International Core Journal of Engineering Volume 12 Issue 6, 2026
ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

Data-Driven Development of Deep Eutectic Solvents: A Review of Machine Learning in Property Prediction and Rational Design

Zhen Yang, Xingchi Deng, Jin Shu, Yichi Zhang, and Shengtao He

North China University of Science and Technology, Tangshan 063000, China

Abstract

Deep Eutectic Solvents (DESs) represent a novel class of emerging green solvents formed through hydrogen-bonding interactions between hydrogen bond donors (HBDs) and hydrogen bond acceptors (HBAs). Boasting core advantages such as facile synthesis, low cost, environmental benignity, and exceptional structural and functional designability, DESs have demonstrated substantial application value and industrial potential across various sectors, including natural product extraction, biomass valorization, hydrometallurgy, acidic gas capture, and electrochemical energy storage. However, several fundamental scientific and technical bottlenecks-including the combinatorial explosion of the chemical design space arising from vast HBD/HBA combinations, the elusive structure-property relationships between microscopic intermolecular interactions and macroscopic physicochemical properties, and the challenges in synergistically optimizing multiple performance indicators in complex application scenarios-have rendered traditional "trial-and-error" R&D models inadequate for the targeted rational design of DESs, thereby significantly impeding their large-scale industrialization. As a data-driven intelligent research paradigm, machine learning (ML) provides a transformative technical route to address these challenges by leveraging its robust capabilities in high-dimensional feature mining, complex non-linear relationship fitting, and global optimization. This review systematically delineates the latest domestic and international advancements in ML-enabled DES research. First, it elucidates the core principles and mainstream methodologies for molecular descriptor selection and high-quality dataset construction in DES property prediction. Second, it summarizes model architectures, algorithmic optimizations, and predictive performance of ML regarding the thermophysical properties as well as chemical and biological activities of DESs. Subsequently, it highlights technical implementations and application outcomes of ML in the intelligent prediction of DES formation, application-oriented reverse design, and multi-objective optimization of extraction and separation processes. Finally, the review provides an in-depth analysis of the core challenges currently facing the field-such as the lack of standardized datasets, insufficient model interpretability, and weak mechanistic insights-while offering perspectives on future research directions. This review aims to provide a systematic theoretical framework and research roadmap for the intelligent rational design, precise property modulation, and industrial application of DESs, facilitating the paradigm shift in green solvent development from traditional empirical trial-and-error toward a data-driven intelligent paradigm.

Keywords

Deep Eutectic Solvents (DES); Machine Learning; Property Prediction; Inverse Design; Multi-objective Optimization; Data-driven.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

1. Introduction

Deep eutectic solvents (DES) are considered a new type of green solvent due to their unique properties, such as ease of synthesis, low cost, environmental friendliness, low volatility, high solubility , high biodegradability and structural design feasibility. Eutectic mixtures are eutectic mixtures formed between compounds , whose eutectic temperature is much lower than the melting enthalpy of pure compounds predicted by ideal solution theory . They usually originate from complex hydrogen bond interactions [1] [2] . The basic composition and formation mechanism of DES were first introduced by Abbott et al. in 2002 through the preparation of eutectic mixtures of choline chloride (ChCl) and urea. This system is liquid at room temperature, has a melting point much lower than that of any component, and exhibits excellent solvent properties [3] . This pioneering work laid the foundation for DES research.

DES are usually composed of hydrogen bond donors (HBDs) and hydrogen bond acceptors (HBAs). HBDs include amides, thioureas, amines, imidazoles, alcohols, acids and phenols. HBAs include quaternary ammonium salts, quaternary phosphate salts, imidazole-based salts, salts based on dihydrogen bond donor - dihydrogen bond acceptor pairings, inner salts, and molecular imidazoles and their analogues [2] . A large number of DESs can be formed by combining the above hydrogen bond donors and hydrogen bond acceptors in various different ratios, and they can be applied to different research and production according to the different properties of DESs . As a substitute for traditional solvents, eutectic solvents have been widely used in various fields such as extraction and separation, biomass conversion, acid gas capture, and electrochemistry.

Eutectic solvents allow for targeted extraction of plant active ingredients, and the extraction of components with different polarities can be achieved by adjusting the composition ratio , as clearly demonstrated in several comparative studies . Wu et al. extracted carotenoids from tomatoes using supercritical fluid extraction. The DES prepared by methyl 1-benzylpiperidin-4-carboxylic acid:acetic acid:heptanoic acid (2:1:1) showed excellent performance, with yields of 408.85±4.97 μg/g lycopene and 68.78±2.00 μg/g carotene, respectively [4] . Wang et al. used betaine-based DES to extract active compounds from peony petals, verifying the selective extraction of components by eutectic solvents. Betaine was used as HBA, and 1,2-propanediol and lauric acid were used as HBDs. When betaine -1,2 -propanediol was used as the extraction solvent, the highest extraction rates of total flavonoids and total anthocyanins were achieved, at 61.65 mg RE/g and 2.15 mg C3GE/g, respectively. When betaine - lauric acid was used as the extraction solvent, the highest total phenol extraction rate was achieved, which was 321.59 mg GAE/g [5] . The separation and extraction of lignin is complex and its mechanism of action is still unclear. Multiple studies have shown that DES can effectively destroy the complex hydrogen bond network and crystal structure in lignocellulose and is used to efficiently separate cellulose, hemicellulose and lignin, and then prepare nanocellulose or platform compounds [6] [7] [8] .

In the field of metal smelting, in order to avoid the equipment corrosion problems caused by traditional acid leaching or alkaline leaching , and in order to reduce costs and reduce environmental pollution, the application of eutectic solvents to hydrometallurgy has become a new research focus [9] . Zhang Yuan used choline chloride-urea and choline chloride-oxalic acid hydrate eutectic solvents to leach zinc in zinc oxide ore. For the choline chloride-oxalic acid hydrate DES leaching agent with a molar ratio of 1:1, the zinc leaching rate can reach more than 90% by controlling the optimal reaction conditions [10] . Nerea's team constructed a eutectic solvent by combining toluenesulfonic acid-hydrate with choline chloride , and used it in molar ratios of 2:1, 1:1 and 1:2 to conduct dissolution experiments on metal oxides such as MnO, Fe2O3, Cu2O, and ZnO. They found that when the molar ratio was 1:2 , these metal oxides all showed high solubility [11] .

