mirrored this accuracy for the training set, with the testing set scoring RMSE of 0.127 km/s, r of 0.981, and R^2 of 0.960. Fig. 2(b) and Fig. S10(b) display the predictions of D using the RF and SVR models. For the RF model, 99.5% of predictions had a relative error less than 5% compared to the calculated values, and all predictions were within a 10% error margin. The SVR model showed similar accuracy, with 99.5% of predictions within a 5% error margin and 100% within

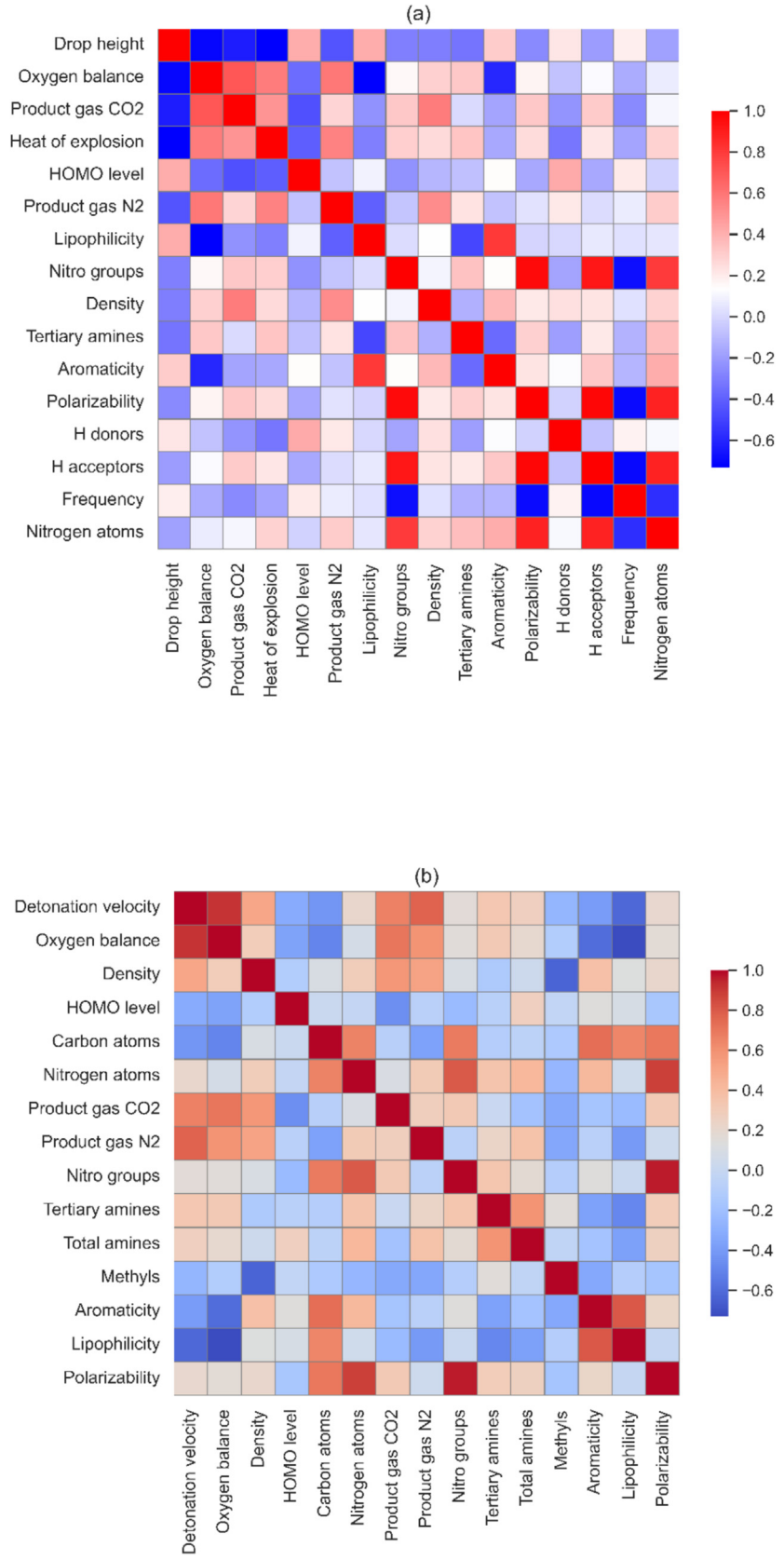
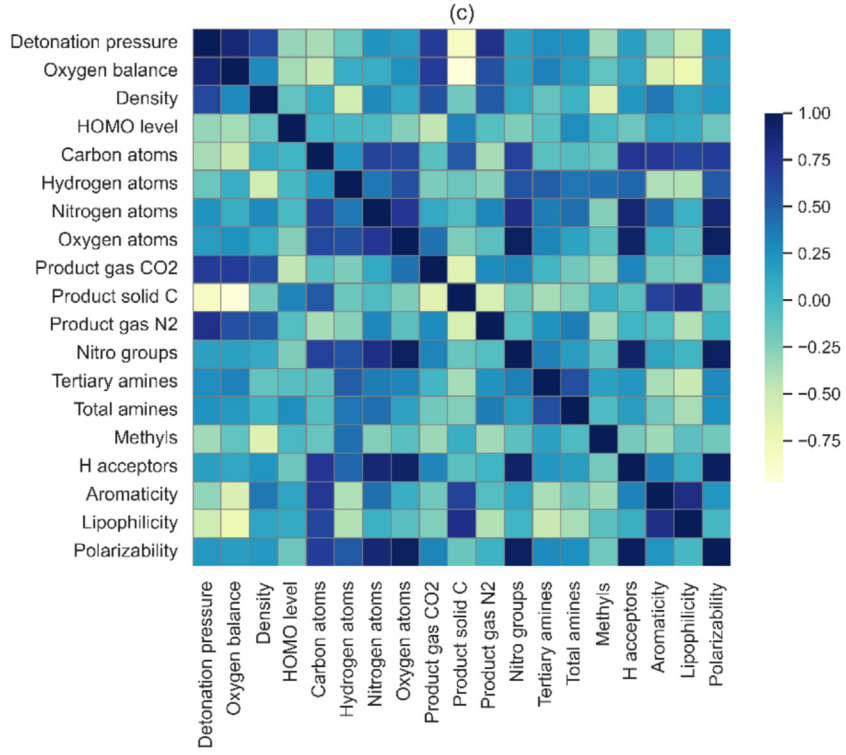


