



OPEN Prediction of speed of sound of deep eutectic solvents using artificial neural network coupled with group contribution approach

Ayat Hussein Adhab¹, Morug Salih Mahdi², Hardik Doshi³✉, Anupam Yadav⁴, R. Manjunatha⁵, Sushil Kumar⁶, Debasish Shit⁷, Gargi Sangwan⁸, Aseel Salah Mansoor⁹, Usama Kadem Radi¹⁰ & Nasr Saadoun Abd¹¹

Predicting the physicochemical properties of deep eutectic solvents (DESs) is crucial for designing new solvents. Heat capacity and speed of sound are important thermodynamic properties in chemical processes. However, experimental data on the speed of sound in DESs is limited. Consequently, a thermodynamic model is needed to estimate the speed of sound in DESs over a wide range of pressures and temperatures. A key challenge in these models is accurately estimating the ideal gas heat capacity. Since the ideal gas heat capacity of DESs is often unavailable, a machine learning (ML) approach, using artificial neural networks (ANNs) coupled with a Group Contribution (GC) method, is a promising technique. The GC approach will be used to estimate critical temperature, volume, and acentric factor of DESs, which can then be input into the ANN model to predict the speed of sound. The results show that using a combination of a GC method and ANNs or CatBoost ML provides a highly accurate prediction of the speed of sound in DESs. Input parameters to the ANN + GC include temperature, acentric factor, molecular weight, and critical volume. The absolute relative deviation (ARD%) and R^2 values of correlated speed of sound for the ANN + GC model have been obtained 0.032% and 0.998, respectively. The ARD% for both the ANN + GC and ML + GC approaches was substantially lower than that of the correlation-based models. Furthermore, cumulative frequency diagrams and the leverage approach were implemented to validate the quality and reliability of the proposed model. The leverage analysis confirmed the accuracy of the data used and the high reliability of the ANN + GC model for estimating the speed of sound in DESs. This analysis indicates that the ANN + GC and ML + GC methods can effectively estimate the speed of sound in DESs based on molecular structure. Therefore, these approaches offer a promising tool for predicting the speed of sound of newly designed DESs when experimental data is unavailable.

Keywords Artificial intelligence, Neural network, Speed of sound, Deep eutectic, Group contribution, Machine learning, CatBoost

List of symbols

ARD (%)	Average relative deviation
b_j	Bias
IF_j	Relative importance of the j^{th} input variable on output variable

¹Department of Pharmacy, Al-Zahrawi University College, Karbala, Iraq. ²College of MLT, Ahl Al Bayt University, Karbala, Iraq. ³Department of Computer Engineering, Faculty of Engineering and Technology, Marwadi University Research Center, Marwadi University, Rajkot, Gujarat 360003, India. ⁴Department of Computer Engineering and Application, GLA University, Mathura 281406, India. ⁵Department of Data Analytics and Mathematical Sciences, School of Sciences, JAIN (Deemed to be University), Bangalore, Karnataka, India. ⁶Department of Computer Application, Chandigarh Engineering College, Chandigarh Group of Colleges-Jhanjeri, Mohali, Punjab 140307, India. ⁷Centre for Research Impact and Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India. ⁸Chitkara Centre for Research and Development, Chitkara University, Baddi, Himachal Pradesh 174103, India. ⁹Gilgamesh Ahliya University, Baghdad, Iraq. ¹⁰Collage of Pharmacy, National University of Science and Technology, Dhi Qar 64001, Iraq. ¹¹Medical Technical College, Al-Farahidi University, Baghdad, Iraq. ✉email: hardik.doshi@marwadieducation.edu.in

i	Input layers
h	Hidden layers
k	Input layers
m	Hidden layers
MAE	Mean average error
n_j	The j^{th} neuron output
N_{atoms}	Total number of atoms
N_i	Number of inputs
N_h	Hidden neurons
n	Output layers
o	Output layers
p_i	Output
P_c	Critical pressure
RMSE	Root Mean Square Deviation
R^2	Coefficient of determination
SD	Standard deviation
T_b	Normal boiling temperature
T_c	Critical temperature
V_c	Critical volume
w_{ij}	Weights of i^{th} neuron in the previous layer to the j^{th} neuron
w	Connection weights
ω	Acentric factor