Given that traditional lithium-ion battery electrolytes are flammable, toxic and environmentally risky, DES -based electrolytes are considered a promising alternative due to their non-flammability, low volatility, wide electrochemical window and good ionic conductivity. Multiple studies have pointed

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

out that DES can not only improve battery safety, but also stabilize the electrode/electrolyte interface through its unique solvation structure, suppress side reactions and thus extend battery cycle life [12] .

Recent studies have shown that eutectic solvents can also be used as catalysts to efficiently catalyze the synthesis of chemical raw materials under mild conditions. Wang Sisi et al. used tetrabutylammonium bromide (TBAB) as a hydrogen bond acceptor and p-toluenesulfonic acid (PTSA) as a hydrogen bond donor to prepare eutectic solvents for the catalytic synthesis of ethylene glycol monoacetate by heating and stirring. Under optimal reaction conditions, the catalyst prepared with a TBAB/PTSA molar ratio of 1:3 achieved an acetic acid conversion rate of 88% and an EGMA selectivity of 94%. After being reused 5 times, the acetic acid conversion rate was 81% and the EGMA selectivity reached 93%, still exhibiting high reactivity and stability [13] .

The large-scale use of fossil energy has led to a continuous increase in the CO2 content in the atmosphere, and efficient and environmentally friendly CO2 capture technology has become a research hotspot. Liu Jiaxing et al. prepared a linear eutectic solvent by using glycerol as a hydrogen bond donor and octyl ethylenediamine (Octen) or 2-ethylhexyl ethylenediamine (EtHexen) as a hydrogen bond acceptor. They studied its CO2 absorption performance through density functional theory calculations and molecular dynamics simulations. The CO2 absorption rates of Glycerol-EtHexen and Glycerol-Octen reached 70.5% and 63.5%, respectively, providing a new idea for green, efficient and low-cost CO2 capture [14] [15] .

Eutectic solvents, as novel green solvents, have been widely used in many research and production processes. However, due to the special nature of their preparation, the number of eutectic solvents with different combinations of HBA and HBD and different concentration ratios is very large. Different combinations of DES have different properties, and how to design or select reasonable eutectic solvents in various fields of research and production has become a major problem [16] [17] . In the face of the problem of redundancy in the chemical design space of eutectic solvents, namely the screening dilemma brought about by the combination of HBD and HBA, computational simulation and machine learning have become key tools for solving the problem . Uzochukwu et al. studied the ionic interaction and hydrogen bond network between HBA and HBD in the choline chloride and glycerol system through molecular modeling , and established a computational framework for predicting the thermodynamic feasibility of new DES, which effectively reduced the huge chemical search space [18] . Abdollahzadeh et al. used seven supervised machine learning algorithms to predict the density of 14 DES with high accuracy based on only a small amount of experimental data, which proved the great potential of data-driven methods in rapidly screening DES formulations that meet specific physical property requirements [19] . Secondly, the design of eutectic solvents is complex in terms of structure-property relationships, that is, there is a lack of research on how micro-interactions determine macro-properties. Yang et al. used tetrabutylammonium chloride (TBAC) as HBA and systematically changed the molar fractions of two HBDs, decanoic acid (DEC) and octanoic acid (OC). They found that small changes in carbon chain length and ratio significantly modulate hydrogen bond strength, thereby affecting the macro-physicochemical properties of DES, and for the first time revealed the possibility of adjusting performance at the molecular level [20] . At the same time, when applying eutectic solvents to production and daily life , the dilemma of multi-objective optimization is particularly prominent in specific applications . That is, in practical applications, it is often necessary to balance multiple properties of DES, which makes it costly to screen DES with suitable physical properties in production research. For example, Xiao et al. pointed out in their review of DES for the separation of lignocellulose components that an ideal DES should have efficient destructive ability, moderate viscosity and good chemical stability [7] . Wang et al. focused on the preparation of nanocellulose by acidic DES (Acid Deep Eutectic Solvents , ADES), and further emphasized that the optimal ADES not only needs sufficient acidity to catalyze hydrolysis, but also needs to maintain a low viscosity to promote the diffusion of reactants and ensure the recyclability of the solvent itself, further highlighting the necessity and complexity of multi-objective synergistic optimization [21] .

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

In summary, DES has attracted great attention in recent years due to its highly tunable properties and the difficulty in predicting its physical properties. However, traditional experimental screening methods are inefficient and resource-intensive. Compared with traditional prediction methods, artificial intelligence can provide faster and more accurate data prediction and processing, and has become an important tool for scientific research. Machine learning (ML) is the core of artificial intelligence. It is a powerful tool for finding rules in high-dimensional data and has been applied in fields such as chemical engineering, advanced materials, and medicine [17] . In order to reduce the design difficulty, reduce the cost of DES screening and maximize the performance of DES, this paper summarizes the application of machine learning models for DES, including solubility prediction, physicochemical property prediction and DES design and screening. At the same time, the application of machine learning in the chemical industry is still in its early stages, and there are still many difficulties and problems to be explored . The article concludes by summarizing the shortcomings of applying machine learning to the design and property prediction of eutectic solvents, pointing the way for further scientific research.

2. Machine Learning Enables DES Property Prediction

Machine learning, as an interdisciplinary field that has emerged rapidly in the last two decades, has demonstrated strong application potential by using statistical algorithms to extract knowledge from massive amounts of historical data, build experience models, and then guide various business decisions [22] . For example, Choi et al. have successfully applied machine learning technology to accurately predict and effectively evaluate the optimal extraction conditions for water coconut [23] . Traditional machine learning research covers multiple directions, including decision trees, random forests, artificial neural networks, and support vector machines. These algorithms and technologies have achieved significant results in different fields [23] [24].