Fig. 3. Hot maps of Pearson correlation coefficients between the input descriptors used for predicting different properties. (a) Hot map of Pearson correlation coefficients between the input descriptors for drop height ($\ln(H_{50})$); (b) Hot map of Pearson correlation coefficients between the input descriptors for detonation velocity (D); (c) Hot map of Pearson correlation coefficients between the input descriptors for detonation pressure (P); (d) Hot map of Pearson correlation coefficients between the input descriptors for Gurney energy (E).



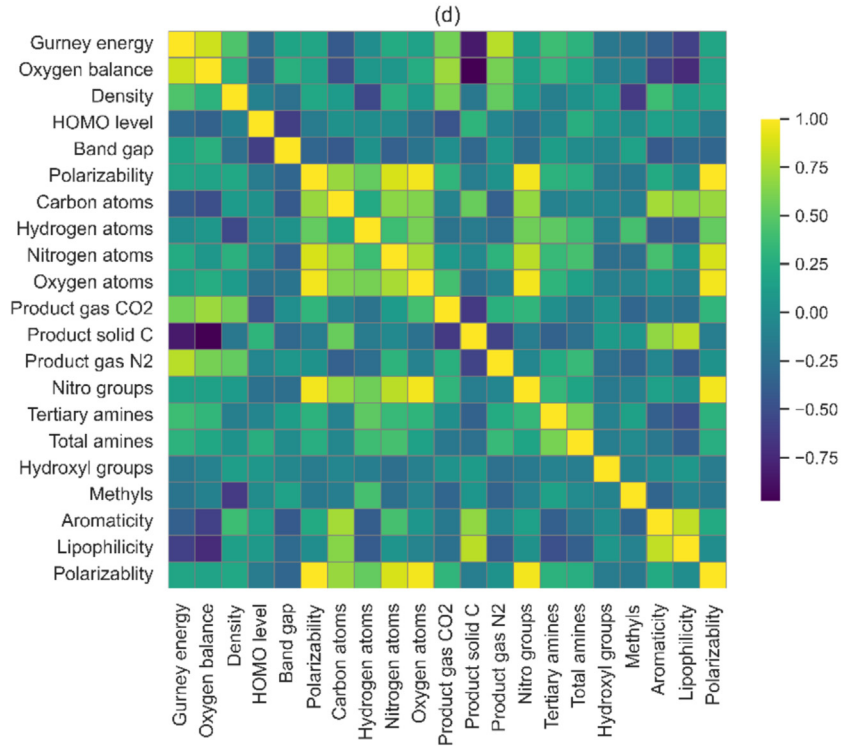


Fig. 3 (*continued*)