Deep eutectic solvents (DESs) are mixtures of a hydrogen bond donor (HBD) and a hydrogen bond acceptor (HBA), which combine to form a solvent with a melting point lower than that of either individual component. Their advantageous properties, including low toxicity and low vapor pressure, have led to significant interest in DESs across a variety of fields¹. These types of solvents are considered an alternative to traditional organic solvents. Similar to ionic liquid (IL), DES physicochemical properties can be tuned by the suitable combination of HBD–HBA compounds. It must be noted that ILs and DESs are different due to the nature of starting materials and the methods for their formation. When considering starting materials and formation methods, ILs are mixtures of cations and anions, requiring synthesis with reagents and solvents. In contrast, DESs consist of HBAs and HBDs, and they can be prepared using single components through a simple heat treatment². DESs are widely used in pharmacology³, gas-capturing processes (especially CO₂ capture)⁴, extraction operations^{5–7}, and water treatment⁸. The low vapor pressure (almost negligible vapor pressure) of DESs is one of their main properties⁹. In industry, environmentally friendlier and greener solvent alternatives for the manufacturing of their products are required¹⁰. Bowen et al. have demonstrated the utility of DESs in protein extraction and purification¹⁰ highlighting their potential as a promising, alternative solvent for protein extraction from diverse raw biomass sources. Furthermore, recent research¹¹ has focused on designing green, highly efficient, and recyclable DESs for separating EVA films from end-of-life (EOL) photovoltaic (PV) modules. The inherent high-temperature stability and acidic nature of DESs can effectively facilitate the separation of EVA film layers within these modules¹¹. Jahanbakhsh-Bonab et al. used molecular dynamics (MD) simulations to examine the physicochemical and structural characteristics of novel DESs¹². Research indicates that methyl- β -cyclodextrin (MBCD)-based DESs can provide predictive insights for their applications in extraction processes. Furthermore, studies have explored the structural and physicochemical properties of chiral DESs, composed of racemic mixtures of menthol with acetic acid, menthol with lauric acid, and menthol with pyruvic acid, specifically for enantioselective extraction processes¹³. Jahanbakhsh-Bonab et al. utilized the MD simulations to examine the structural and dynamical properties of DESs-based boron nitride nanotube (BNNT) nanofluid¹⁴. The impact of nanotube diameter on the physicochemical parameters of DES-based systems has been investigated. Results indicate that adding Boron Nitride Nanotubes (BNNTs) to DESs increases viscosity due to a reduction in the diffusion coefficient of the DES components. Understanding the structure of DES-based nanofluids is crucial for comprehending the properties of these novel solvents in chemical processes. Moreover, MD simulations have been employed to examine the effects of external electric fields (EEFs) on the structural and transport properties of DESs composed of a 2:1 molar ratio of glycerol (Gly) and choline chloride (ChCl)¹⁵. They calculated several key physicochemical properties of Gly/ChCl DESs in the absence of external electric fields (EEFs), including viscosity, self-diffusion coefficient, isothermal compressibility, and density. They also employed the radial distribution function (RDF), coordination number, and number of hydrogen bonds to analyze the arrangement of the DES components. Their findings indicated that the correlation of movement between glycerol (Gly) and chloride (Cl) ions decreases as the strength of the EEF increases. Furthermore, in recent years, MD simulations have been increasingly used to investigate the performance of DESs in separating acid gases from natural gas¹⁶. The performance of DESs was compared with that of methyl diethanolamine (MDEA) system. The results show that, the diffusion coefficient of H₂S and CO₂ follow the trends Nano-DES < DES-MDEA < DES < MDEA < aqueous MDEA, and DES < DES-MDEA < MDEA < Nano-DES < aqueous MDEA in the liquid phase. Also, the performance of an amine-based DES composed of a 6:1 molar ratio of methyl diethanolamine (MDEA) and choline chloride (ChCl) for the natural gas (NG) sweetening process was investigated using MD simulations¹⁷. The effect of pressure on the performance of amine-based DESs for natural gas sweetening has been examined. The results suggest that DESs based on N-methyldiethanolamine (MDEA) can compete with conventional amine solvents in natural gas sweetening processes. P. Jahanbakhsh-Bonab et al. simulated the absorption capability of DESs based on MDEA using MD simulations¹⁸. Also, the effect of temperature on absorption capability was studied. The thermophysical properties of monoethanolamine (MEA)-based DESs for H₂S extraction from biogas were investigated¹⁹. The

impact of pressure on the performance of amine-based DESs for the extraction of H₂S and CO₂ was studied. Esfahani et al. used choline chloride-based DESs for the extraction of 1-butanol or 2-butanol from azeotropic n-heptane + butanol mixtures. This shows a difference in application and the type of DES used⁷. In summary, the prediction of thermodynamic properties of DESs plays an important role in chemical processes. The first-order derivative thermodynamic properties such as density, and second-order derivative thermodynamic properties such as heat capacity and speed of sound of the system, must be estimated in the pre-design of a new chemical process. Predictive models are more cost-effective than experimentally measuring the properties of a large number of potential designs (DESs) for a chemical process. M. Taherzadeh et al. proposed a correlation-based model to correlate/predict the heat capacity of 28 DESs over the wide temperature range from 278 to 363 K²⁰. The proposed correlation was developed based on molecular weight, temperature, acentric factor, and critical pressure. The absolute average relative deviation (AARD%) of the model for all of the studied data points was 4.7%. Leron and Li created a model specifically for estimating the heat capacity of DES²¹. Naser et al. developed a model for the specific heat capacity calculation of 15 DESs²². Zhang et al. proposed an empirical equation for the heat capacity estimation of two DESs²³. H. Peyrovedin et al. proposed a correlation-based model aimed at estimating the speed of sound in 39 different deep eutectic solvents (DESs) across a broad temperature range. This model likely utilizes empirical correlations derived from experimental data to provide accurate predictions of sound speed in these solvents, which is essential for understanding their thermophysical properties and potential applications in various fields²⁴. Their results indicated that the AARD% of the model is about 5.4%. Lapeña et al. correlated the speed of sound, heat capacity, density, isentropic compressibility, and viscosity of two DESs using a linear correlation²⁵. Peng and Minceva utilized the perturbed chain polar statistical associating fluid theory (PCP-SAFT) model to predict the density and viscosity of DESs²⁶. Lashkarbolooki et al. utilized the ANN to predict the heat capacity of binary ionic liquids mixtures²⁷. They collected 1571 binary heat capacity data points for ILs. A neural network with one hidden layer containing 16 neurons successfully predicted IL binary heat capacities²⁷. Thermodynamic models are a primary approach for predicting phase equilibrium and calculating second-order derivative thermodynamic properties

