

Screening High-Efficiency Deep Eutectic Solvents for Oxidative-Extractive Desulfurization via a Machine Learning Framework to Analyze Intermolecular Interactions

Changqing He^{1,2,†}, Jiahao Xu^{1,2,†}, Anupa Gunawardane^{1,2}, Yang Xue^{1,2}, Xueqian Liu^{1,2}, Lin He^{1,2,*} and Hong Sui^{1,2}

¹ School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China

² National Engineering Research Center of Distillation Technology, Tianjin 300072, China

* Correspondence: linhe@tju.edu.cn

† These authors contributed equally to this work.

How To Cite: He, C.; Xu, J.; Gunawardane, A.; et al. Screening High-Efficiency Deep Eutectic Solvents for Oxidative-Extractive Desulfurization via a Machine Learning Framework to Analyze Intermolecular Interactions. *Smart Chemical Engineering* **2026**, *2*(3), 11. <https://doi.org/10.53941/sce.2026.100011>

Received: 30 July 2026

Revised: 19 September 2026

Accepted: 21 September 2026

Published: 28 September 2026

Abstract: Oxidative-extractive desulfurization (OEDS) represents a highly efficient pathway for removing recalcitrant thiophenic sulfides from liquid fuels. Deep eutectic solvents (DESs) have emerged as promising green extractants due to their tunable multi-site non-covalent interactions and favorable environmental profile. However, exploring the vast chemical space of potential DES formulations demands rational, data-driven methodologies to identify molecular affinities while bypassing costly experimental trial-and-error. Herein, we report a novel hybrid algorithm termed PDI-DMPNN+SME. By coupling solvation free energy predictions with substructural masking, this framework quantitatively decodes the affinity and repulsion aimed at target molecules. The algorithm systematically evaluates both the extraction affinity of candidate DESs toward oxidized thiophenic species and the internal eutectic feasibility between solvent components. Through high-throughput screening of 19,291 binary candidates, an environmentally benign citric acid-glycerol DES system was identified as a promising extractant candidate. Under optimized process conditions, this system achieved a single-stage dibenzothiophene (DBT) removal efficiency of 99.95% in model oil and 93.72% in real diesel, maintaining satisfactory performance (98.50% and 90.80%, respectively) after the tested three consecutive regeneration cycles. Density functional theory (DFT) calculations reveal that the total interaction energy between the citric acid-glycerol system and dibenzothiophene sulfone (DBTO₂) reaches −31.24 kcal/mol. Energy decomposition analysis indicates that electrostatic and dispersion interactions contribute comparably to this stabilization, jointly overcoming Pauli repulsion to form a thermodynamically favorable extractant-solute complex. This work establishes an intelligent, green paradigm for the rational design of functional solvents in advanced separation processes.

Keywords: desulfurization; deep eutectic solvent; high-throughput screening; machine learning; extractant design



1. Introduction

Deep desulfurization of transportation fuels is the core process of petroleum refining, driven by increasingly stringent global environmental regulations. Traditional hydrodesulfurization can effectively remove aliphatic organic sulfur compounds (such as sulfides and thiols). However, its desulfurization effect on aromatic sulfides with large steric hindrance is limited, especially thiophene and its derivatives (Figure 1a) [1]. The realization of ultra-low sulfur content by hydrodesulfurization requires harsh operating conditions (high temperature of 300 to 400 °C and high hydrogen partial pressure). It leads to high energy consumption and catalyst deactivation [2,3].

In order to overcome these thermodynamic and operational bottlenecks, non-hydrodesulfurization technology receives much attention. Among them, oxidative extraction desulfurization (OEDS) technology has become an attractive alternative due to its mild operating conditions and excellent selectivity. Unoxidized thiophene compounds have weak molecular polarity. Due to this, it is thermodynamically unfavorable to extract them directly from hydrocarbon-rich fuel phases. In a typical OEDS process, hydrogen peroxide (H_2O_2) selectively converts weakly polar thiophene sulfides into highly polar sulfoxides or sulfones. This conversion occurs under the catalysis of organic acids (such as acetic acid) (Figure 1b) [4,5]. As the extractant continuously sequesters these polar species into the solvent phase, the chemical oxidation equilibrium is driven forward, achieving deep desulfurization.

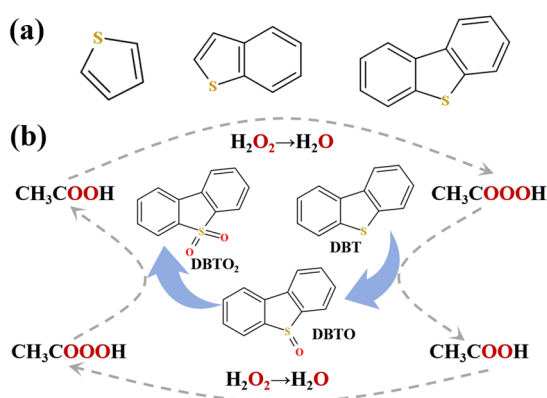


Figure 1. (a) Representative thiophenic sulfides present in oil. (b) Catalytic oxidation pathway of thiophenic sulfides using H_2O_2 under acetic acid catalysis.

The development of high-performance, sustainable extractants is fundamental to the industrial upgrading of OEDS. Deep eutectic solvents (DESs), typically formed by eutectic mixtures of hydrogen-bonding donors and acceptors [6], have emerged as viable alternatives to hazardous conventional organic solvents [7] (e.g., *N,N*-dimethylformamide, acetonitrile). The dense hydrogen-bonding networks within DESs provide customizable active sites for the selective recognition and binding of oxidized thiophenes. Existing experimental efforts have explored various DES classes, including *p*-toluenesulfonic acid- [8], oxalate- [4], choline- [5], polyethylene glycol- [9,10], carboxylic acid- [11], sultone- [12], and metal-based [13,14] DES systems. The maximum sulfur removal rate exceeds 99%, which is obtained using choline chloride/*p*-toluenesulfonic acid and *L*-proline/*p*-toluenesulfonic acid DESs [5,8].

Despite these advances, reported DES formulations represent a tiny fraction of the practically infinite compositional landscape resulting from combinatorial hydrogen-bonding donors/acceptors pairing [15–17]. Designing an optimal DES requires balancing multi-component synergistic binding, target selectivity, and environmental safety [18]. Traditional trial and error screening consumes resources, and its scope is limited. The machine learning paradigm provides a transformative way to accelerate the exploration of chemical space [19].

An algorithm for analyzing intermolecular interactions, namely PDI-DMPNN+SME, was proposed in this study. Specifically, quantum chemistry-derived polarity prior information was integrated into the directed message passing framework (DMPNN). Meanwhile, substructure mask explanation was employed to quantify non-covalent solute-solvent interactions at the atomic group resolution. Previous machine learning models typically dealt with the formation of DESs or solute solubility alone. In contrast, the proposed framework jointly evaluated solute-solvent affinity and solvent-solvent hydrogen bond stability. This enabled multi-dimensional DES screening. The effectiveness of this method was verified by screening 19,291 binary DES combinations. The best candidate system was further validated through extraction experiments. Moreover, its potential capture mechanism was clarified by density functional theory (DFT) calculations. Overall, this method provided a new perspective for the discovery of green desulfurizers.

2. Methods

2.1. Virtual Screening Workflow

The DES screening methodology integrated three major sequential filters: non-covalent affinity modeling, environmental assessment, and experimental validation (Figure 2). The initial candidate set consisted of 19,291 binary DES combinations extracted from the experimental literature (Table S2) [20]. Affinity screening relied on the PDI-DMPNN+SME coupling algorithm. In each binary system, the component was divided into Component-1 (primary extractant) and Component-2 (eutectic partner). Screening was carried out through two prediction stages:

- (i) Extraction affinity selection (Component-1): The affinity of each candidate for the oxidative thiophene species relative to the non-polar hydrocarbon matrix (n-octane/toluene) was evaluated. Candidates with high priority extraction affinity were separated into Component-1.
- (ii) Eutectic feasibility selection (Component-2): The intermolecular affinity between the candidate Component-2 and the selected Component-1 was calculated to ensure the formation of a good eutectic.