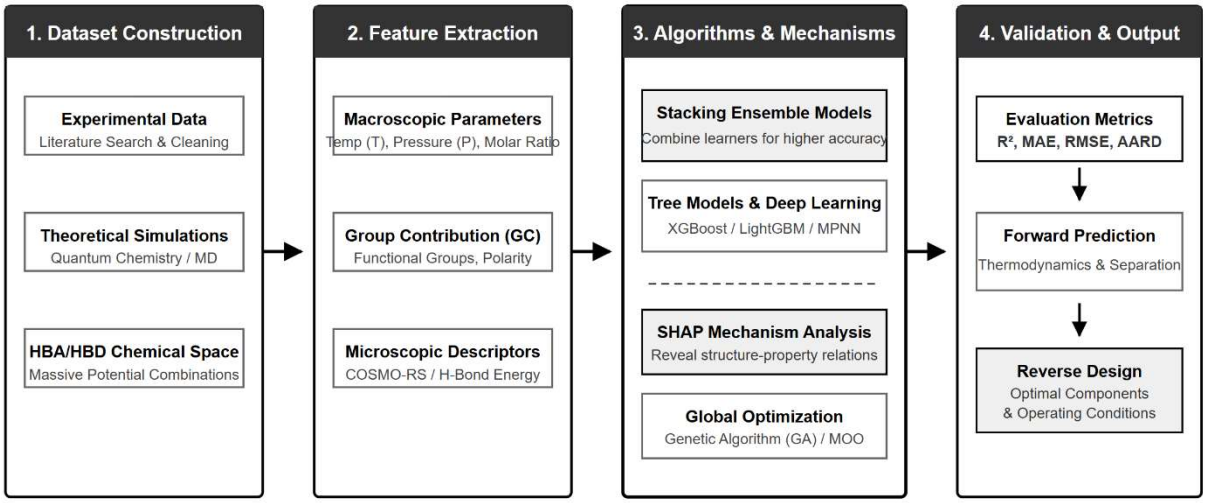

Fig. 1 Machine Learning Workflow for Property Prediction and Directional Design of DES

The physicochemical properties and chemical/biological activities of DES constitute the core foundation for its application in extraction and separation, biocatalysis, carbon capture and other fields. Accurate and rapid prediction of various properties of DES is the key prerequisite for realizing its directional rational design. Traditional experimental measurement methods have the disadvantages of being time-consuming, laborious and costly, which makes it difficult to meet the actual needs of large-scale DES system screening [25] [26] . Machine learning technology, by constructing a quantitative prediction model between input features and target properties, can significantly reduce the workload of experiments, improve the efficiency and accuracy of DES property research, and is the core technical support for promoting the transformation of DES research from traditional

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

experience exploration to modern rational design [25] . In order to intuitively demonstrate this data-driven intelligent research paradigm, this paper systematically summarizes the whole process of DES property prediction and directional design based on machine learning , as shown in Figure 1. This process covers the construction of high-quality datasets, extraction of multi-dimensional feature descriptors, construction of core algorithm models and mechanism analysis, and finally realizes the complete closed loop of target property prediction and reverse design. The following chapters will elaborate on this main line.

2.1 DES Molecular Descriptor and Dataset Construction

As the core input feature of machine learning models, the scientific selection and reasonable construction of molecular descriptors directly affect the final prediction performance of the model. High-quality datasets are the important foundation for model training, validation and optimization. Based on the system composition and structural characteristics of DES, the input features currently used for machine learning model construction are mainly divided into three categories: experimental parameters, theoretical descriptors, molecular fingerprints and group contributions [25] :

The main parameters include the composition ratio, test temperature, and pressure of DES. These parameters can directly reflect the regulatory effect of experimental conditions on the properties of DES . Abbott et al. laid the foundation for DES research in 2004 and systematically recorded basic experimental parameters including temperature, viscosity and conductivity, providing a benchmark for subsequent property modeling [27] . Liu et al. further combined experimental parameters (such as composition ratio, test temperature, and pressure) with the UNIFAC model in 2025, realizing high-throughput screening of aromatic extraction processes and verifying the predictive ability of theoretical models [28] .

Covering COSMO-RS parameters and quantum chemical parameters (bond length, bond angle, molecular dipole moment, etc.), the structural features and intermolecular interaction rules of DES can be revealed from the molecular microscopic level. Among them, the σ-profile descriptor derived from COSMO-RS quantifies the molecular surface charge density into the S_1 - S_{10} region , which has become an important technological advancement in DES machine learning feature engineering [29] [30] . In addition, Liu et al. (2025) used the UNIFAC method to perform high-throughput screening and successfully predicted the component ratio and phase equilibrium behavior of DES in the aromatic extraction process of coking gasoline [28] .

Molecular fingerprints encode molecular structural information to reflect its core features. Among them, extended connection fingerprint (ECFP) can efficiently encode molecular substructure information. Le et al. (2020) achieved a breakthrough in reverse deducing molecular structure from ECFP fingerprints using the Neuraldecipher method, further improving the interpretability of fingerprint features [30] . Klamt et al. (1975) proposed the UNIFAC group contribution method to predict the overall properties of the system based on the functional group characteristics of each component of DES. Both provide important structural feature support for the prediction of DES properties [31] .

In the field of dataset construction, researchers mainly construct dedicated eutectic solvent (DES) datasets suitable for predicting different physical properties by integrating experimental data from existing literature, conducting independent experimental tests, and combining molecular simulation data. For example, researchers collected 1410 DES density data points from publicly available literature using bibliometric methods and constructed a database containing 166 DES at different temperatures and pressures [32] , providing high-quality, high-dimensional data support for the construction of machine learning models.

2.2 ML Prediction of Thermophysical Properties

Significant progress has been made in both domestic and international research regarding dataset construction and machine learning model application for predicting the thermal properties of eutectic solvents (DES). To visually illustrate the application strategies of different machine learning models

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

in DES property prediction in recent years, Table 1 systematically summarizes the model construction parameters and performance evaluation metrics of representative studies. When evaluating model accuracy, many existing studies comprehensively consider metrics such as the coefficient of determination (R^2), mean absolute relative deviation (AARD), mean squared error (MSE), and root mean square error (RMSE) to ensure the generalization ability and reliability of the prediction model. The following will discuss this in detail with specific physical properties:

Table 1. Machine learning-based prediction model for typical thermophysical properties of DES