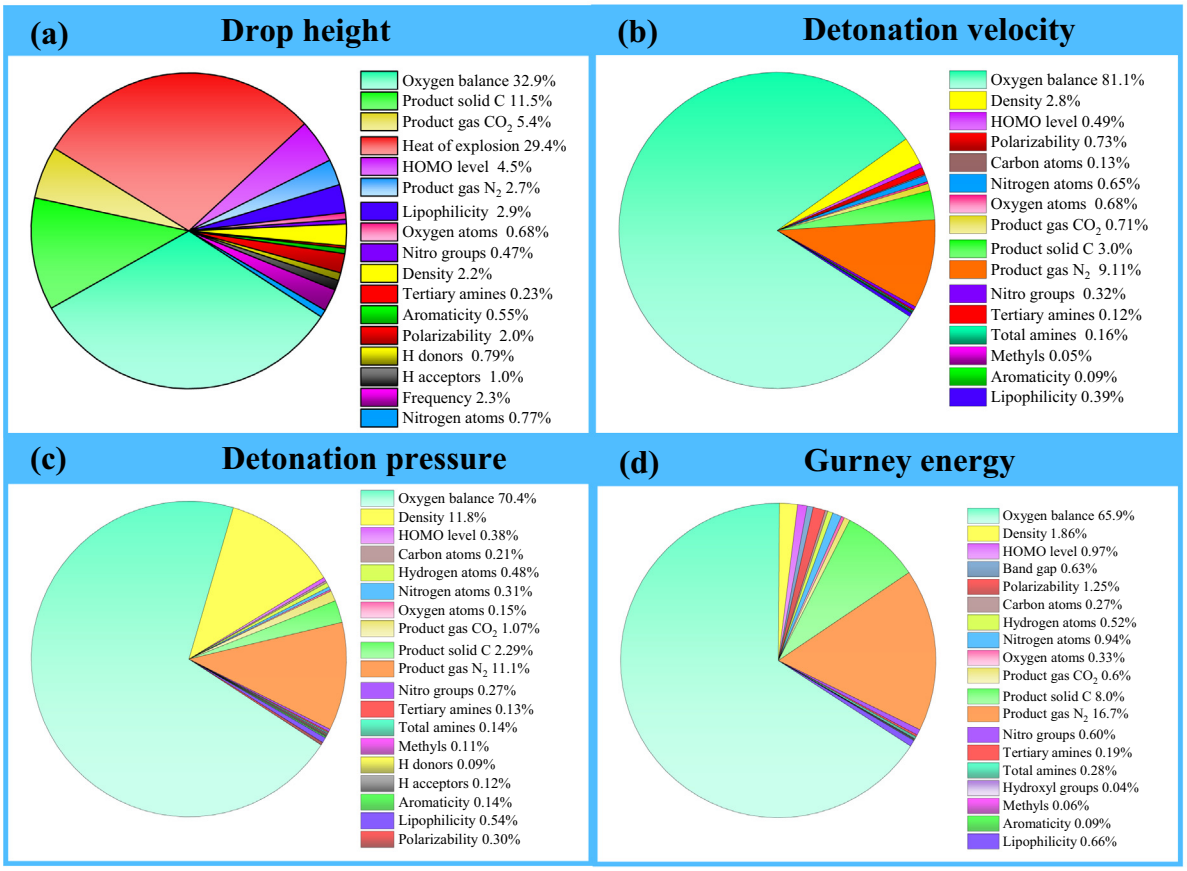


Fig. 4. The pie chart of importance ranking of feature descriptors (a) ln(H_{50}), (b) D, (c) P, and (d) E in predictions of drop height and detonation properties of EMs.

10% error margin compared to the calculated values. Figs. S11(b) and S14(b) illustrate the residual distributions of predictions from the RF and SVR models, which approximately follow a normal distribution.