Theory and methodology

Data collection

As mentioned in the introduction section, the experimental data on the speed of sound are scarce.

In this study, 415 experimental data points of 38 DESs over a wide range of temperatures have been selected in the literature. The data points have been randomly divided into two sets containing 300 training and 115 testing data points. The most common train-test splits in the literature are 70:30 and 80:20, offering a good balance by providing enough data for both training and testing, and are often selected for their robustness and reliability in various contexts⁴⁴. The training data points have been considered to develop the model, and the test data points have been utilized to check the model performance.

The input layer is critical in machine learning models. Predicting the speed of sound in DESs requires appropriate input features, likely drawn from various thermophysical properties. However, a challenge arises with newly introduced DESs, as their thermophysical properties are often unknown. This presents a problem for training accurate machine learning models to predict the speed of sound in these novel solvents. The group contribution methods can be utilized to estimate the thermo-physical properties of DESs such as critical pressure, critical temperature and critical volume, and acentric factor. In this work, the modified Lydersen and Joback–Reid GC method^{45,46} is used to estimate the thermo-physical properties of DESs. In Table 1 the groups parameters of the Lydersen and Joback–Reid method have been presented.

Valderrama et al. estimated the critical points, normal boiling temperature, and acentric factor of ILs using the Joback method⁴⁵. As shown in Table 1, the ion parameters have been considered based on the Valderrama et al. method. The normal boiling temperature (T_b), critical temperature (T_c), critical pressure (P_c), critical volume (V_c), and acentric factor (ω) are estimated using Eqs. (1)–(5) as follows:

$$T_b (K) = 198 + \sum_k N_k t b_k \quad (1)$$

$$T_c (K) = T_b \left[0.584 + 0.965 \left\{ \sum_k N_k t c_k \right\} - \left\{ \sum_k N_k t b_k \right\}^2 \right]^{-1} \quad (2)$$

$$P_c (\text{bar}) = \left[0.113 + 0.0032 N_{atoms} - \sum_k N_k p c_k \right]^{-2} \quad (3)$$

$$V_c \left(\frac{\text{cm}^3}{\text{mol}} \right) = 17.5 + \sum_k N_k v c_k \quad (4)$$

where N_{atoms} is the total number of atoms in the molecule. The acentric factors of ILs were estimated as follows⁴⁶:

$$\omega = \frac{(T_b - 43)(T_c - 43)}{(T_c - T_b)(0.7T_c - 43)} \log \left(\frac{P_c}{P_b} \right) - \left(\frac{T_c - 43}{T_c - T_b} \right) \log \left(\frac{P_c}{P_b} \right) + \log \left(\frac{P_c}{P_b} \right) - 1 \quad (5)$$

In Table 2 critical properties and acentric factor of DESs have been reported.

Group	tb_k	tc_k	pc_k	vc_k	
Without rings					
-CH3	23.58	0.0141	-0.0012	65	
-CH2-	22.288	0.0189	0.0000	56	
>CH-	21.74	0.0164	0.0020	41	
>C<	18.25	0.0067	0.0043	27	
=CH2	18.18	0.0113	-0.0028	56	
=CH-	24.96	0.0129	-0.0006	46	
=C<	24.14	0.0117	0.0011	38	
-O-	[-O] ⁻	22.42	0.0168	0.0015	18
>C=O		94.97	0.0284	0.0028	55
>N-	[>N<] ⁺	11.74	0.0169	0.0074	9
-CN		125.66	0.0496	-0.0101	91
-F	[F] ⁻	-0.03	0.0111	-0.0057	27
-Cl	[Cl] ⁻	38.13	0.0105	-0.0049	58
-Br	[Br] ⁻	66.86	0.0133	0.0057	71
-I	[I] ⁻	93.84	0.0068	-0.0034	97
-OH		92.88	0.0741	0.0112	28
With ring					
-CH2-		27.15	0.0100	0.0025	48
>CH-		21.78	0.0122	0.0004	38
>C<		21.32	0.0042	0.0061	27

Table 1. Groups considered in the Joback–Reid method.