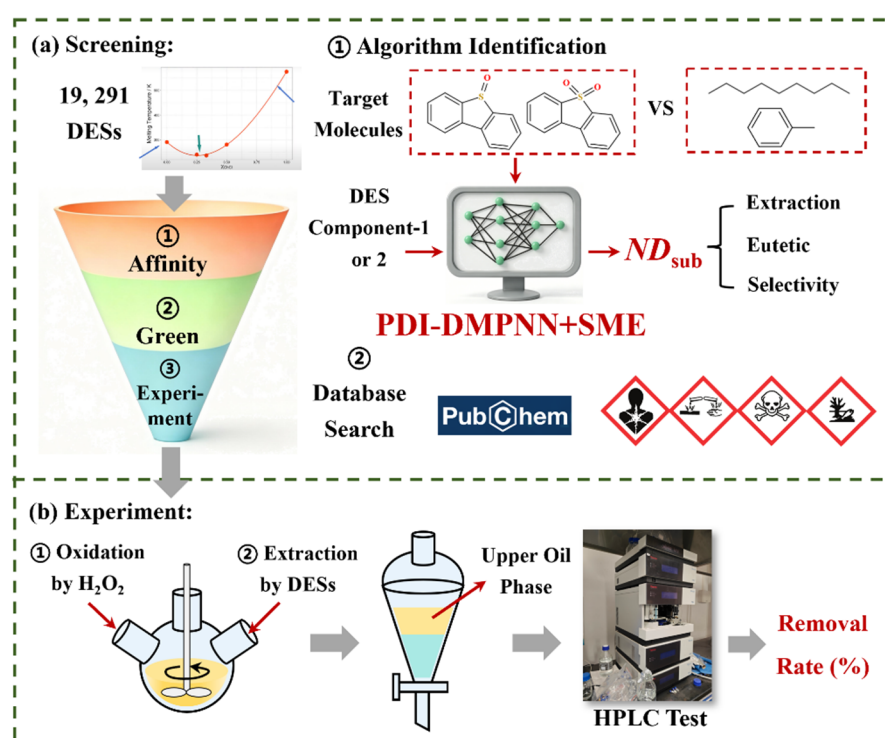


Figure 2. Overview of DES (a) screening and (b) experimental workflow in this study.

The molecules screened by two rounds of affinity screening were cross-matched with the PubChem database. This cross-matching aimed to exclude compounds with acute toxicity, flammability, or ecological hazards. A streamlined subset was thus obtained for experimental preparation and OEDS evaluation. Irritating substances were conditionally retained.

2.2. PDI-DMPNN+SME Algorithmic Architecture

The core algorithm (Figure 3) combined the prior data introducing-directed message passing neural network (PDI-DMPNN) with the substructure mask explanation (SME).

Module 1: Prior Data Introducing-Directed Message Passing Neural Network (PDI-DMPNN)

To ground graph representation learning in physical chemistry principles (“like dissolves like”), molecular graph embeddings were augmented with prior data (the absolute difference between the molecular polarity descriptors of the solute and the solvent, Equation (1)):

$$\text{Prior_data} = |\text{Polarity_Descriptor}_{\text{solute}} - \text{Polarity_Descriptor}_{\text{solvent}}| \quad (1)$$

These descriptors, derived from DFT calculations, included the molecular polarity index (MPI) [21], electrostatic potential (ESP) statistical distributions [22], and molecular surface area metrics (Table 1). Feature values were scaled via min-max normalization.

The graph network was configured with 10 message-passing steps, a hidden dimension of 1000, and a 10-layer feedforward prediction head utilizing ReLU activation functions. The model was trained using the Adam optimizer with a batch size of 50 over 200 epochs at a learning rate of 10^{-4} on a dataset of 997,682 solvation free energy (ΔG_{sol}) data points from Vermeire et al. [23]. The data were randomly split into 60% for training, 20% for validation, and 20% for testing. All results were obtained from a single final prediction on the test set. The test set was not accessed at any intermediate stage. Random seeds for data splitting and model initialization were fixed to 0. No explicit regularization was applied: dropout was set to 0.0, and no gradient clipping was used. Hyperparameter optimization was performed within the 80% training-validation set using 5-fold cross-validation with 30 training epochs, searching over learning rate, network layers, and prior data categories. The optimal hyperparameters were finalized, and the final model was trained for a maximum of 200 epochs with early stopping based on validation RMSE. Performance was evaluated using Root Mean Square Error (RMSE, Equation (2)) and the coefficient of determination (R^2 , Equation (3)).

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

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

Module 2: Substructure Mask Explanation (SME)

Substructural fragments were systematically generated via the BRICS (Breaking of Retrosynthetically Interesting Chemical Substructures, Section S3) fragmentation protocol [24]. In the SME module, specific fragment nodes and associated edges within the molecular graph were zero-masked. The ΔG contribution of masked substructure (D_{sub} , Equation (4)) and its normalized index ($ND_{\text{sub}} \in [-1, 1]$, Equation (5)) were computed as [25]:

$$D_{\text{sub}} = \Delta G_{\text{sol_whole}} - \Delta G_{\text{sol_masked}} \quad (4)$$

$$ND_{\text{sub}} = \frac{e^{D_{\text{sub}}} - e^{-D_{\text{sub}}}}{e^{D_{\text{sub}}} + e^{-D_{\text{sub}}}} \quad (5)$$

Here, negative ND_{sub} values denoted attractive non-covalent interactions, whereas positive values denoted repulsive interactions. Absolute values approaching 1 signified strong interaction tendencies.

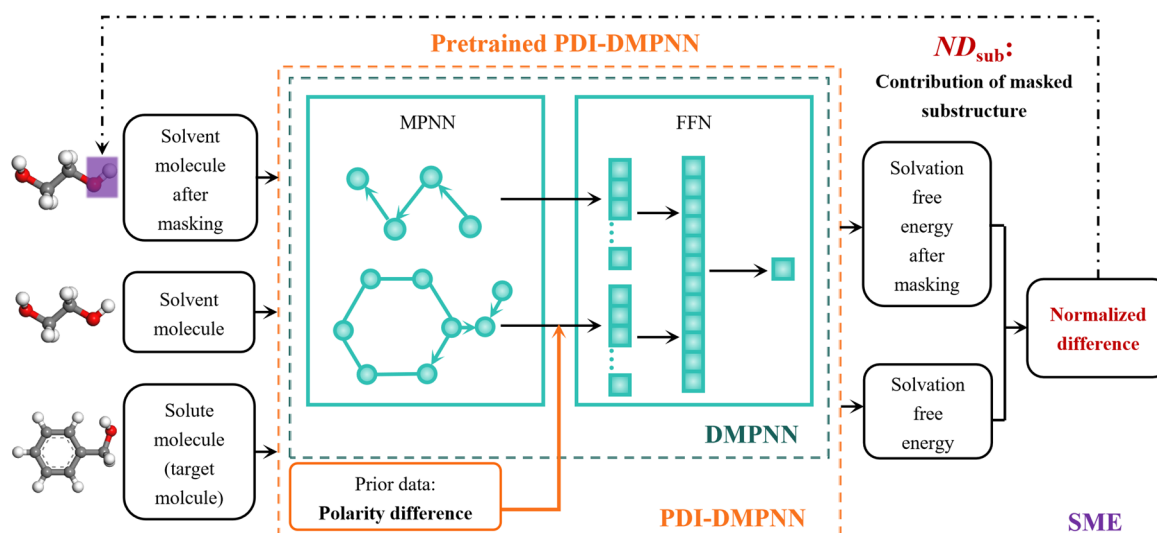


Figure 3. Architecture of the PDI-DMPNN+SME coupling algorithm. Reprinted with permission from Ref. [26]. 2026, Elsevier.