In the prediction process of P, all four models exhibited remarkable performance. Each model achieved a RMSE less than 8.351×10^8 Pa, with r exceeding 0.98 and R^2 above 0.96, as detailed in Table 2. Mirroring the case with D, the RF model demonstrated the best consistency with P calculated through the KJ equation for the training set, marked by RMSE of 4.084×10^8 Pa, r of 0.996, and R^2 of 0.990. For the training set, the SVR model also showed excellent consistency with P as calculated, recording the RMSE of 8.072×10^8 Pa, r of 0.987, and R^2 of 0.973 in the testing set. Fig. 2(c) and Fig. S10(c) illustrate the predictions of P by the RF and SVR models. With the RF model, 94.6% of predictions had a relative error less than 5% compared to the calculated values, and 99.1% of predictions had errors less than 10%. The SVR model had 90.5% of predictions within a 5% error range and 99.5% of predictions within 10%. Figs. S11(c) and S14(c) present the residual distributions of predictions from both RF and SVR models, which generally follow a normal distribution.

In the process of predicting E, all four models displayed nice performance with their RMSE values all being less than 377 kJ/kg, r exceeding 0.92, and R^2 over 0.85, as indicated in Table 2. Similar to the predictions of D and P, the RF model showed the highest consistency with E_G values calculated using the KJ equation for the training set, evidenced by its scoring metrics of RMSE of 136.144 kJ/kg, r of 0.991, and R^2 of 0.981. The SVR model illustrated this high level of consistency for the training set, with the testing set achieving an RMSE of 363.897 kJ/kg, r of 0.951, and R^2 of 0.899. Fig. 2(d) and Fig. S10(d) demonstrate the precise predictions of E_G using the RF and SVR models. For the RF model, 87.9% of the predictions had a relative error less than 5% compared to the calculated values, and 96.9% of the predictions had an error less than 10%. In the case of the SVR model, 72.2% of predictions had less than 5% relative error, and 90.1% of the predictions were within 10% error margin. Figs. S11(d) and S14(d) depict the residual distributions of predictions from the RF and SVR models, which also follow the normal distributions.

Feature importance ranking in RF prediction of drop height (H_{50}): Scientists have been relentlessly pursuing the development of novel compounds, aiming to reduce sensitivity while maintaining high explosive performance. After training and evaluating four models for predicting ln(H_{50}), D, P, and E, the RF model, due to its superior performance, was selected for feature importance analysis of each attribute. The feature importance is assessed based on their contribution to improving model performance during the construction of decision trees in the random forest. Fig. 4 displays the ranking of feature importance for determining impact sensitivity and detonation properties. Fig. 4(a) shows the importance ranking of feature descriptors for determining ln(H_{50}). Key features include oxygen balance (OB), heat of explosion (Q), HOMO level

Table 3
The optimal ranges of critical features of energetic materials for balancing impact sensitivity and detonation performances.

Features	Optimal ranges
Oxygen balance	−40% to −30%
Density	1.66 to 1.72 g/cm^3
HOMO level	−6.34 to −6.01 eV
Lipophilicity	−1.0 to 0.1 and 4.49 to 5.59

(the negative values of the ionization potential), product solid C, product gas CO_2, lipophilicity, product gas N_2, density, polarizability, and frequency. Each of these 10 features contributes over 2% of the importance percentage. Among these, OB and Q are crucial in determining impact sensitivity, where higher OB and higher Q lead to lower ln(H_{50}), contradicting the requirement for high explosive performance (Fig. 3). Furthermore, the ionization potential is an important characteristic for breaking covalent bond in energetic materials, playing a significant role in determining impact sensitivity (as shown in Fig. 3a and 4a).

Feature importance ranking in RF predictions of detonation velocity (*D*), detonation pressure (*P*), and Gurney energy (*E*): In predicting *D*, *P*, and *E*, the common features of the top four important characteristics are OB, the presence of gaseous N_2 in the reaction products, the presence of solid C, and density. OB is ranked as the most significant feature, accounting for 81.1%, 70.4%, and 65.9% of the feature importance pie charts in predicting *D*, *P*, and *E*, respectively. The presence of gaseous N_2 in the reaction products is the second most important feature, contributing 9.11%, 11.1%, and 16.7% to the importance percentages in predicting *D*, *P*, and *E*. In the prediction of *D* and *E*, the presence of solid C is ranked the third, while it is ranked the fourth in determining *P*, accounting for 3%, 2.29%, and 8% in the feature importance pie charts for *D*, *P*, and *E*, respectively. The *r* presented in Fig. 3 indicates that the presence of solid C has a negative impact on achieving high values of *D* and *P*. These results are consistent with recent findings that a significant reduction in *D* and *P* is caused by the formation of large clusters of solid carbon in the explosion products due to the lack of oxygen in the structure of energetic materials [52]. In predicting *D* and *E*, density is ranked the fourth, while it is the third in deter-