The GC method serves as an effective approach to estimate the thermophysical properties of DESs based on their molecular structures. This is particularly useful when experimental data is lacking, allowing researchers to generate necessary property estimates that can inform the modeling processes. In this work, the GC approach has been integrated with ANNs to predict the speed of sound in DESs.

Thoroughly analyzing and preparing input data is vital for building robust machine learning models. Each of these steps contributes to improving the quality of the input data, which in turn enhances the model's ability to learn and generalize from the training data. The input data have been analyzed and reported in Table 3⁵⁸.

In the next section, the ANN + GC method has been presented.

ANN methodology

The ANN can be likened to a black box with multiple inputs and outputs. The number of neurons can vary widely, ranging from fewer than ten to tens of thousands. These neurons can be organized into one, two, three, or more layers, with the two-layer network being the most commonly used ANN design for chemical applications. In the ANN methodology, inputs are fed into the input layer and then transmitted to the hidden layer using a specific transfer function. The transformed inputs are subsequently relayed to the output layer to estimate the desired properties. The estimated values are then compared to experimental data to analyze the error using the objective function (OF). Finally, the results are fed back into the system, and this process is repeated using a trial-and-error approach to minimize the objective function (OF) error. The number of neurons in the hidden layer is adjusted during these trials to achieve the best possible outputs with the lowest OF. This adjustment is guided by error analysis. Initially, one neuron is used to estimate the error with the training subset. Next, two neurons are evaluated, and this process continues, incrementing the number of neurons until the optimal configuration is found based on the minimum error value obtained from the testing subsets. Consequently, if the desired error is not achieved, the number of neurons in the hidden layer is increased. It must be noted that all input data are divided into two subsets namely training and testing. In Fig. 1 the schematic diagram of the neural network has been shown.

It must be noted that the weight parameters between neurons in hidden layers are utilized to develop the network. Summations of outputs of the previous layer in hidden layers must be weighted and added with bias. Hyperbolic tangent sigmoid transfer function is utilized in the hidden layer as follows:

$$n_j = \frac{2}{1 + \exp(-2Z)} - 1 \quad (6)$$

$$Z = \sum_{i=1}^r w_{ij} p_i + b_j \quad (7)$$

where, n_j is the j^{th} neuron output, w_{ij} refers to weights of i^{th} neuron in the previous layer to the j^{th} neuron, b_j is the bias, and p_i is output. In this work, the weight and Bias of output and hidden layer have been reported