This study employed the PDI-DMPNN+SME coupling algorithm to conduct two sequential affinity predictions, which assessed the extraction and eutectic capabilities of the DES, respectively. The first prediction targeted various oxidized thiophene sulfur species (Figure S2) to evaluate the affinity of the DES, thereby reflecting its extraction performance. The second prediction focused on a single component within the DES to

assess the intermolecular affinity, which ensured the stability of the eutectic system. For the quantitative characterization of extraction performance, the parameter was introduced. This parameter was defined as the difference between the target molecule and the oil phase components (n-octane, toluene) as shown in Equation (6).

$$\Delta ND_{\text{sub}} = ND_{\text{sub_target}} - ND_{\text{sub_oil}} \quad (6)$$

Table 1. Molecular Polarity Descriptors Derived from DFT Calculations.

Category	Microscopic Physicochemical Data	Definitions	Equations
MPI	Molecular polarity index	The integral of the absolute value of ESP on the molecular surface, divided by the molecular surface area.	$\text{MPI} = \frac{1}{A} \iint_S \text{ESP} dS$
ESP statistics	ESP maximum	Global maximum value of ESP.	-
	ESP minimum	Global minimum value of ESP.	-
	Kurtosis of ESP distribution	Kurtosis is one of the statistics that measures peakness or flatness in the ESP distribution.	$\text{Kurtosis} = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2}$
	Skewness of ESP distribution	Skewness measures the degree of symmetry or asymmetry of the ESP distribution.	$\text{Skewness} = \frac{E[(X - \mu)^3]}{(E[(X - \mu)^2])^{3/2}}$
	Quantiles (0.25, 0.50, 0.75) of ESP distribution	Tree quartiles to divide the ESP distribution into four equal parts.	-
Local and global molecular surface area	Global surface area	Global molecular surface area.	-
	Surface area ($i < \text{ESP} < i + 10$ kcal/mol)	The molecular surface area corresponds to $ \text{ESP} $ ranging from i to $i + 10$ kcal/mol ($i = 0, 10, 20, 30, 40, 50, 60, 70$).	-
Remarks: A: global surface area; S: molecular surface; μ : mean value; E: expectation operator.			

ΔND_{sub} served as a metric for the transfer tendency of target molecules from the oil phase to the DES phase. Dibenzothiophene (DBT) served as the model compound in the experimental section. The oxidation products of dibenzothiophene, especially the sulfoxide and sulfone forms, were named dibenzothiophene sulfoxide (DBTO) and dibenzothiophene sulfone (DBTO₂), respectively.

Candidate substances did not use a fixed threshold, but were sorted according to the model output results. The top 20 candidate substances with the largest absolute negative values of ΔND_{sub} and ND_{sub} were selected, respectively. Then, the comprehensive evaluation was carried out based on environmental friendliness, raw material accessibility, and experimental operability.

2.3. Experimental Methods

The experimental procedure was carried out according to the following stages (Figure 2b):

Model Oil and DES Preparation: Model oil containing 1.437 g DBT dissolved in 500 mL n-octane/toluene mixture was prepared. Candidate DES systems were synthesized by mixing constituents at prescribed molar ratios at 100 °C under constant stirring (800 r/min) for 4 h until clear, homogeneous liquids were obtained. After optimizing experimental conditions using model oil, 1.437 g of DBT was added to the industrial diesel to simulate real diesel before desulfurization. The DBT concentration was approximately 436.5 mg/kg, while the negligible background sulfur content was far lower than this added sulfur source.

Oxidative-Extractive Desulfurization: Oxidation was initiated with 10 mL of model oil or real diesel. The molar ratio of hydrogen peroxide to acetic acid was 0.6. The reaction mixture was stirred at 500 rpm for 2 h at the target temperature. Subsequently, a certain amount of DES was added. The extraction was continued for 2 h under the same conditions. After phase separation, the upper fuel phase was analyzed by high-performance liquid chromatography (HPLC). An Agilent LC1260 system with a diode array detector was used for HPLC analysis. The detection wavelength and reference wavelength were set to 254 nm and 360 nm, respectively, and the bandwidth was 40 nm. The injection volume was 5 μ L. The total running time was 30 min. Quantitative analysis was performed by the external standard method. A five-point calibration curve was generated using DBT standards. The DBT removal rate (Equation (7)) was determined by the DBT concentration in the oil phase before (c_{before}) and after (c_{after}) extraction. This study also calculated the s of DES in three cycles of real diesel. The recovery rate (Equation (8)) was defined as the ratio of the mass of recovered DES (m_{recovery}) to the mass of initial DES (m_{initial}).

$$\text{Removal_Rate} = \left(1 - \frac{c_{\text{after}}}{c_{\text{before}}}\right) \times 100\% \quad (7)$$

$$\text{Recovery_Rate} = \frac{m_{\text{recovery}}}{m_{\text{initial}}} \times 100\% \quad (8)$$

Solvent Regeneration and Recycling: Spent DES phases were anti-solvent regenerated via ultrapure water addition, followed by rotary evaporation to remove residual moisture. Reclaimed solvents were subjected to consecutive OEDS cycles to assess operational longevity.

Chemicals were of analytical reagent grade, except for the high-purity standards used in HPLC. All reagents were purchased from Tianjin Heowns Biochemical Technology Co., Ltd. (Tianjin, China) and Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). The performance of DES was evaluated as the average of three independent experiments.

2.4. Computational and Quantum Chemical Details

DFT calculations were conducted using Gaussian 09, and the analyses were executed using Multiwfn 3.8 [27]. Initial conformer searching for multi-molecular clusters was executed via the Molclus platform using the MMFF94 force field, followed by preliminary geometry optimization at the semi-empirical GFN2-xTB level [28]. High-precision geometry optimizations were performed at the B3LYP-D3/6-31G(d) level with Grimme's empirical dispersion correction. Single-point energy calculations and wavefunction evaluations were conducted at the M062X/6-311+G(d,p) level.

The analysis includes molecular surface ESP distribution analysis, hydrogen-bonding energy calculation [29], and SobEDAw interaction energy decomposition [30].

Intermolecular hydrogen-bonding energy (a. u.) was estimated from bond critical point electron density (ρ) via Emamian's correlation (Equation (9)). It represented the energy difference between the two individual molecules in the post-complexation geometry and the same two molecules at infinite separation. The interaction energy (ΔE) between DBTO₂ and DES components was defined as the total energy minus the energy of each isolated component (Equation (10)). Using SobEDAw energy decomposition analysis, the interaction is decomposed into four physical and chemical natures (Equation (11)): electrostatic (ΔE_{els}), exchange repulsion (ΔE_{xrep}), orbital (ΔE_{orb}) and dispersion (ΔE_{disp}).

$$\text{Hydrogen_bonding_energy} = -223.08 \times \rho + 0.7423 \quad (9)$$

$$\Delta E = E_{\text{total}} - E_{\text{DBTO}_2} - E_{\text{DES}} \quad (10)$$

$$\Delta E = \Delta E_{\text{els}} + \Delta E_{\text{xrep}} + \Delta E_{\text{orb}} + \Delta E_{\text{disp}} \quad (11)$$

Table S1 could be consulted to obtain the mathematical definition of this article for easy reading.

3. Results and Discussion

3.1. DESs Screening

The predictive performance of PDI-DMPNN on the test set was depicted in Figure 4a. Incorporating physical polarity descriptors into the graph architecture decreased the prediction RMSE from 0.281 (unconstrained DMPNN) to 0.083 kcal/mol (a 70.46% accuracy gain, Figure S4), achieving an R^2 of 0.999. The feature importance analysis (Figure S3) confirmed that the molecular surface area index was the main control descriptor. This result provided a quantitative mathematical basis for the classical “like dissolves like” heuristic, where the solvation affinity fundamentally depends on the polarity between the solute and the solvent species.

Through the BRICS fragmentation method, 10,523 constituent substructures were extracted from the candidate compound library and evaluated using the SME. As shown in Figure 4b, the functional groups exhibited different ΔND_{sub} distributions. Oxygen-containing compounds (alcohols, phenols, polyols) showed strong negative potential. This indicates that the model could correctly and preferentially select multi-site electrostatic donors/acceptors with strong solvation ability. On the contrary, the non-polar hydrocarbon fragments (alkanes, cycloalkanes) were clustered closely around $\Delta ND_{\text{sub}} \approx 0$ due to the negligible electrostatic contribution. Among the top single substructures showing the highest extraction selectivity ($\Delta ND_{\text{sub}} \ll 0$), the polyhydroxy fragment stood out as the main candidate compound (Figure 4c). Considering the data-driven affinity index, environmental friendliness, and commercial availability, glycerol was finally selected as Component-1.