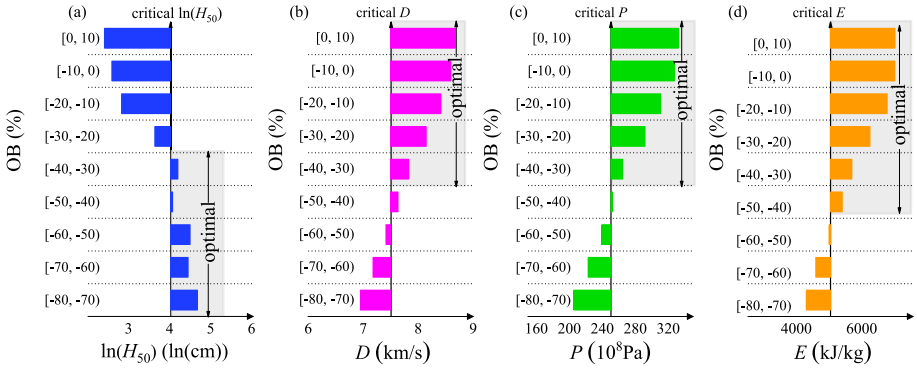


Fig. 5. The optimal ranges of oxygen balance in balancing impact sensitivity and detonation performances of EMs.

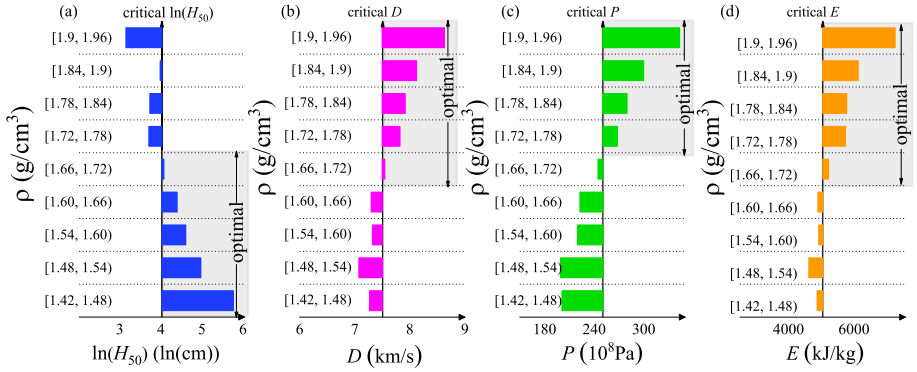


Fig. 6. The optimal ranges of density in balancing impact sensitivity and detonation performances of EMs.

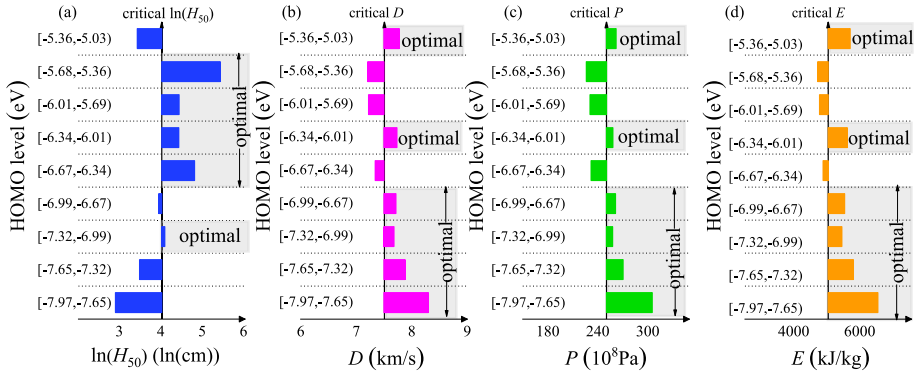


Fig. 7. The optimal ranges of HOMO level in balancing impact sensitivity and detonation performances of EMs.