Target nature	Algorithm	Input features	Dataset size	Performance validation metrics and model performance	References
density	Group Contribution (GC) Method + Multilayer Perceptron (MLP)	35 functional groups, temperature	1410 data points	AARD < 0.56%, R^2 = 0.99	[39]
viscosity	Message Passing Neural Network (MPNN) + Graph Attention Mechanism (GAT)	Molecular energy, molecular weight of heavy atoms, etc.	166 types of DES (5790 or 4423 data points)	R^2 = 0.9945, revealing the key influence of energy and molecular weight.	[40] [42]
Heat capacity	Artificial Neural Networks (ANN), CNNs, and Long Short-Term Memory Networks (LSTM)	DES structural characteristics and experimental conditions	Approximately 2,696 data points (covering 55 DES systems)	An interpretable AI model was constructed to reveal key structural features affecting heat capacity.	[35]
Speed of sound	Hybrid machine learning prediction framework	High-dimensional feature vectors and computational chemical descriptors based on molecular signatures	Approximately 415 data points (covering 38 types of DES)	It enables rapid and accurate prediction of sound speed and elucidates microscopic structure-property relationships.	[37]

As the core parameter for energy transfer and process design of eutectic solvent (DES), accurate prediction of heat capacity is of great significance for the industrial application and process optimization of DES. Heat capacity prediction models constructed based on machine learning algorithms such as artificial neural networks (ANN), convolutional neural networks (CNN), and long short-term memory networks (LSTM) can fully explore the complex nonlinear relationship between input features and heat capacity, and show excellent prediction performance. Some studies have also constructed interpretable artificial intelligence models, which, while achieving accurate prediction of DES heat capacity, further reveal the key structural features and experimental conditions affecting DES heat capacity [33] [34] , providing theoretical support for the targeted control of DES heat capacity.

In the study of the thermophysical properties of deep eutectic solvents (DES), sound velocity, as a key indicator reflecting intermolecular interactions and microstructural features, plays a crucial role in the interpretation of structure-property relationships at the microscopic level. To overcome the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

limitations of traditional experimental methods in rapidly acquiring sound velocity data, researchers have constructed a hybrid machine learning prediction framework based on **molecular signatures** . This framework successfully achieves rapid and accurate prediction of DES sound velocity by encoding the molecular structural features of DES into high-dimensional feature vectors and combining them with high-precision computational chemical descriptors. This achievement not only provides a new perspective for deconstructing the structure-property relationship between DES structure and sound velocity at the molecular level, but also lays an important reference methodology for predicting other thermophysical properties that are highly dependent on microstructure (such as thermal conductivity, sound velocity dispersion, etc.) [35] [36] .

Accurate density prediction is also crucial for mixing and separation design in industrial processes. Researchers constructed a large database containing 1410 density data points and combined the group contribution (GC) method and multilayer perceptron (MLP) models to successfully reduce the mean absolute relative deviation (AARD) to below 0.56% and achieve a high fitting degree of R^2=0.99. These models not only improved the prediction accuracy, but also improved the applicability to different DES systems through feature engineering (such as considering 35 functional groups and temperature factors) [37] .

Viscosity is a core parameter that restricts the industrial application of DES. Traditional experimental methods for measuring viscosity are time-consuming, laborious, and difficult to cover a large-scale composition space. In recent years, researchers have significantly improved the prediction accuracy of machine learning models by constructing large-scale, high-quality datasets (such as collecting 5790 sample points or 4423 data points). For example, a study used a multi-scale prediction framework based on message passing neural network (MPNN) and graph attention mechanism (GAT) to achieve a high-precision prediction of R^2=0.9945 on a database containing 166 DES, and revealed the key influence of molecular energy and heavy atom molecular weight on viscosity through SHAP analysis [38] [39] [40] .

2.3 ML Prediction of Chemical and Biological Activities

In the development of applications for eutectic solvents (DES), machine learning (ML) technology is becoming a key bridge connecting molecular structure and functional properties, showing significant advantages, especially in solute solubility prediction and enzyme activity assessment.

Solubility is a core metric for evaluating DES as an extractant or separation medium. Studies have shown that by integrating molecular descriptors (such as σ values calculated by COSMO-RS, molecular polarity, etc.) with machine learning models, highly efficient solvents can be rapidly screened without conducting expensive experiments.

Solubility prediction model: Based on deep learning and nuclear magnetic resonance data, the study successfully predicted the solubility of multiple active pharmaceutical ingredients (APIs) in different DES, which confirmed the practicality of machine learning in high-throughput screening [41] [42] .

CO_2 capture: For carbon capture applications, machine learning models can not only predict the solubility of CO_2 in DES, but also quantify the influence of operating conditions such as temperature and pressure on solubility, providing data support for process design [43] .

Lignin dissolution: By combining real characterization data from gel permeation chromatography (GPC) and HSQC NMR, machine learning models can predict the solubility of lignin in different DES, providing an efficient computational tool for biomass pretreatment [45] .

The activity of enzymes in non-aqueous environments directly determines their potential for application in biocatalysis. The application of machine learning in this regard is mainly reflected in evaluating the effect of DES on enzyme conformational stability and activity. Researchers have successfully built a predictive model for the activity of enzymes such as laccase in DES by integrating conductor-like shielding model-real solvent (COSMO-RS) parameters with machine learning technology, achieving a rapid assessment of the biocompatibility of DES [44] .

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

2.4 Model Algorithm and Performance Comparison

In DES property prediction, researchers often use a variety of machine learning algorithms for comparison in order to find the optimal prediction model.

Model comparison: When comparing on the same dataset, support vector regression (SVR) is stable under some high-dimensional features, but it is easily limited by computational cost on large-scale datasets. Random forest (RF) improves generalization ability through random sampling and feature selection, but it may have bias when dealing with extreme values (such as high viscosity) [38] .

Advantages of ensemble learning: Gradient boosting trees (GBDT) and their optimized implementations (such as XGBoost and LightGBM) perform well when dealing with large-scale datasets. Studies show that XGBoost, after introducing regularization terms, sparsity awareness and other techniques, not only has a fast training speed but also high accuracy [46] [47] ; while LightGBM, through histogram algorithm and leaf growth strategy, further outperforms XGBoost in terms of memory efficiency and training speed, and is especially suitable for scenarios with high feature dimensions and large data volume [48] .

3. Machine Learning-assisted DES Design and Performance Optimization

EDS, as a new generation of green solvents, possesses core advantages such as simple synthesis, designable structure, and environmental friendliness, and has broad application prospects in chemical fields such as gas capture, natural product extraction, and biomass pretreatment. The macroscopic properties of DES are determined by the composition, molar ratio, and microscopic interactions of hydrogen bond acceptors and donors. However, traditional trial-and-error development methods struggle to address the chemical space explosion caused by massive HBA/HBD combinations, failing to achieve targeted matching of DES structure, performance, and application. Machine learning, with its powerful nonlinear fitting, high-dimensional data mining, and global optimization capabilities, provides a new paradigm for the rational design, precise performance control, and application process optimization of DES, effectively overcoming the efficiency bottlenecks and generalization limitations of traditional methods.