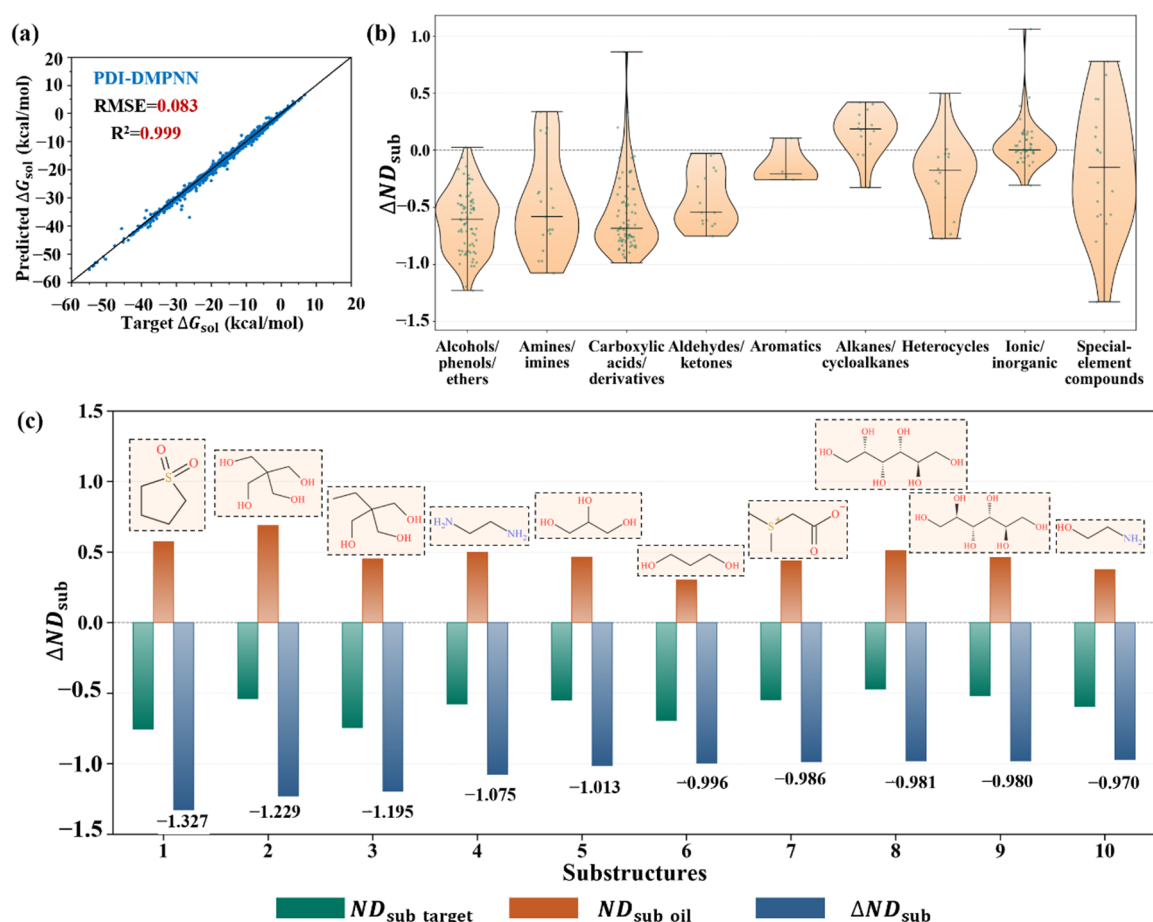


Figure 4. (a) Prediction accuracy of PDI-DMPNN. (b) Distribution violin plots for different substructure categories among components with selective extraction capability. (c) Top ten components with selective extraction capability.

The model was used to screen potential candidates for Component-2 to ensure a eutectic formation with the selected target molecule glycerol. As shown in Figure 5a, the violin plot showed the ΔND_{sub} distribution of different substructure categories in detail. The functional groups such as alcohols/phenols/ethers, carboxylic acids/derivatives, and heterocycles exhibited strong negative distribution centers. This trend indicated that the model successfully captured the physical and chemical properties of the system. A significantly higher eutectic contribution was endowed to the substructure that could form a wide hydrogen-bonding network with the polyhydroxy structure of glycerol. However, high theoretical affinity must be balanced against practical green chemistry constraints. Figure 5b categorized the top-performing eutectic candidates based on their safety and environmental profiles. Although some amine and amide derivatives exhibit excellent affinity, they were strictly excluded due to their practical hazards such as flammability, corrosiveness, or acute toxicity (Table S3). In contrast, components such as urea, lactic acid, and citric acid were considered ideal candidates, and they fitted perfectly into the green or irritant-only category. By combining data-driven performance with the principles of green chemistry, Figure 5c showed the three best Component-2 candidate components that were finally selected. Urea ($\Delta ND_{sub} = -1.872$), lactic acid ($\Delta ND_{sub} = -1.140$), and citric acid ($\Delta ND_{sub} = -1.085$) all showed strong predictive eutectic ability. Therefore, these three harmless and highly compatible components were selected as the final Component-2 candidate components paired with glycerol.

Accordingly, the above three DES systems (Table 2) were selected as extractants. Subsequent experiments focused on the optimization of the OEDS process using these three systems.

Table 2. Composition of Selected DESs.

Number	Component-1	Component-2	Molar Ratio
DES-1	Glycerol	Lactic acid	2:1
DES-2	Glycerol	Citric acid	2:1
DES-3	Glycerol	Urea	2:1

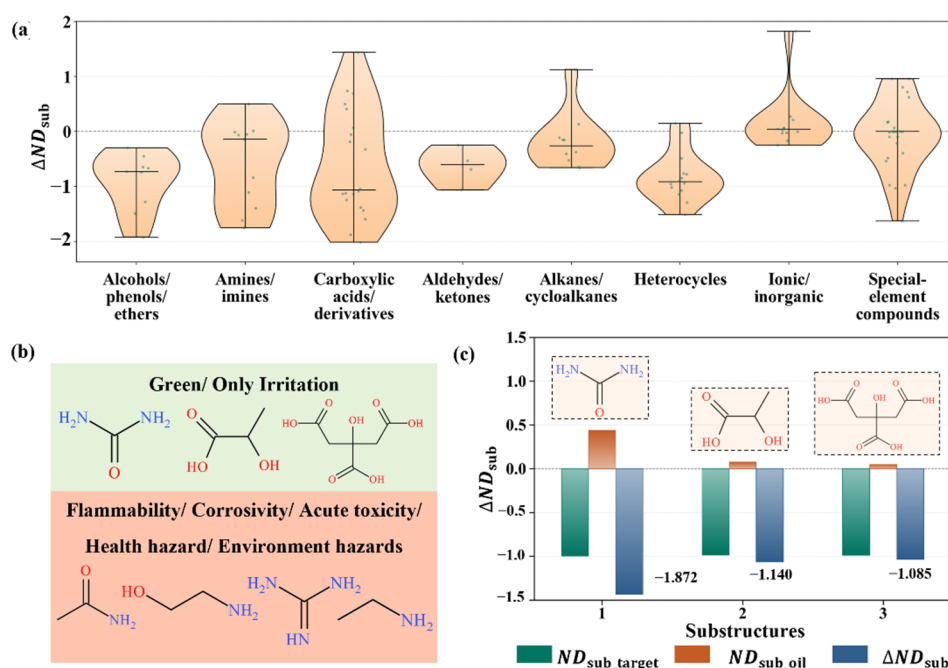


Figure 5. (a) Distribution violin plots for different substructure categories among eutectic components. (b) Harmless and harmful components among superior eutectic components. (c) Molecular structures of the three selected eutectic components.