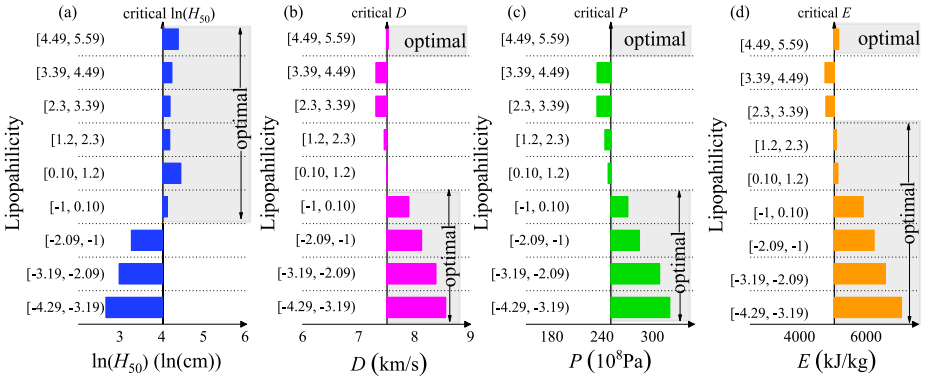


Fig. 8. The optimal ranges of lipophilicity in balancing impact sensitivity and detonation performances of EMs.

mining P, accounting for 2.8%, 11.8%, and 1.86% in the feature importance pie charts for D, P, and E, respectively. Additionally, it is found that the HOMO level and lipophilicity also influence the D, P, and E of these molecules.

Guidelines for balancing detonation performances and impact sensitivity: Based on the ranking of feature importance and statistical data, it is proposed to consider the impact sensitivity (ln(H_{50}) = 4.025 cm) of 2,3,4-trinitrotoluene and detonation properties of 2,4,6-trinitrobenzene-1,3,5-triamine (D = 7.507 km/s, P = 250.6 × 10^8 Pa, and E = 5048 kJ/kg) as the benchmark to balance the impact sensitivity and detonation performances of energetic materials using certain key features. The optimal ranges for all identified crucial characteristics are presented in Table 3. Figs. 5–8 display the relationships between OB, density, HOMO level, and lipophilicity across various ranges with the average impact sensitivity and average detonation performances. As observed in Figs. 3 and 4, OB, density, HOMO level, and lipophilicity are important parameters causing the contradiction between high sensitivity and high performance.

A 0% OB ranks the first among all features in providing optimal detonation performances, while also ranking the first in reducing impact sensitivity (excluding heat of explosion). As shown in Fig. 5, the ideal range for OB to balance impact sensitivity and detonation performances is between −40% and −30%. Fig. 6 indicates that higher densities correlate with higher impact sensitivities and better detonation performances, suggesting a density range from 1.66 to 1.72 g/cm^3 for balancing both. Fig. 3 and Table S2 demonstrate that lipophilicity and HOMO level also influence impact sensitivity and detonation performances. Lipophilicity refers to the ability of a molecule to interact with lipids or non-polar solvents. High lipophilicity means that molecules are more likely to attract each other in a non-polar environment, and this mutual attraction causes the molecules to pack more tightly together. The van der Waals forces between molecules become stronger, which increases the overall stability of the material. Since highly lipophilic energetic materials tend to lack high-energy polar groups such as nitro groups, the energy released during detonation is significantly lower than that of compounds containing a large number of nitro groups. The suggested ranges are −6.34 to −6.31 eV for HOMO level, and −1.0 to 0.1 and 4.49 to 5.59 for lipophilicity as shown in Figs. 7 and 8, respectively.

4. Conclusions

In this study, quantum chemical computations and machine learning methods were utilized to better predict the impact sensitivity and detonation performances of energetic materials, thereby facilitating their rational design by providing optimal ranges for key features. Specifically, electronic and vibrational parameters of 222 energetic materials were calculated based on density functional theory, and other feature descriptors were obtained using the open-source toolkit RDKit. These parameters and features were then used to train four ML models: AdaBoost, SVR, RF, and KRR. The training and testing sets were divided in a 4:1 ratio, and over 4000 cycles of grid search and cross-validation were conducted to optimize the hyperparameters for predicting ln(H_{50}), D, P, and E of energetic materials. The performance metrics and distribution of prediction residuals demonstrated that the RF model exhibited the best performance in predicting impact sensitivity. The RF and SVR models showed optimal performance in training and testing sets for predicting detonation performances, respectively. Moreover, through feature importance analysis, the optimal ranges of key features balancing the impact sensitivity and detonation performances of energetic materials were identified as: (1) Oxygen balance: 70% to −30%, (2) Density: 1.54 to 1.72 g/cm^3, (3) HOMO level: −0.28 to −0.20 eV, (4) Lipophilicity: −1.0 to 5.59.