3.1 Intelligent Prediction of DES Formation Capability

The stable formation of DES is a prerequisite for its subsequent performance regulation and engineering applications. Essentially, HBA and HBD interact through intermolecular hydrogen bonds to form a homogeneous system with a melting point significantly lower than that of a single component. Traditional DES screening methods suffer from limitations such as long development cycles, high reagent consumption, and weak generalization ability of novel components. Constructing machine learning classification models based on microscopic features is the core path to achieving high-throughput prediction of DES formation capability.

Starting from microscopic features such as hydrogen bond patterns, a high-performance binary classification prediction model for DES formation can be constructed. In the study of Usman L. Abbas et al. [49] , experimental verification data and molecular dynamics simulation results of hundreds of DES and non-DES systems were systematically mined, and core hydrogen bond features directly related to DES formation were extracted, including the number and lifetime of intramolecular/intermolecular hydrogen bonds of HBA and HBD, the number of hydrogen bond donor/acceptor sites, hydrogen bond binding energy, and the number of hydrogen bond active functional groups such as hydroxyl/carboxyl groups. Key structural features such as molecular topological polar surface area, molecular size, and van der Waals volume were also included. A binary classification model was constructed based on algorithms such as gradient boosting tree, random forest, and support vector machine, which achieved high-throughput and high-accuracy prediction of whether a given HBA-HBD combination can form a stable DES. The model had a prediction accuracy of over 92% on the independent test set and an area under the receiver operating characteristic curve

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

(AUC) of 0.94, which was significantly better than the traditional thermodynamic phase equilibrium model.

Feature importance analysis based on machine learning interpretability method can quantitatively reveal the key mechanism of DES formation. By ranking the contribution of model input features by SHapley Additive exPlanations (SHAP) value, Gini coefficient and other methods, it was found [49] that the hydrogen bond acceptance ability of HBA anion, the number of active hydrogen bond functional groups of HBD, and the hydrogen bond binding energy between HBA and HBD are the three core features that determine whether the system can form stable DES, with a cumulative contribution of over 70%. At the same time, the optimal range of hydrogen bond strength and the indirect regulatory effects of molecular size matching degree, alkyl chain steric hindrance and component molar ratio on DES formation were clarified, breaking through the limitations of the qualitative description of DES formation mechanism in traditional research, and providing a quantitative theoretical basis for the rational design of new DES systems.

3.2 Application-Oriented DES Reverse Engineering

Traditional DES (Device Engineering System) development follows a forward trial-and-error model of "component synthesis - structural characterization - performance testing - application verification," which struggles to quickly match the performance requirements of specific application scenarios and cannot address the design requirements of collaborative optimization of multiple performance indicators. Therefore, application-scenario-oriented DES reverse engineering has become a core component of machine learning-assisted design. To overcome the multi-objective optimization dilemma in complex applications, researchers have coupled high-precision prediction models with global optimization algorithms. Table 2 summarizes the typical strategies and core achievements of current machine learning in DES reverse engineering across different chemical engineering application scenarios.

Table 2. Typical Cases of DES Machine Learning Reverse Engineering Design for Specific Application Scenarios

Application scenarios	Target performance	Prediction/Generation Algorithms	Global optimization strategy	Core design achievements	References
Gas capture	CO_2 solubility prediction and capacity enhancement	GBR, RF, DNN, SDVAE	Genetic Algorithm (GA)	The test set R^2 reached a maximum of 0.97; a novel system with a 35.3% increase in absorbance capacity was identified; the accuracy of anion and cation reconstruction was >98.8%.	[52]
Biomass pretreatment	Simultaneous optimization of lignin yield, purity, and molecular weight	Gradient Boosting Machine (GBM)	Multi-objective optimization algorithm	Training was conducted using 944 datasets, with training set $R^2 > 0.95$; directional design of a DES for high-purity, high-molecular-weight separation was achieved.	[53]
Natural product extracts	Extraction rates of total phenols, total flavonoids, and anthocyanins	XGBoost	Currently unavailable (focusing on positive performance prediction)	The test set R^2 reached 0.73~0.93, accurately predicting the extraction efficiency of betaine DES.	[54]

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

High-precision forward performance prediction models are a prerequisite for realizing reverse design of DES. For the core performance indicators of different application scenarios, relevant studies have constructed a series of machine learning prediction models: In the field of gas capture, Jiasi Sun et al. [50] and Yuan Tian et al. [51] systematically integrated thousands of sets of CO_2 solubility experimental data in DES, and constructed machine learning models including gradient boosting regression (GBR), RF, and deep neural network (DNN) with HBA/HBD RDKit molecular descriptor, system temperature, and pressure as input features, realizing high-precision prediction of CO_2 solubility in DES. The test set determination coefficient R^2 can reach up to 0.97. At the same time, the model can also be extended to the solubility prediction of industrial waste gases such as H_2S and SO_2; In the field of natural product extraction, Shenglin Wang et al. [52] used the viscosity, pH, polarity of DES, and the structural parameters and molar ratio of HBA/HBD as inputs to construct an XGBoost model, which accurately predicted the extraction rate of total phenols, total flavonoids and anthocyanins from peony petals by betaine DES. The R^2 of the test set reached 0.73~0.93. In the field of biomass pretreatment, Weijin Zhang et al. [53] constructed a gradient boosting machine (GBM) model based on 944 sets of experimental data, which realized the synchronous prediction of lignin yield, purity and molecular weight after DES pretreatment. The R^2 of the training set all exceeded 0.95, which laid the foundation for the reverse design of DES for lignin separation.