DESs	HBA	HBD	HBA:HBD molar ratio	Tc (K)	Pc (bar)	Vc (cm ³ /mol)	ω	Exp. data
DES1	1-Ethyl-3-methylimidazolium	Ethylene glycol	2:1	670.98	36.65	355.99	0.666	47
DES2	1-Ethyl-3-methylimidazolium	Ethylene glycol	1:1	651.23	39.77	308.96	0.7476	
DES3	1-Ethyl-3-methylimidazolium	Ethylene glycol	1:2	632.35	43.77	264.25	0.8293	
DES4	Benzyl-tributyl-ammonium-chloride	Ethylene glycol	1:3	657.28	31.24	364.48	0.9659	48
DES5	Benzyl-tributyl-ammonium-chloride	Diethylene glycol	1:3	720.58	25.62	480.62	0.9994	
DES6	Benzyl-tributyl-ammonium-chloride	Triethylene glycol	1:3	778.21	22.07	589.83	1.0507	
DES7	Benzyl-tributyl-ammonium-chloride	Glycerol	1:3	749.11	25.67	433.69	1.3146	
DES8	Benzyl-trimethyl-ammonium-chloride	Ethylene glycol	1:3	618.43	41.08	270.56	0.8745	
DES9	Benzyl-trimethyl-ammonium-chloride	Diethylene glycol	1:3	678.15	31.88	377.22	0.908	
DES10	Benzyl-trimethyl-ammonium-chloride	Triethylene glycol	1:3	733.31	26.6	478.88	0.9593	49
DES11	Benzyl-trimethyl-ammonium-chloride	Glycerol	1:3	708.07	32.89	333.89	1.2232	
DES12	Benzyl-tripropyl-ammonium-chloride	Phenol	1:3	701.16	37.82	380.25	0.5152	
DES13	Benzyl-tripropyl-ammonium-chloride	Ethylene glycol	1:3	644.1	33.78	334.18	0.9375	
DES14	Benzyl-tripropyl-ammonium-chloride	Lactic acid	1:3	721.27	33.15	384.56	0.9166	
DES15	Betaine	Glycerol	1:3	735.27	27.58	401.61	1.2862	50
DES16	Betaine	Lactic acid	1:2	668.5	44.09	281.96	0.7863	
DES17	Betaine	Lactic acid	1:5	683.07	47.23	259.82	0.8755	
DES18	Betaine	Levulinic acid	1:2	701.24	38.94	356.12	0.6195	
DES19	Betaine	Lactic acid/water	1:1:1	637.98	61.84	206.94	0.5794	51
DES20	Choline-Chloride	Citric acid/water	2:1:6	659.71	92.43	146.46	0.5139	
DES21	Choline-Chloride	Urea	1:2	644.44	49.54	254.37	0.6509	
DES22	Choline-Chloride	Ethylene glycol	1:2	602	40.99	259.67	0.9155	52
DES23	Choline-Chloride	Glycerol	1:2	680.67	33.46	315.17	1.2254	
DES24	Choline-Chloride	Fructose	2:1	742.22	27.03	424.87	1.2278	
DES25	Choline-Chloride	Glucose	2:1	738.99	27.23	422.14	1.2163	
DES26	Choline-Chloride	1,2 propanediol	1:3	620.93	38.44	284.11	0.929	
DES27	Choline-Chloride	Levulinic acid	1:2	702.19	35.4	376.78	0.7301	53
DES28	Choline-Chloride	Malonic acid	1:1	689.82	37.16	335.84	0.8577	54
DES29	Choline-Chloride	Glutaric acid	1:1	713.43	32.24	397.17	0.8782	
DES30	Choline-Chloride	Oxalic acid	1:1	676.24	40.44	303.06	0.8531	55
DES31	Dodecanoic	Octanoic acid	1:3	737.07	24.71	559.27	0.7649	56
DES32	Dodecanoic	Decanoic acid	1:2	773.88	21.55	656.4	0.8307	
DES33	Menthol	Octanoic acid	1:1	717.72	28.79	493.39	0.6173	
DES34	Menthol	Decanoic acid	1:1	739.17	26.26	549.11	0.6568	57
DES35	Menthol	Salicylic acid	4:1	744.23	33.56	445.77	0.5733	
DES36	Menthol	Ethylene glycol	1:1	654.33	38.54	319.91	0.751	50
DES37	Proline	Levulinic acid	1:2	745.61	42.88	333.41	0.7044	
DES38	Proline	Lactic acid	1:1	721.95	48.54	272.6	0.8243	

Table 2. The list of studied DESs and their thermo-physical properties that estimated by the modified Lydersen and Joback–Reid^{45,46}, and references of speed of sound experiential data.

in Table 3. The speed of sound of new-designed DESs can be predicted using the weight, Bias, and input values (obtained by the GC approach). The results of the ANN + GC model have been described in section “ANN+GC method”.

Machine learning model

Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform tasks without explicit instructions. Instead of being programmed with specific rules, machines learn from data and improve their performance over time. In this study the Categorical Boosting (CatBoost) machine learning approach has been used^{59,60}. It excels at handling categorical features and often achieves state-of-the-art results in various machine learning tasks, particularly those involving tabular data. It's designed to be easy to use, robust, and provides excellent accuracy. CatBoost directly handles categorical features without needing extensive preprocessing like one-hot encoding (though it can still benefit from good feature engineering)⁶¹. It uses a special method to encode categorical features called Target Statistics (also known as Ordered Target Encoding). This method reduces overfitting compared to naive target encoding. For each categorical feature, CatBoost calculates the average value of the target variable for each category. However, to prevent overfitting (a common problem with target encoding), it uses a more sophisticated approach to calculate these statistics. It avoids using the target value of the current row when calculating the

Parameters	U (m/s)	Tc (K)	Vc (cm ³ /mol)	w (-)	MW (g/mol)
Minimum	1166	602	146.46	0.5139	59.39
Maximum	2138	778.21	656.4	1.3146	190.61
Range	972	176.21	509.94	0.8007	131.22
Mean	1666.1	694.6	368.77	0.8625	125.00
Median	1688.5	701.2	356.055	0.8554	121.84
Standard deviation	208.99	45.35	108.34	0.216	30.289
Variance	43,679.34	2056.84	11,739.3	0.0468	917.46
Mid range	1652	690.105	401.43	0.9142	125
Interquartile range	244.77	77.99	140.76	0.2549	50.03
Sum of squares	11,880,780.5	76,103.2	434,351.36	1.734	34,863.54
Mean absolute deviation	160.65	38.55	83.28	0.1672	24.76
Root mean square	1679.11	696.04	383.95	0.888	128.53
Skewness	-0.7036	-0.164	0.643	0.483	0.1129
Kurtosis	2.965	2.393	3.751	2.893	2.767
Coefficient of variation	0.1254	0.0652	0.2938	0.251	0.2423
Relative standard deviation	12.54%	6.529%	29.38%	25.1%	24.23%

Table 3. Statistical description of the data set used for modeling.