3.2. OEDS Process Optimization and Performance Evaluation

Figure 6 illustrated the single-factor optimization experiments conducted on the model oil system under varying conditions. Figure 6a compared the desulfurization rates of DBT achieved by three DESs under initial conditions. The citric acid-glycerol DES exhibited the highest desulfurization efficiency at 93.25%, followed by the urea-glycerol DES at 90.87% and the lactic acid-glycerol DES at 81.41%. Consequently, the citric acid-glycerol system was determined for further optimization.

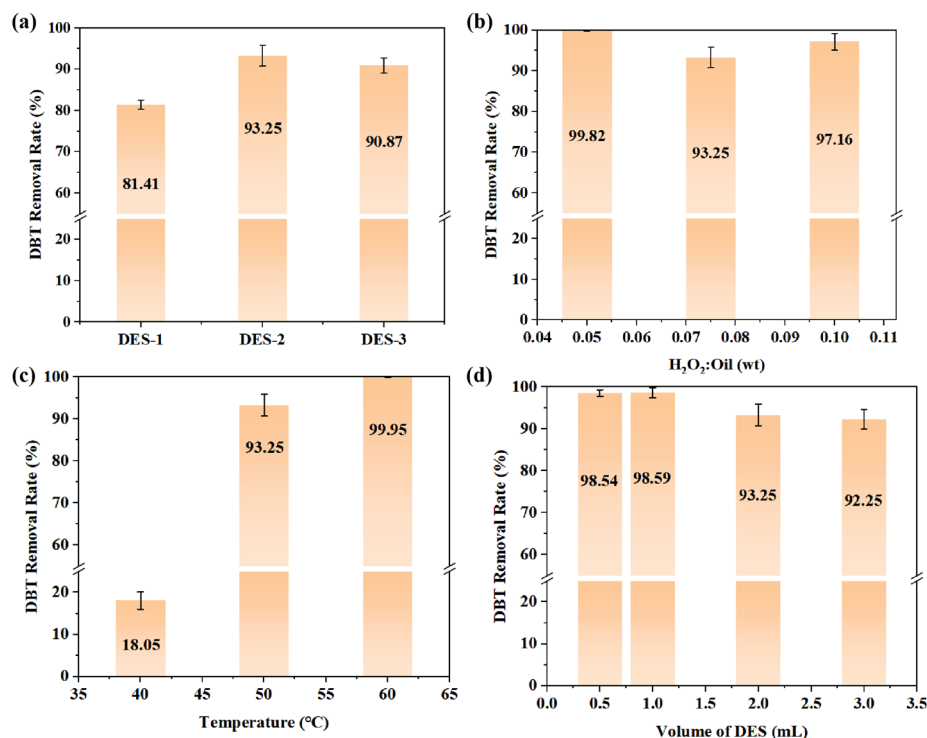


Figure 6. Influence of various factors on the DBT removal rate: (a) DES type; (b) H_2O_2 :Oil (wt); (c) temperature (°C); (d) DES volume.

Regarding the hydrogen peroxide dosage (Figure 6b), the desulfurization efficiency reached a peak of 99.82% at a mass ratio of 0.05. High oxidation selectivity and minimal water generation preserved the hydrogen-bonding network within the DES. This facilitated the solvation of sulfones and promoting the oxidation equilibrium toward the product. When the ratio increased to 0.075, the efficiency decreased to 93.25%. This decrease was due to the decomposition of excess hydrogen peroxide and the increase of water. As a competitive hydrogen-bonding donor and acceptor, water destroyed the supramolecular structure of DES and reduced the distribution coefficient of sulfone products. When the ratio was further increased to 0.1, the efficiency rose slightly to 97.16%. The decline was attributed to the partial destruction of the hydrogen-bonding network. Therefore, the optimum mass ratio of hydrogen peroxide to model oil was determined to be 0.05.

The effect of temperature on the desulfurization process was shown in Figure 6c. Temperature directly affected the viscosity and interfacial mass transfer efficiency of DES through the thermal response of the hydrogen-bonding network. At 40 °C, citric acid-glycerol DES exhibited a viscosity of up to 12,850 mPa·s. At this time, the rigidity of the three-dimensional hydrogen-bonding network structure was strong. This limited the molecular movement, hindered the dispersion of the oil phase and the DES phase, and inhibited the mass transfer process of DBT and the oxidant. Therefore, the high viscosity reduced the diffusion rate of the oxidation reaction, resulting in a desulfurization rate of only 18.05%. When the temperature rose to 50 °C, the viscosity decreased to 2563 mPa·s. The enhanced thermal motion promoted the reversible cleavage of hydrogen bonds, reduced the rigidity of the network, and improved the fluidity of the DES. This transformation reduced the mass transfer resistance, promoted the synergistic effect between DBT oxidation and sulfone extraction. The desulfurization efficiency was increased to 93.25%. When the temperature reached 60 °C, the viscosity further decreased to 956 mPa·s. Under this condition, the diffusion coefficient of the sulfone molecule reached the maximum, and the mass transfer resistance of the reactants, oxidants, and products at the interface was the smallest. Thus, the desulfurization efficiency was promoted to be close to 100%. These results indicated that a moderate increase in temperature could significantly reduce the viscosity of the solvent. The reduction in viscosity was achieved through the disruption of hydrogen bonds, which enhanced mass transfer and ultimately promoted the synergistic effect between sulfide oxidation and extraction.

The optimized volume ratio of DES to model oil (Figure 6d) showed that when the ratio was 0.5:10, the desulfurization efficiency reached 98.54%. This good performance highlighted the high affinity and strong extraction ability of citric acid-glycerol DES for sulfone compounds. This could provide sufficient hydrogen-bonding sites and capture almost all oxidation products even at a very low addition amount. When the ratio was slightly increased to 1:10, the efficiency was only increased to 98.59%, and the increase was negligible. It was confirmed that the extraction process was close to the thermodynamic limit and the mass transfer tended to be saturated. However, when the amount of DES was further increased to 2:10 and 3:10, the efficiency decreased to 93.25% and 92.25%, respectively. This decrease was mainly attributed to the significant viscosity difference between DES (viscosity of 2563 mPa·s at 50 °C) and model oil (<1 mPa·s). Excessive DES led to an increase in the overall viscosity of the system, which inhibited phase dispersion and droplet breakage under stirring conditions. Thus, the effective interface area and mass transfer coefficient were reduced. In addition, the high viscosity medium limited the diffusion of sulfur oxide molecules in the liquid phase and hindered the dynamic update of the active extraction sites. After achieving a balance between extraction performance and solvent economy, 0.5:10 was selected as the best volume ratio of DES to oil. Operating at this minimum ratio could effectively reduce the industrial operation cost related to solvent recovery.

In summary, the DBT removal efficiency approached 99.95%, under the optimized conditions of citric acid-glycerol DES as the extractant, a H₂O₂:oil mass ratio of 0.05, a temperature of 60 °C and a DES-to-oil volume ratio of 0.5:10. HPLC analysis showed that unoxidized DBT was not detected in the extracted DES-2 phase (Figure S5). This further confirmed that the mechanism of the DES system was oxidation-extraction, rather than simple physical dissolution. According to Figures S7 and S8, the removal rates of DBT by only glycerol, only citric acid and DES-2 were 88.59%, 98.11% and 99.95%, respectively. Although glycerol and citric acid used alone had a certain desulfurization function, DES-2 had the highest removal rate of DBT. This showed that under the same component dosage, glycerol and citric acid did not play a role alone after the formation of DES, but produced a significant synergy, especially conducive to deep desulfurization.