CRediT authorship contribution statement

Wei-Hong Liu: Writing – original draft, Formal analysis. **Qi-Jun Liu:** Writing – review & editing, Project administration, Conceptualization. **Fu-Sheng Liu:** Writing – review & editing. **Zheng-Tang Liu:** Software.

Declaration of competing interest

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

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. 2682024GF019).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jechem.2024.10.035.

References

[1] W. Huang, Y.X. Tang, G.H. Imler, D.A. Parrish, J.M. Shreeve, J. Am. Chem. Soc. 142 (2020) 3652.
[2] H. Li, L. Zhang, N. Petrutik, K.C. Wang, Q. Ma, D. Shem-Tov, F.Q. Zhao, M. Gozin, ACS Cent. Sci. 6 (2019) 54.
[3] M. Abd-Elghany, T.M. Klapötke, Phys. Sci. Rev. 3 (2018) 20170103.
[4] J.H. Zhang, L.A. Mitchell, D.A. Parrish, J.M. Shreeve, J. Am. Chem. Soc. 137 (2015) 10532.

[5] L. Hu, P. Yin, G. Zhao, C.L. He, G.H. Imler, D.A. Parrish, H.X. Gao, J.M. Shreeve, J. Am. Chem. Soc. 140 (2018) 15001.
[6] W. Qian, J. Huang, S.T. Guo, B.W. Duan, W.Y. Xie, J. Liu, C.Y. Zhang, Energ. Mater. Front. (2023). https://doi.org/10.1016/jenmf.2023.05.002.
[7] W.L. Yuan, L. He, G.H. Tao, J.M. Shreeve, Acc. Mater. Res. 2 (2021) 692.
[8] National Research Council, Division on Engineering, Physical Sciences, National Materials Advisory Board, & Committee on Materials Research for Defense After Next. (2003). Materials research to meet 21st-century defense Needs. National Academies Press.
[9] T.R. Ravindran, R. Rajan, V. Venkatesan, J. Phys. Chem. C. 123 (2019) 29067.
[10] G. Zhao, P. Yin, R. Staples, J.M. Shreeve, Chem. Eng. J. 412 (2021) 128697.
[11] S.V. Sysolyatin, A.A. Lobanova, Y.T. Chernikova, G.V. Sakovich, Russ. Chem. Rev. 74 (2005) 757.
[12] A.T. Nielsen, A.P. Chafin, S.L. Christian, D.W. Moore, M.P. Nadler, R.A. Nissan, D. J. Vanderah, R.D. Gilardi, C.F. George, J.L. Flippen-Anderson, Tetrahedron 54 (1998) 11793.
[13] D. Fischer, J.L. Gottfried, T.M. Klapötke, K. Karaghiosoff, J. Stierstorfer, T.G. Witkowski, Angew. Chem. 128 (2016) 16366.
[14] N.V. Muravyev, D.R. Wozniak, D.G. Piercey, J. Mater. Chem. A 10 (2022) 11054.
[15] D. Mathieu, T. Alaime, J. Mol. Graph. Model. 62 (2015) 81.
[16] P. Politzer, J.S. Murray, J. Mol. Model. 21 (2015) 1.
[17] M.J. Kamlet, H.G. Adolph, Propell. Explos. Pyrot. 4 (1979) 30.
[18] C.Y. Zhang, Y.J. Shu, Y.G. Huang, X.D. Zhao, H.S. Dong, J. Phys. Chem. B. 109 (2005) 8978.
[19] J.S.M.P.L. Peter, Mol. Phys. 93 (1998) 187.
[20] J.S. Li, J. Phys. Chem. B. 174 (2010) 728.
[21] J.S. Li, J. Phys. Chem. B. 114 (2010) 2198.
[22] S.H. Ge, X.L. Cheng, L.S.W.X.D. Yang, J. Mol. Struc.: THEOCHEM. 809 (2007) 55.
[23] S.J. Ye, K. Tonokura, M. Koshi, Combust. Flame. 132 (2003) 240.
[24] A.A.L. Michalchuk, M. Trestman, S. Rudić, P. Portius, P.T. Fincham, C.R. Pulham, C.A. Morrison, J. Mater. Chem. A. 7 (2019) 19539.