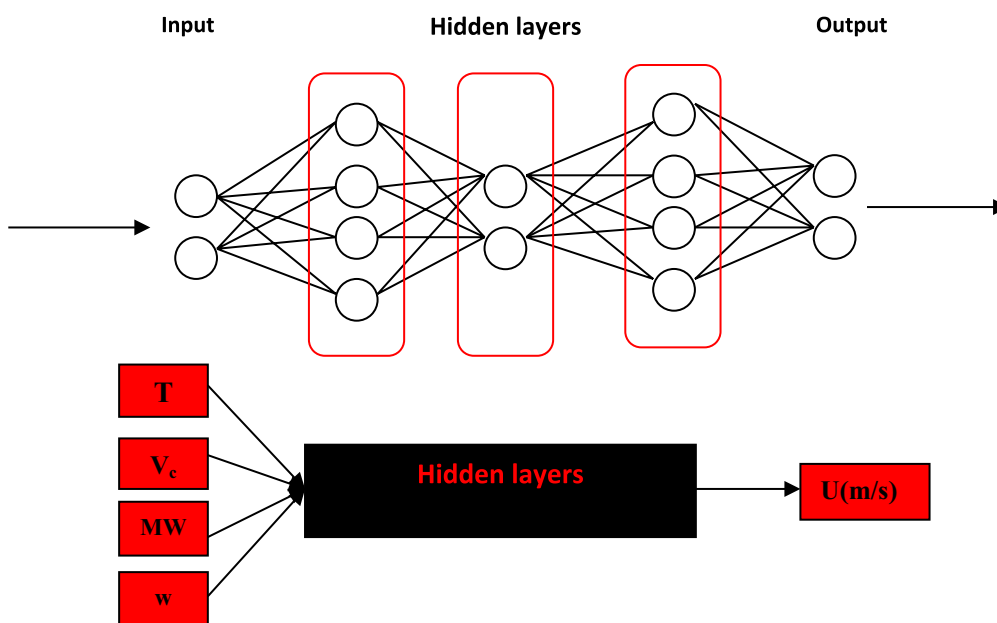


Fig. 1. The neural network diagram.

target statistic for that row. This is done by calculating the target statistic based on the rows that came before the current row in the dataset's order. Random permutations of the data are used to further reduce overfitting. CatBoost implements a variation of gradient boosting known as ordered boosting. This helps to address gradient bias that can occur in traditional gradient boosting algorithms, especially when dealing with categorical features. Gradient bias arises because the model is trained using the same data that was used to calculate the gradients, leading to an overestimation of the model's performance. Ordered boosting aims to correct this bias. Like other gradient boosting methods, CatBoost iteratively builds an ensemble of decision trees. Each tree is trained to correct the errors made by the previous trees. The optimization process is driven by gradients calculated from a loss function. Within CatBoost, categorical boosting entails the utilization of categorical columns, incorporating permutation techniques such as one hot max size (OHMS) and target-based statistics. This method employs a greedy approach for each new split of the current tree, enabling CatBoost to unveil the exponential expansion of feature combinations⁶¹. The CatBoost method follows these steps:

- Formation of a random subset of the records
- Converting labels to integers
- Transforming categorical features into numeric values, as outlined below:

$$\text{Average Target} = \frac{\text{Count in Class} + \text{Prior}}{\text{Total Count} + 1} \quad (8)$$

In the equation provided, the parameter “Count in Class” aggregates the targets. Furthermore, each target is assigned a value of one and linked to specific categorical features, while the term “Total Count” tallies all previous instances⁶². In Fig. 2, schematic of the CatBoost tree construction has been depicted.

Statistical error analysis

Statistical error analysis involves examining the errors and uncertainties in statistical measurements and data analysis. This is a crucial process in many fields, including science, engineering, economics, and social sciences, as it helps in assessing the reliability and validity of conclusions drawn from data. In this work the model results have been evaluated using the average relative deviation percent (ARD%), standard deviation (SD), mean absolute error (MAE), root mean square deviation (RMSE), and R^2 values as follows:

$$\text{ARD} (\%) = \frac{100}{N} \sum_{i=1}^N \left| \frac{U_i^{exp} - U_i^{calc}}{U_i^{exp}} \right| \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (U_i^{exp} - U_i^{calc})^2}{N}} \quad (10)$$

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^N \frac{(U_i^{exp} - U_i^{calc})^2}{U_i^{exp}}}{N - 1}} \quad (11)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |U_i^{exp} - U_i^{calc}| \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (U_i^{exp} - \bar{U}^{exp})^2}{\sum_{i=1}^N (U_i^{exp} - \bar{U}^{exp})^2} \quad (13)$$

where U_i^{exp} and U_i^{calc} are experimental and estimated speed of sound of DESs. \bar{U}^{exp} is the average value of the experimental speed of sound. In the next section the results of ANN + GC and ML + GC approaches have been presented.

Results and discussion

In the previous section the ANN and ML methodologies for the prediction of the speed of sound of DESs have been described and depicted in Figs. 1 and 2. In this section the ANN + GC and ML + GC approaches have been described separately.