Based on the forward prediction model, by coupling a global optimization algorithm, reverse recommendation and virtual screening of DES components and ratios can be achieved. Currently, the most widely used optimization algorithms include Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Gradient-based Particle Swarm Optimization (GBPSO). These algorithms can achieve global optimization in high-dimensional chemical space and avoid getting trapped in local optima. Jiasi Sun et al. [50] coupled the CO_2 solubility prediction model with GA to construct a complete DES reverse design framework. From thousands of potential HBA/HBD combinations, they screened out a new system with a CO_2 absorption capacity that was 35.3% higher than that of traditional DES. Yuan Tian et al. [51] targeted the biomass lignin separation scenario, using lignin yield, purity, and molecular weight as multi-objective optimization functions. They combined the GBM prediction model and optimization algorithm to reverse optimize the HBA/HBD type, molar ratio, and reaction conditions of the DES, realizing the directional design of DES for high-purity, high-molecular-weight lignin separation. In addition, Jiasi Sun et al. [50] also used a syntax-directed variational autoencoder (Syntax-Directed Variational Autoencoder). SDVAE transforms the molecular structure of DES into chemical space coordinates and directly generates novel DES molecules that meet the target performance through a generative deep learning model. Its anion and cation reconstruction accuracy reaches 99.0% and 98.8%, respectively, breaking through the limitations of the existing HBA/HBD component library and greatly expanding the chemical design space of DES.

DES (Digital Extraction System) to simultaneously meet multiple performance indicators in practical applications, multi-objective inverse design has become a current research focus. By employing multi-objective optimization algorithms such as non-dominated sorting genetic algorithms, combined with machine learning multi-output prediction models, the constraints between different performance indicators such as high CO_2 solubility and low viscosity, high extraction rate and low toxicity, and high lignin yield and high molecular weight can be effectively balanced. This achieves synergistic optimization of the overall performance of DES, allowing for the selection of the optimal DES system that meets the multi-dimensional requirements of industrial applications.

3.3 ML Optimization of Extraction and Separation Processes

The practical application effect of DES in the field of extraction and separation depends not only on its own physicochemical properties, but also on process parameters such as extraction temperature, reaction time, liquid-solid ratio, ultrasonic power, and solvent water content. The optimization of process parameters directly determines the extraction efficiency, separation purity and process economy of the target product. Machine learning provides key technical support for this step.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

In terms of replacing and improving the performance of traditional optimization methods, traditional DES extraction and separation process optimization mostly adopts response surface method (RSM). This method can only characterize the linear and second-order interaction between parameters, and it is difficult to accurately describe the complex nonlinear coupling relationship between multiple factors. The experimental workload increases exponentially in high-dimensional parameter scenarios, and it is easy to get trapped in local optima. Machine learning methods with artificial neural networks (ANN) as the core, and the RSM-ANN-GA hybrid optimization framework [54] , effectively solve the above defects. Related studies have shown that the ANN model has significantly better prediction accuracy and optimization stability than the traditional RSM model for active ingredient extraction rate. In high-dimensional process parameter optimization scenarios , machine learning can achieve global optimization by combining a small amount of experimental data with an active learning strategy, which greatly reduces the experimental workload and has more prominent advantages.

In terms of multi-objective optimization and mechanism insight, the value of machine learning is not limited to the precise optimization of single objectives such as extraction rate and separation efficiency, but can also take into account multiple optimization objectives such as product purity, solvent recovery rate, process energy consumption, and production cost, so as to achieve a synergistic balance between the economy, environmental protection and efficiency of DES extraction and separation process. At the same time, based on the feature importance and SHAP value analysis of interpretable machine learning [53] , the key influencing factors and micro-mechanisms of the extraction and separation process can be quantitatively analyzed, the differential control mechanism of different process parameters to target response can be clarified, and the optimization of DES extraction and separation process can be upgraded from "experience trial and error" to "theoretical guidance", providing universal theoretical guidance for the process design and industrial scale-up of similar systems.

4. Conclusion and Future Perspectives

Deep eutectic solvents, as a class of highly tunable green solvents, have demonstrated broad potential in extraction and separation, biomass valorization, gas capture, electrochemical energy storage, and catalytic synthesis. However, the vast number of possible hydrogen bond acceptor/hydrogen bond donor combinations and molar ratios makes conventional trial-and-error screening inefficient, costly, and difficult to generalize. Machine learning offers a promising data-driven strategy for accelerating DES property prediction, formation screening, rational design, and process optimization.

This review summarizes recent advances in machine-learning-assisted DES research. For property prediction, machine learning models can establish complex nonlinear relationships between molecular descriptors, experimental conditions, theoretical descriptors, and target properties, enabling efficient prediction of density, viscosity, heat capacity, sound velocity, solubility, CO_2 absorption capacity, lignin dissolution ability, and enzyme activity. For DES design, classification models based on hydrogen-bonding characteristics and molecular structural features can predict DES formation capability, while the integration of predictive models with global optimization algorithms provides effective support for reverse design and multi-objective optimization in specific application scenarios.

Nevertheless, several challenges remain. Current DES datasets are still limited in scale, fragmented across different studies, and insufficiently standardized, which restricts model reproducibility, transferability, and generalization. Moreover, many existing models mainly capture statistical correlations rather than explicit physicochemical mechanisms, and their extrapolation ability for new DES systems or complex industrial conditions remains limited. Future research should focus on constructing standardized DES databases, integrating experimental data with quantum chemical and molecular simulation descriptors, and developing interpretable, physics-informed, and experimentally validated machine learning models.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

Overall, machine learning is promoting the transition of DES research from empirical screening toward data-driven and application-oriented rational design. With continued improvements in data quality, model interpretability, and experimental validation, machine learning is expected to further advance the intelligent development and industrial application of DESs in green and sustainable chemical engineering.

References

[1] Yin, Y., Li, S. H., Han, J., et al. (2025). Research progress on deep eutectic solvents. *Shandong Chemical Industry*, *54*, 95–97.

[2] Hou, Y. C., Yao, C. F., & Wu, W. Z. (2018). Deep eutectic solvents: green solvents for separation applications. *Acta Physico-Chimica Sinica*, *34*, 873–885.