To verify the cycle stability of DES and its applicability in real fuels, Figure 7a,b showed the desulfurization rates of model oil (containing DBT) and real diesel in three cycles, respectively. In the three cycles, the desulfurization rate of the model oil was maintained at 99.95%, 99.30% and 98.50%, respectively. The desulfurization rates of real diesel were 93.72%, 92.86% and 90.80%, respectively. The recovery rates of the three-stage recovery process in the real diesel system were measured to be 97.11%, 90.97% and 85.20%. The reason for the low desulfurization performance of real diesel oil was that its composition is complex: higher viscosity and density reduce the mass transfer efficiency. Meanwhile, the steric hindrance of macromolecules such as polycyclic

aromatic hydrocarbons also hindered the contact between sulfides and DES. Despite the above differences, both systems showed only slight performance degradation after three cycles. This small attenuation was mainly attributed to two factors: one is the high viscosity of DES, which led to the loss of trace solvent in the phase separation process. In addition, the residual trace oxidized sulfide in the regenerated DES gradually accumulates, which affected the subsequent extraction effect. These results indicated that citric acid-glycerol DES exhibited good cycle stability and had broad application potential in the field of industrial continuous desulfurization.

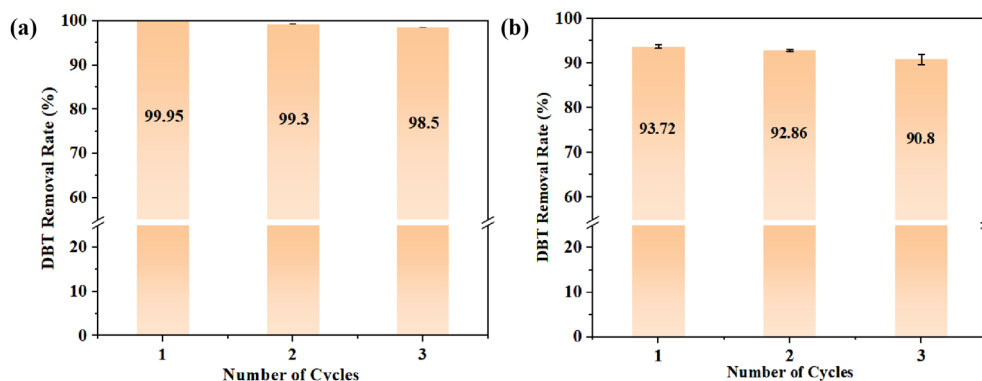


Figure 7. Changes in DBT removal rate after three cycles: (a) model oil; (b) real diesel.

Figure 8 illustrated the comprehensive characterization of DES-2. NMR ^1H and FTIR showed some characteristic functional groups in citric acid and glycerol. Compared with the standard FTIR spectra (Figures S9 and S10), the $\text{C}=\text{O}$ stretching vibration was blue-shifted from 1720 cm^{-1} of citric acid to about 1735 cm^{-1} . In addition, the peak shape of the $\text{C}-\text{O}$ stretching vibration region ($1200\text{--}1000\text{ cm}^{-1}$) was not a simple superposition of the peaks of glycerol and citric acid. These changes clearly indicated that the original hydrogen-bonding network in the raw material was destroyed, and a new DES system was formed. The glass transition temperature T_g was $-41.6\text{ }^\circ\text{C}$. No sharp melting endothermic peak was observed. This network limited the rearrangement of molecules to regular lattices during the cooling process, thereby inhibiting crystallization. When the ratio of glycerol and citric acid was 2:1 (Figure S10), the heat transition temperature of $-41.6\text{ }^\circ\text{C}$ was much lower than the melting point of the single component (153 and $18.1\text{ }^\circ\text{C}$). It retained higher citric acid content while ensuring low T_g and good fluidity. This was the material basis for a good removal rate. Section 3.3 continued to analyze the mechanism from the perspective of intermolecular non-covalent interactions.

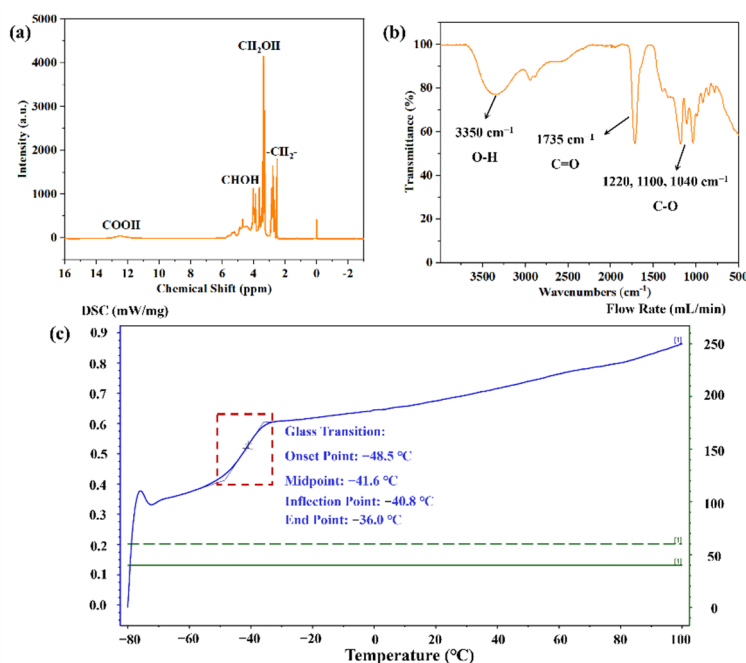


Figure 8. Comprehensive characterization of DES-2 (glycerol to citric acid molar ratio of 2:1): (a) NMR ^1H spectrum, (b) FTIR spectrum, and (c) DSC heating curves.

3.3. Quantum Chemical Insights into Intermolecular Interactions

Figure 9a showed the ESP surface and its distribution of DBT and its oxidized derivatives DBTO and DBTO₂. Figure 9b showed the ESP distribution of each component (citric acid and glycerol) of DES-2 in detail. The blue, red, and white regions represented negative, positive, and near-neutral electrostatic potentials, respectively. Before oxidation, the polarity of thiophene sulfur atoms in DBT was weak, and the surface electrostatic potential was mainly concentrated near the neutral value (minimum negative electrostatic potential $E_{\min} = -14.74$ kcal/mol). This reflected that its tendency to form medium to strong hydrogen bonds was negligible. After oxidation to DBTO and DBTO₂, the $E_{\min}$ values decreased sharply to -49.06 and -40.77 kcal/mol, respectively. This change indicated that the sulfoxide group had a stronger ability to act as a hydrogen-bonding acceptor than the thiophene group.

The citric acid molecule contains three carboxyl groups and one hydroxyl group. These functional groups intertwine with the glycerol molecule to form a stable three-dimensional hydrogen-bonding network. The network underlies the excellent performance of the citric acid-glycerol system. This structure endowed the system with multi-site recognition ability. Quantitative analysis of ESP (Figure 9b) confirmed that the minimum negative ESPs ($E_{\min}$) of citric acid and glycerol were -32.16 and -28.42 kcal/mol, respectively. Their maximum positive ESPs ($E_{\max}$) were 56.89 and 48.36 kcal/mol, respectively. These results highlighted the dual roles of both molecules as appropriate hydrogen-bonding donors and acceptors.