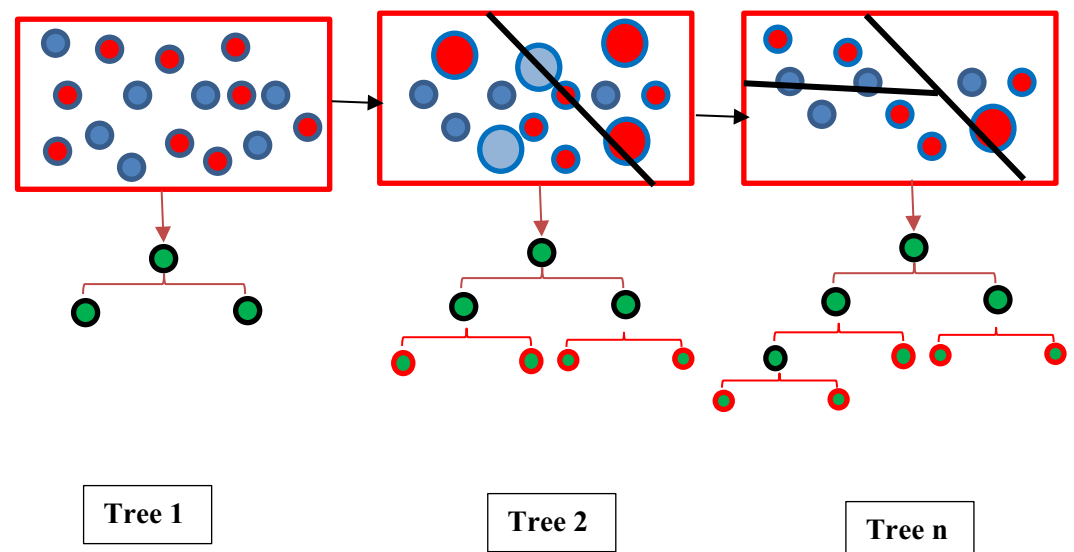


Fig. 2. Schematic of the CatBoost tree construction.

ANN + GC method

The number of independent variables (input layer) plays a crucial role in ANN and ML methods⁶³. In Table 1, thermodynamic properties of DESs containing T_c , V_c , MW, and ω have been reported. In the case of ANN method, different input properties and different numbers of neurons in one and two hidden layers have been considered, because one hidden layer only does not lead to adequate results^{64–67}. The number of the hidden layer and neuron in the hidden layer and output are obtained using a trial and error algorithm. In this work the Levenberg–Marquardt algorithm^{68,69} is used to optimize the aforementioned parameters. The results show that, one hidden layer containing 16 neurons and four input properties containing the critical volume (V_c), molecular weight (Mw), temperature, and acentric factor (ω) are the optimum values. As described in section “Theory and methodology”, 300 training and 115 testing data points of the speed of sound have been considered.

As mentioned in Eq. (7), the weight parameters between neurons in hidden layers are essential to develop the network. In Table 4 the weight parameters of neurons have been reported.

The optimum values of neurons in the hidden layer are evaluated using the average relative deviation percent (ARD%). When the optimum network architecture was determined, the input data of the ten DESs were fed to the network to predict their speed of sound. In Fig. 3 the flowchart of the proposed ANN model has been depicted.

As shown in Fig. 3, the model can predict the speed of sound of DESs using independent variables containing T , V_c , ω , and MW. The inputs containing V_c and ω can be estimated using GC approaches^{45,46}. Experimental literature data is used as the training dataset for sound speed. An ANN with one hidden layer containing sixteen neurons is employed to develop the model. Four input variables can be fed into the saved file to generate predicted sound speed. Using the “saved network” and these four inputs, the sound speed of DESs can be accurately predicted. The complete MATLAB code, including all source files used in the programming, is provided in the Supplementary Material. The correlated and predicted results of the ANN + GC approach have been shown in Fig. 4.

Figure 4 shows that the ANN + GC approach can correlate the speed of sound of DESs over a wide range of temperatures, satisfactory. The ARD% and R^2 of the correlated speed of sound have been obtained 0.032% and 0.9988, respectively. Figure 4b shows the prediction of the speed of sound of ten DESs using the ANN + GC approach. The results are in good agreement with experimental data. Figure 5 shows a simultaneous comparison between the experimental data and ANN + GC data.

As shown in Fig. 5, the ANN + GC method can correlate the experimental data accurately. Distributions of the deviation points for the ANN + GC method are shown in Fig. 6.

As shown in Fig. 6, the deviations between the ANN + GC predictions and experimental data do not exceed 4 m/s. Error analysis indicates that the proposed network is suitable for engineering calculations. In this study, the predictive performance of the ANN + GC model was assessed using R^2 , ARD%, SD, MAE, and RMSE metrics; see Table 5.