[3] Abbott, A. P., Capper, G., Davies, D. L., et al. (2002). Novel solvent properties of choline chloride/urea mixtures. *Chemical Communications*, *1*, 70–71. https://doi.org/10.1039/b108270g

[4] Wu, C., Chen, Q. C., Ni, Q., et al. (2025). Genetic algorithm-back propagation artificial neural network model-and response surface methodology-based optimization of crude Lophatherum gracile Brongn. polysaccharides extraction and their anti-NAFLD effects via regulating liver-adipose tissue crosstalk. *International Journal of Biological Macromolecules*, *320*, 145806. https://doi.org/10.1016/j.ijbiomac.2025.145806

[5] Wang, S., Wei, J., Ge, H., et al. (2025). Molecular simulation and machine learning assisted in exploring betaine-based deep eutectic solvent extraction of active compounds from peony petals. *Separation and Purification Technology*, *361*, 131550. https://doi.org/10.1016/j.seppur.2025.131550

[6] Cai, J., Xue, J., Li, Y., et al. (2024). Research progress on high-value-added application of lignocellulosic biomass based on deep eutectic solvent pretreatment. *Biomass Conversion and Biorefinery*, *14*, 30913–30928. https://doi.org/10.1007/s13399-023-01542-9

[7] Xiao, T., Hou, M., Guo, X., et al. (2024). Recent progress in deep eutectic solvent (DES) fractionation of lignocellulosic components: a review. *Renewable and Sustainable Energy Reviews*, *192*, 114243. https://doi.org/10.1016/j.rser.2024.114243

[8] Lin, B., Liu, X. L., & Bi, C. Y. (2025). Research progress on pretreatment of lignocellulose with deep eutectic solvents. *Jiangxi Chemical Industry*, *41*, 11–14.

[9] Yu, Q. Y., Long, Y., Liu, Y. Z., et al. (2025). Application progress of deep eutectic solvents in hydrometallurgical leaching. *Green Mining and Metallurgy*, *41*, 23–29.

[10] Zhang, Y. (2015). Study on leaching of zinc oxide ore by deep eutectic solvent-type ionic liquids [Master's thesis]. Kunming University of Science and Technology.

[11] Rodriguez Rodriguez, N., Machiels, L., & Binnemans, K. (2019). p-Toluenesulfonic acid-based deep-eutectic solvents for solubilizing metal oxides. *ACS Sustainable Chemistry & Engineering*, *7*, 3940–3948. https://doi.org/10.1021/acssuschemeng.8b06146

[12] Zhou, K., Dai, X., Li, P., et al. (2024). Recent advances in deep eutectic solvents for next-generation lithium batteries: safer and greener. *Progress in Materials Science*, *146*, 101338. https://doi.org/10.1016/j.pmatsci.2024.101338

[13] Wang, S. S., Li, W. S., Guo, B., et al. (2026). Efficient catalytic synthesis of ethylene glycol monoacetate with deep eutectic solvents under mild conditions. *Acta Petrolei Sinica (Petroleum Processing Section)*, 1–15.

[14] Chen, S. M., Liu, J. C., Zhong, Z. X., et al. (2025). Development and application of deep eutectic solvents in carbon dioxide capture. *Chemical Industry and Engineering Progress*, *44*, 5377–5390.

[15] Liu, J. X., & Hua, E. (2025). Analysis of CO2 absorption by glycerol-alkylethylenediamine deep eutectic solvents based on DFT theory and MD simulation. *Low-Carbon Chemistry and Chemical Engineering*, *50*, 121–132.

[16] Sun, J., Sato, Y., Sakai, Y., et al. (2023). A review of ionic liquids and deep eutectic solvents design for CO2 capture with machine learning. *Journal of Cleaner Production*, *414*, 137695. https://doi.org/10.1016/j.jclepro.2023.137695

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

[17] Tian, Y., Zhang, H., Qiao, Y., et al. (2025). Intelligent prediction of ionic liquids and deep eutectic solvents by machine learning. *Chinese Journal of Chemical Engineering*.

[18] Uzochukwu, M. I., Oyegoke, T., Momoh, R. O., et al. (2023). Computational insights into deep eutectic solvent design: modeling interactions and thermodynamic feasibility using choline chloride & glycerol. *Chemical Engineering Journal Advances*, *16*, 100564. https://doi.org/10.1016/j.cejadv.2023.100564

[19] Abdollahzadeh, M., Khosravi, M., Hajipour Khire Masjidi, B., et al. (2022). Estimating the density of deep eutectic solvents applying supervised machine learning techniques. *Scientific Reports*, *12*, 4954. https://doi.org/10.1038/s41598-022-08946-7

[20] Yang, F., Zheng, Q., Tan, H., et al. (2024). Insight into the role of hydrogen bond donor in deep eutectic solvents. *Journal of Molecular Liquids*, *399*, 124332. https://doi.org/10.1016/j.molliq.2024.124332

[21] Wang, Y., Liu, H., Ji, X., et al. (2023). Production of nanocellulose using acidic deep eutectic solvents based on choline chloride and carboxylic acids: a review. *International Journal of Biological Macromolecules*, *245*, 125227. https://doi.org/10.1016/j.ijbiomac.2023.125227

[22] Lee, K. M., Yoo, J., Kim, S. W., et al. (2019). Autonomic machine learning platform. *International Journal of Information Management*, *49*, 491–501. https://doi.org/10.1016/j.ijinfomgt.2019.07.011

[23] Choi, H. J., Naznin, M., Alam, M. B., et al. (2022). Optimization of the extraction conditions of Nypa fruticans Wurmb. using response surface methodology and artificial neural network. *Food Chemistry*, *381*, 132086. https://doi.org/10.1016/j.foodchem.2022.132086

[24] Wan, H. Y., Zhang, Y. Y., Yu, L., et al. (2020). Optimization of ultrasonic extraction process of two active components in safflower based on random forest model. *Journal of Chinese Medicinal Materials*, *43*, 673–677.

[25] (2025). Application of machine learning in deep eutectic solvents: advances in property prediction and molecular design. *Journal of Molecular Liquids*, *437*, 128317. https://doi.org/10.1016/j.molliq.2025.128317

[26] Abbas, U. L., Zhang, Y., Tapia, J., et al. (2024). Machine learning-assisted design of deep eutectic solvents based on unravelled hydrogen bonding patterns. *Engineering*, *39*, 74–83. https://doi.org/10.1016/j.eng.2023.12.008

[27] Abbott, A. P., Boothby, D., Capper, G., et al. (2004). Deep eutectic solvents formed between choline chloride and carboxylic acids: versatile alternatives to ionic liquids. *Journal of the American Chemical Society*, *126*, 9142–9147. https://doi.org/10.1021/ja0482667

[28] Liu, Y., et al. (2025). Screening deep eutectic solvent systems with UNIFAC: application to aromatics extraction from pyrolysis gasoline. *Chemical Engineering Science*, 123130. https://doi.org/10.1016/j.ces.2025.123130