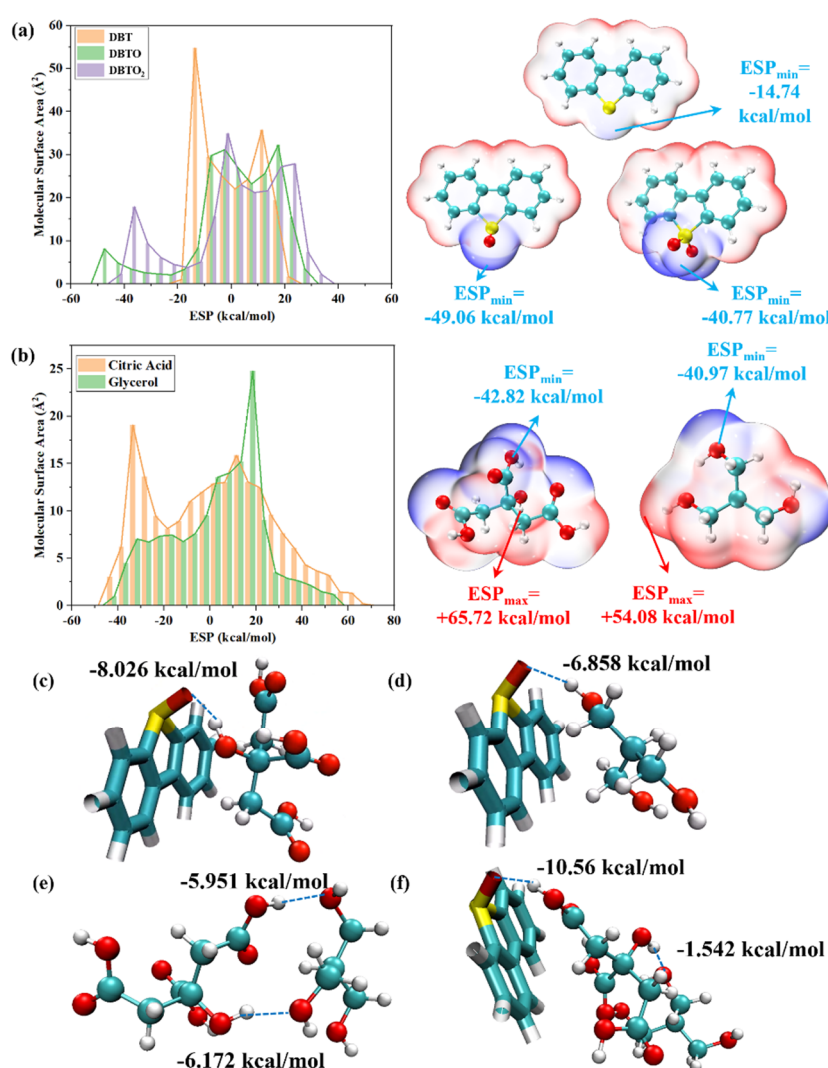


Figure 9. Analysis of non-covalent interactions in the target system: (a) Molecular surface ESP distributions of DBT before and after oxidation. (b) Molecular surface ESP distributions for the two components of DES-2. (c) Hydrogen-bonding energy in the extraction of DBTO₂ by citric acid. (d) Hydrogen-bonding energy in the extraction of DBTO₂ by glycerol. (e) Internal hydrogen-bonding energy within the DES-2 eutectic. (f) Hydrogen-bonding energy in the extraction of DBTO₂ by DES-2.

DFT calculations showed that the hydrogen-bonding energies of DBTO₂ with citric acid and glycerol were −6.053 and −5.154 kcal/mol, respectively (Figure 9c,d). The hydrogen-bonding energy in DES-2 ranged from −5.951 to −6.172 kcal/mol (Figure 9e). In the ternary system of citric acid, glycerol, and DBTO₂, the optimized configuration showed that the hydrogen-bonding energy of DES-DBTO₂ was −6.368 kcal/mol (Figure 9f). The hydrogen-bonding energy of the ternary system was slightly higher than that of the single DES component, but the difference was not significant. Besides, the enhancement of the hydrogen bond between DBTO₂ and DES was far less than that of DBTO (Figure S12). This was because the polarity directions of the two S=O bonds of sulfone were similar and symmetrically distributed. These polarities were partially offset during vector superposition. Consequently, the overall polarity of sulfone is significantly weaker than that of sulfoxide.

Therefore, it was not sufficient to evaluate the capture ability of DES to DBTO₂ only by hydrogen-bonding energy. This study supplemented the SobEDAw energy decomposition analysis, and the results were shown in Table 3. The total interaction energy of the DES-2 two-component with the DBTO₂ was −31.24 kcal/mol. This was significantly larger than the absolute value of the interaction energy (−23.33 and −17.31 kcal/mol) of any single DES-2 component. It was shown that the two-component DES system formed a structure with stronger overall binding ability to sulfone through synergistic reconstruction between components. For DBTO₂, the electrostatic contribution (−35.24 kcal/mol) was comparable to the dispersion contribution (−33.91 kcal/mol). Together, they provided a strong attraction term sufficient to overcome the exchange repulsion (+54.47 kcal/mol) caused by the Pauli exclusion principle, so that the molecular complex could be stabilized.

Table 3. Energy decomposition analysis between DBTO₂ and DES-2 components.

Target Molecule	DES-2 Molecule	ΔE (kcal/mol)	ΔE_{els} (kcal/mol)	ΔE_{xrep} (kcal/mol)	ΔE_{orb} (kcal/mol)	ΔE_{disp} (kcal/mol)
DBTO ₂	Citric acid	−23.32	−23.89	37.51	−13.12	−23.84
DBTO ₂	Glycerol	−17.31	−21.97	37.19	−10.07	−22.46
DBTO ₂	Citric acid/ Glycerol	−31.24	−35.24	54.47	−16.56	−33.91

4. Conclusions

This study demonstrates a machine learning framework (PDI-DMPNN+SME) for the rapid, accurate discovery of DESs. By coupling high-precision solvation free energy predictions ($R^2 > 0.999$) with atomic-group substructure masking, the algorithm evaluates both solute extraction selectivity and solvent eutectic stability. Virtual screening of 19,291 binary candidates identified a citric acid-glycerol DES system as a promising green extractant candidate. Under optimized conditions, this system achieved an exceptional DBT removal efficiency of 99.95% in model oil and 93.72% in real diesel, with satisfactory cyclic reusability over the tested cycles. Quantum chemical analysis revealed that the high performance of this DES stems from synergistic non-covalent interactions between the two DES components and the DBTO₂. Electrostatic and dispersion contributions were found to be comparable, jointly overcoming Pauli repulsion to stabilize the extractant-solute complex, with a total interaction energy of −31.24 kcal/mol. Overall, this methodology bridges theoretical machine learning models with functional solvent synthesis, offering a design paradigm for complex liquid separation systems.

Supplementary Materials

The additional data and information can be downloaded at: <https://media.sciltp.com/articles/others/2609281452183583/SCE-26070289-SI.pdf>. Table S1: The main mathematical definitions and full names. Table S2: Hydrogen bond acceptor and donor components of the binary DES candidates. Table S3: Environmental assessment of Component-2 candidates. Figure S1: Substructures of aspirin generated via the BRICS method. Figure S2: Target various oxidized thiophenic sulfur compounds. Figure S3: Prediction accuracy of PDI-DMPNN evaluated at 30 epochs varies with the introduction of different prior data types. Figure S4: Comparative parity plots of DMPNN and PDI-DMPNN. Figure S5: HPLC detected the spectrum of DBT in DES-2 phase after OEDS. Figure S6: Oil phase HPLC spectrum after OEDS using glycerol. Figure S7: Oil phase HPLC spectrum after OEDS using citric acid. Figure S8: Standard FTIR spectrum of glycerol. Figure S9: Standard FTIR spectrum of citric acid. Figure S10: Change of thermal transition temperature with the molar ratio of glycerol to citric acid. Figure S11. TGA curve of DES-2 (glycerol to citric acid molar ratio of 2:1). Figure S12: (a) Hydrogen-bonding energy in the extraction of DBTO by citric acid. (b) Hydrogen-bonding energy in the extraction of DBTO by glycerol.

Author Contributions

C.H. is responsible for methodology, software, writing-original draft. J.X., A.G., Y.X. and X.L. perform validation and investigation. L.H. and H.S. contribute to project administration, methodology, writing-review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

We acknowledge the financial support from National Natural Science Foundation of China (NSFC, Grant No. 22178252). Furthermore, the experimental and computational work in this research run on the Huawei Cloud AI Compute Service. We appreciate the stable compute supply from this platform.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Custom scripts applied in this study are publicly available on GitHub at <https://github.com/hechangqing1997/PDI-DMPNN-SME> (access on 27 September 2026).