[25] A.A. Michalchuk, J. Hemingway, C.A. Morrison, J. Chem. Phy. 154 (2021) 1.
[26] D. Mathieu, T. Alaime, J. Phys. Chem. A. 118 (2014) 9720.
[27] R. Claveau, J. Glorian, D. Mathieu, Phys. Chem. Chem. Phys. 25 (2023) 10550.
[28] G. Li, C.Y. Zhang, J. Hazard. Mater. 398 (2020) 122910.
[29] A. Velidandi, P.K. Gandam, M.L. Chinta, S. Konakanchi, A.R. Bhavanam, R.R. Baadhe, M. Sharma, J. Gaffey, Q.D. Nguyen, V.K. Gupta, J. Energy. Chem. 81 (2023) 42.
[30] K. Luo, X. Chen, H.R. Zheng, Z.C. Shi, J. Energy. Chem. 74 (2022) 159.
[31] N. Parikh, M. Karamta, N. Yadav, M.M. Tavakoli, D. Prochowicz, S. Akin, A. Kalam, S. Satapathi, P. Yadav, J. Energy. Chem. 66 (2022) 74.
[32] T. Yu, C. Wang, H. Yang, F. Li, J. Energy. Chem. 90 (2023) 191.
[33] P.K. Kanti, P. Sharma, K.V. Sharma, M.P. Maiya, J. Energy. Chem. 82 (2023) 359.
[34] X.Y. Liu, X.Q. Zhang, X. Chen, G.L. Zhu, C. Yan, J.Q. Huang, H.J. Peng, J. Energy. Chem. 68 (2022) 548.
[35] C. Chen, D.Y. Liu, S.Y. Deng, L.X. Zhong, S.H.Y. Chan, S.Z. Li, H.H. Hng, J. Energy. Chem. 63 (2021) 364.
[36] Y.B. Wu, C. He, W.X. Zhang, J. Energy. Chem. 82 (2023) 375.
[37] Y. Lu, D. Wei, W. Liu, J. Meng, X.M. Huo, Y. Zhang, Z.Q. Liang, B. Qiao, S.L. Zhao, D.D. Song, Z. Xu, J. Energy. Chem. 77 (2023) 200.
[38] T.Y. Liu, X. Zhao, X.F. Liu, W.J. Xiao, Z.J. Luo, W.T. Wang, Y.F. Zhang, J.C. Liu, J. Energy. Chem. 81 (2023) 93.
[39] M. Chen, Z.H. Yin, Z.C. Shan, X.K. Zheng, L. Liu, Z.H. Dai, J. Zhang, S.Z. Frank Liu, Z. Xu, J. Energy. Chem. 94 (2024) 254.
[40] C.B. Storm, J.R. Stine, J.F. Kramer, Chem. Phys. Energy. Mater. 309 (1990) 605.
[41] F.W. Marrs, J.V. Davis, A.C. Burch, G.W. Brown, N. Lease, P.L. Huestis, M.J. Cawkwell, V.W. Manner, J. Chem. Inf. Model. 63 (2023) 753.
[42] D. Mathieu, Ind. Eng. Chem. Res. 56 (2017) 8191.
[43] S. Beaucamp, D. Mathieu, V. Agafonov, Acta Crystallogr. B. 63 (2007) 277.
[44] M.J. Kamlet, S.J. Jacobs, J. Chem. Phys. 48 (1968) 23.
[45] G.F. Kinney, K.J. Graham, Explosive shocks in air, Springer Berlin, Heidelberg, 2013, doi: 10.1007/978-3-642-86682-1.
[46] X.N. Huang, C.Y. Li, K.Y. Tan, Y.S. Wen, F. Guo, M. Li, Y.L. Huang, C.Q. Sun, M. Gozin, L. Zhang, Iscience. 24 (2021) 1.
[47] D.C. Elton, Z. Boukouvalas, M.S. Butrico, M.D. Fuge, P.W. Chung, Sci. Rep. 8 (2018) 9059.
[48] A.D. Becke, Phys. Rev. A. 38 (1988) 3098.
[49] J.P. Perdew, Phys. Rev. B. 33 (1986) 8822.
[50] F. Weigend, R. Ahlrichs, Phys. Chem. Chem. Phys. 7 (2005) 3297.
[51] M. Frisch, F. Clemente. "Gaussian 09, revision a. 01, mj frisch, gw trucks, hb schlegel, ge scuseria, ma robb, jr cheeseman, g." Scalmani, V. Barone, B. Mennucci, GA Petersson, H. Nakatsuji, M. Caricato, X. Li, HP Hratchian, AF Izmaylov, J. Bloino, G. Zhe (2009) 20–44.
[52] C.Y. Li, Y.L. Huang, C.Q. Sun, L. Zhang, Chin. J. Energ. Mater. 28 (2020) 854.