[29] Wojciechowski, J. P., et al. (2021). Using COSMO-RS to predict solvatochromic parameters for deep eutectic solvents. *ACS Sustainable Chemistry & Engineering*, *9*, 12113–12127. https://doi.org/10.1021/acssuschemeng.1c03252

[30] Le, T., Winter, R., Noé, F., et al. (2020). Neuraldecipher - reverse-engineering ECFP fingerprints to their molecular structures. ChemRxiv. https://doi.org/10.26434/chemrxiv.12965555.v1

[31] Klamt, A., et al. (1975). UNIFAC group contribution method for activity coefficients. *AIChE Journal*, *21*, 108–115. https://doi.org/10.1002/aic.690210115

[32] Roosta, A., Haghbakhsh, R., Duarte, A. R. C., et al. (2023). Machine learning coupled with group contribution for predicting the density of deep eutectic solvents. *Fluid Phase Equilibria*, *565*, 113672. https://doi.org/10.1016/j.fluid.2023.113672

[33] Li, Y., Wang, X., Zhou, L., et al. (2024). Interpretable artificial intelligence model for estimating heat capacity of deep eutectic solvents. *Chemical Engineering Transactions*, *96*, 71–78.

[34] Zhang, H., Liu, J., Chen, W., et al. (2024). Machine learning approach for mapping heat capacity of deep eutectic solvents for sustainable energy applications. *Journal of Molecular Liquids*, *399*, 126736. https://doi.org/10.1016/j.molliq.2024.126736

[35] Wang, J., Li, M., Zhang, Y., et al. (2024). Learning to listen: predicting speed of sound in deep eutectic solvents through molecular features. *AIChE Journal*, *70*, e18013. https://doi.org/10.1002/aic.18013

ISSN: 2414-1895 DOI: 10.6919/ICJE.202606_12(6).0011

[36] Luan, V. T. A. (2024). Integrating computational tools to address the thermophysical behavior of deep eutectic solvents in gas separation applications [Joint master's thesis]. Universitat Rovira i Virgili & Universidade Federal do Rio de Janeiro.

[37] Halder, A. K., Haghbakhsh, R., Ferreira, E. S. C., et al. (2025). Machine learning-driven prediction of deep eutectic solvents heat capacity for sustainable process design. *Journal of Chemical Theory and Computation*, *21*, 1359–1369. https://doi.org/10.1021/acs.jctc.5c00098

[38] Wu, T., & Lin, X. Q. (2025). Accurate viscosity prediction of deep eutectic solvents achieved by multiscale feature mining. *AIChE Journal*, *71*, e18007. https://doi.org/10.1002/aic.18007

[39] Mood, M., et al. (2024). Accurate machine learning for the prediction of the viscosities of deep eutectic solvents. *Journal of Chemical Theory and Computation*, *20*, 5344–5357. https://doi.org/10.1021/acs.jctc.4c00528

[40] Zhang, Y., et al. (2022). Machine learning coupled with group contribution for predicting the density of deep eutectic solvents. *ChemistryOpen*, *11*, e202200097. https://doi.org/10.1002/open.202200097

[41] Cysewski, P., Jeliński, T., Przybyłek, M., et al. (2025). Navigating the deep eutectic solvent landscape: experimental and machine learning solubility explorations of syringic, p-coumaric, and caffeic acids. *International Journal of Molecular Sciences*, *26*, 10099. https://doi.org/10.3390/ijms262010099

[42] (2025). Data-driven classification of solubility space in deep eutectic solvents: deciphering driving forces using PCA and k-means clustering. *Journal of Molecular Liquids*, *437*, 128317. https://doi.org/10.1016/j.molliq.2025.128317

[43] (2024). Predicting the solubility of carbon dioxide in water and brine using advanced machine learning models. *International Journal of Greenhouse Gas Control*, *138*, 104726. https://doi.org/10.1016/j.ijggc.2024.104726

[44] (2024). Predicting laccase activity in deep eutectic solvents by combining COSMO-RS and machine learning techniques. *Journal of Chemical Technology & Biotechnology*, *99*, 4235–4245. https://doi.org/10.1002/jctb.7521

[45] (2025). Machine learning for deep eutectic solvents: advances in property prediction and molecular design. *Journal of Molecular Liquids*, *437*, 128317. https://doi.org/10.1016/j.molliq.2025.128317

[46] (2024). Comparison of machine learning models and performance. *Chemical Engineering Journal*, *475*, 144939. https://doi.org/10.1016/j.cej.2024.144939

[47] Chen, T., & Guestrin, C. (2016, August 13–17). XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

[48] Microsoft. (2024). *Comparison of LightGBM and XGBoost* [Technical documentation]. LightGBM Technical Documentation.

[49] Abbas, U. L., Zhang, Y., Tapia, J., et al. (2024). Machine-learning-assisted design of deep eutectic solvents based on uncovered hydrogen bond patterns. *Engineering*. https://doi.org/10.1016/j.eng.2023.12.008

[50] Sun, J., Sato, Y., & Kansha, S. Y. (2023). A review of ionic liquids and deep eutectic solvents design for CO2 capture with machine learning. *Journal of Cleaner Production*, *414*, 137695. https://doi.org/10.1016/j.jclepro.2023.137695

[51] Tian, Y., Zhang, H., Qiao, Y., et al. (2025). Intelligent prediction of ionic liquids and deep eutectic solvents by machine learning. *Chinese Journal of Chemical Engineering*.

[52] Wang, S., Wei, J., Ge, H., et al. (2025). Molecular simulation and machine learning assisted in exploring betaine-based deep eutectic solvent extraction of active compounds from peony petals. *Separation and Purification Technology*, *361*, 131550. https://doi.org/10.1016/j.seppur.2025.131550

[53] Zhang, W., Cao, Z., Chen, Q., et al. (2025). Machine learning-assisted separation of lignin from hydrothermal treatment of biomass by deep eutectic solvents. *Bioresource Technology*, 133537. https://doi.org/10.1016/j.biortech.2025.133537

[54] Liu, C., Lei, J., Liu, X., et al. (2025). Novel ternary deep eutectic solvent coupled with in-situ-ultrasound synergistic extraction of flavonoids from Epimedium wushanense: machine learning, mechanistic investigation, and antioxidant activity. *Ultrasonics Sonochemistry*, *121*, 107547. https://doi.org/10.1016/j.ultsonch.2025.107547