Conflicts of Interest

The authors declare that the source code described in this paper has been applied for software copyright registration (2025SR0349859), and there are no other potential competing financial interests.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

Abbreviations

BRICS	Breaking of retrosynthetically interesting chemical substructures
DBT	Dibenzothiophene
DBTO	Dibenzothiophene sulfoxide
DBTO ₂	Dibenzothiophene sulfone
DFT	Density functional theory
ESP	Electrostatic potential
MPI	Molecular polarity index
OEDS	Oxidative-extractive desulfurization
PDI-DMPNN	Prior data introducing directed message passing neural network
DES	Deep eutectic solvent
SME	Substructure mask explanation

References

1. Tanimu, A.; Alhooshani, K. Advanced hydrodesulfurization catalysts: A review of design and synthesis. *Energy Fuels* **2019**, *33*, 2810–2838. <https://doi.org/10.1021/acs.energyfuels.9b00354>.
2. Macías, M.J.; Ancheyta, J. Simulation of an isothermal hydrodesulfurization small reactor with different catalyst particle shapes. *Catal. Today* **2004**, *98*, 243–252. <https://doi.org/10.1016/j.cattod.2004.07.038>.
3. Cheng, F.; Chen, J.; Gou, X.L. MoS₂-Ni Nanocomposites as Catalysts for Hydrodesulfurization of Thiophene and Thiophene Derivatives. *Adv. Mater.* **2006**, *18*, 2561–2564. <https://doi.org/10.1002/adma.200600912>.
4. Lü, H.; Li, P.; Deng, C.; et al. Deep catalytic oxidative desulfurization (ODS) of dibenzothiophene (DBT) with oxalate-based deep eutectic solvents (DESSs). *Chem. Commun.* **2015**, *51*, 10703–10706. <https://doi.org/10.1039/c5cc03324a>.
5. Yin, J.; Wang, J.; Li, Z.; et al. Deep desulfurization of fuels based on an oxidation/extraction process with acidic deep eutectic solvents. *Green Chem.* **2015**, *17*, 4552–4559. <https://doi.org/10.1039/c5gc00709g>.
6. El Achkar, T.; Greige-Gerges, H.; Fourmentin, S. Basics and properties of deep eutectic solvents: A review. *Environ. Chem. Lett.* **2021**, *19*, 3397–3408. <https://doi.org/10.1007/s10311-021-01225-8>.

7. Hansen, B.B.; Spittle, S.; Chen, B.; et al. Deep eutectic solvents: A review of fundamentals and applications. *Chem. Rev.* **2020**, *121*, 1232–1285. <https://doi.org/10.1021/acs.chemrev.0c00385>.
8. Hao, L.; Wang, M.; Shan, W.; et al. L-proline-based deep eutectic solvents (DESSs) for deep catalytic oxidative desulfurization (ODS) of diesel. *J. Hazard. Mater.* **2017**, *339*, 216–222. <https://doi.org/10.1016/j.jhazmat.2017.06.050>.
9. Lima, F.; Gouvenaux, J.; Branco, L.C.; et al. Towards a sulfur clean fuel: Deep extraction of thiophene and dibenzothiophene using polyethylene glycol-based deep eutectic solvents. *Fuel* **2018**, *234*, 414–421. <https://doi.org/10.1016/j.fuel.2018.07.043>.
10. Li, C.; Li, D.; Zou, S.; et al. Extraction desulfurization process of fuels with ammonium-based deep eutectic solvents. *Green Chem.* **2013**, *15*, 2793–2799. <https://doi.org/10.1039/c3gc41067f>.
11. Li, J.J.; Xiao, H.; Tang, X.D.; et al. Green carboxylic acid-based deep eutectic solvents as solvents for extractive desulfurization. *Energy Fuels* **2016**, *30*, 5411–5418. <https://doi.org/10.1021/acs.energyfuels.6b00471>.
12. Hadj-Kali, M.K.; Mulyono, S.; Hizaddin, H.F.; et al. Removal of Thiophene from Mixtures with *n*-Heptane by Selective Extraction Using Deep Eutectic Solvents. *Ind. Eng. Chem. Res.* **2016**, *55*, 8415–8423. <https://doi.org/10.1021/acs.iecr.6b01654>.
13. Mao, C.; Zhao, R.; Li, X.; et al. Trifluoromethanesulfonic acid-based DESSs as extractants and catalysts for removal of DBT from model oil. *RSC Adv.* **2017**, *7*, 12805–12811. <https://doi.org/10.1039/c6ra28448e>.
14. Li, C.; Zhang, J.; Li, Z.; et al. Extraction desulfurization of fuels with ‘metal ions’ based deep eutectic solvents (MDESSs). *Green Chem.* **2016**, *18*, 3789–3795. <https://doi.org/10.1039/c6gc00366d>.
15. López-Flores, F.J.; Ramírez-Márquez, C.; González-Campos, J.B.; et al. Machine learning for predicting and optimizing physicochemical properties of deep eutectic solvents: Review and perspectives. *Ind. Eng. Chem. Res.* **2025**, *64*, 3103–3117. <https://doi.org/10.1021/acs.iecr.4c03610>.
16. Mohan, M.; Jetli, K.D.; Smith, M.D.; et al. Accurate machine learning for predicting the viscosities of deep eutectic solvents. *J. Chem. Theory Comput.* **2024**, *20*, 3911–3926. <https://doi.org/10.1021/acs.jctc.3c01163>.
17. Liu, X.; Chen, J.; Qiu, Y.; et al. Machine learning boosted eutectic solvent design for CO₂ capture with experimental validation. *AIChE J.* **2025**, *71*, e18631.
18. Odegova, V.; Lavrinenko, A.; Rakhmanov, T.; et al. DESignSolvents: An open platform for the search and prediction of the physicochemical properties of deep eutectic solvents. *Green Chem.* **2024**, *26*, 3958–3967. <https://doi.org/10.1039/d3gc04533a>.
19. Butler, K.T.; Davies, D.W.; Cartwright, H.; et al. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555. <https://doi.org/10.1038/s41586-018-0337-2>.
20. Omar, K.A.; Sadeghi, R. Database of deep eutectic solvents and their physical properties: A review. *J. Mol. Liq.* **2023**, *384*, 121899. <https://doi.org/10.1016/j.molliq.2023.121899>.
21. Liu, Z.; Lu, T.; Chen, Q. Intermolecular interaction characteristics of the all-carboatomic ring, cyclo [18]carbon: Focusing on molecular adsorption and stacking. *Carbon* **2021**, *171*, 514–523. <https://doi.org/10.1016/j.carbon.2020.09.048>.
22. Murray, J.S.; Politzer, P. The electrostatic potential: An overview. *WIREs Comput. Mol. Sci.* **2011**, *1*, 153–163. <https://doi.org/10.1002/wcms.19>.
23. Vermeire, F.H.; Green, W.H. Transfer learning for solvation free energies: From quantum chemistry to experiments. *Chem. Eng. J.* **2021**, *418*, 129307. <https://doi.org/10.1016/j.cej.2021.129307>.
24. Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; et al. On the Art of Compiling and Using ‘Drug-Like’ Chemical Fragment Spaces. *ChemMedChem* **2008**, *3*, 1503–1507. <https://doi.org/10.1002/cmdc.200800178>.
25. Wu, Z.; Wang, J.; Du, H.; et al. Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking. *Nat. Commun.* **2023**, *14*, 2585. <https://doi.org/10.1038/s41467-023-38192-3>.
26. He, C.; Xue, Y.; Zhou, H.; et al. General applicability for developing interfacial regulation materials: An artificial-intelligence-driven approach based on non-covalent interactions. *Chem. Eng. J.* **2026**, *547*, 179552. <https://doi.org/10.1016/j.cej.2026.179552>.
27. Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592. <https://doi.org/10.1002/jcc.22885>.
28. Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671. <https://doi.org/10.1021/acs.jctc.8b01176>.
29. Emamian, S.; Lu, T.; Kruse, H.; et al. Exploring nature and predicting strength of hydrogen bonds: A correlation analysis between atoms-in-molecules descriptors, binding energies, and energy components of symmetry-adapted perturbation theory. *J. Comput. Chem.* **2019**, *40*, 2868–2881. <https://doi.org/10.1002/jcc.26068>.
30. Lu, T.; Chen, Q. Simple, efficient, and universal energy decomposition analysis method based on dispersion-corrected density functional theory. *J. Phys. Chem. A* **2023**, *127*, 7023–7035. <https://doi.org/10.1021/acs.jpca.3c04374>.