As shown in Table 5, the total ARD%, MAE, SD, RMSE, and R^2 values have been obtained 0.032%, 1.5656, 0.0549, 2.227, and 0.9988, respectively. The results of the ANN + GC approach show good agreement with experimental data. Models with high R^2 values nearing 1 and low values for ARD%, RMSE, MAE, and SD are considered more accurate in predicting the speed of sound. In this study, the ARD% for the training and testing phases of the ANN + GC model were 0.024% and 0.053%, respectively. The overall R^2 value approached unity at 0.9988. These results indicate that the ANN + GC model can accurately correlate the speed of sound in DESs across a wide temperature range. In the next section, the ML + GC model has been studied.

Neurons	Hidden layer				Output layer
	V_c	w	Mw	T	
1	-2.24536	-0.48558	1.098708	-1.16422	0.630794
2	0.944691	-2.52083	-0.67325	0.373787	0.758028
3	-1.01834	-1.29702	1.427834	-1.75557	0.977823
4	-1.74537	1.232633	1.03008	-1.4877	-0.99896
5	1.276081	-0.35078	-1.66558	-1.82055	0.730877
6	-1.16913	1.762558	1.645331	0.812039	0.225133
7	-0.31927	-0.60319	2.715422	-0.02673	0.9799
8	1.163699	1.667596	0.089376	-1.92274	0.05536
9	-0.93382	-0.68464	2.549148	-0.03313	-0.04095
10	1.244022	1.623924	0.463342	-1.85488	0.602695
11	-0.47005	1.138715	-1.54125	-1.98668	-0.54431
12	1.003801	-0.67099	-1.64259	1.919385	-0.00381
13	1.557248	-2.16729	-0.7101	0.462178	0.801705
14	-0.27949	1.406818	1.204804	2.081152	0.149322
15	-1.77034	1.655671	1.199766	0.724712	0.690356
16	-1.22579	-1.24965	2.101388	0.599993	0.477281

Table 4. The weight of hidden and output layers for 16 neurons.

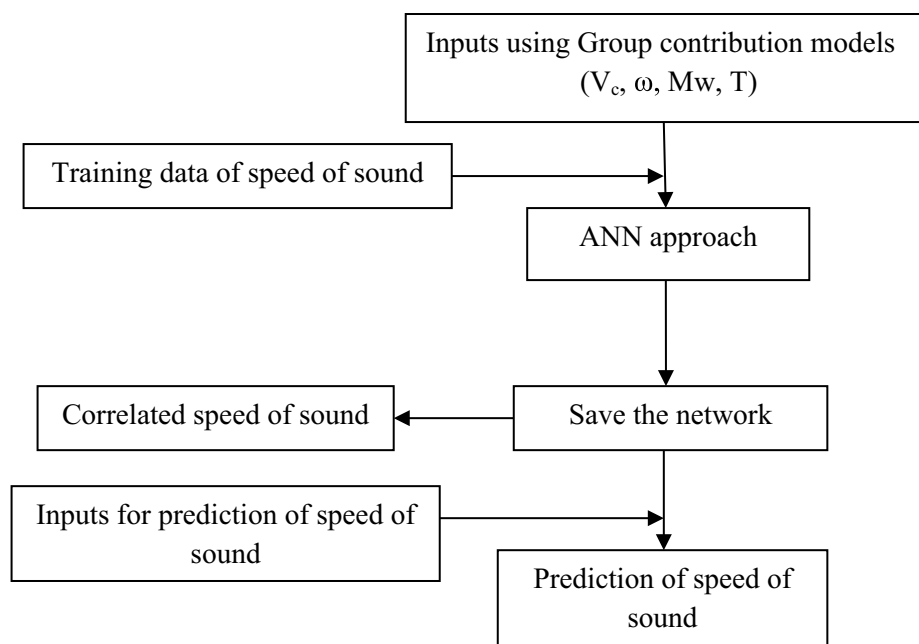


Fig. 3. Flowchart of proposed model.

ML + GC method

Similar to the ANN + GC method, the inputs for the ML + GC model included V_c , ω , M_w , and T . Additionally, 300 training and 115 testing data points of the speed of sound were used to develop the machine learning approach. The statistical metrics for the CatBoost model are summarized in Table 6, which presents evaluations on the training and testing subsets (300 and 115 data points, respectively), as well as the complete dataset consisting of 415 points. In this study, the predictive performance of the models was assessed using R^2 , ARD%, SD, MAE, and RMSE metrics. Comparing model predictions with experimental data across both training and testing datasets provides valuable insights into the models' accuracy and generalization capability; see Table 6.

The greater the alignment between the predicted value and the experimental data, the higher the accuracy of the predictive model. In Figs. 7 and 8, the error distribution plot of the presented model vs the predicted speed of sound has been depicted. This visual representation demonstrates the robust agreement between the experimental data and the forecasts produced by the CatBoost ML method.

Figures 7 and 8 illustrate a strong correlation between the model-predicted data and the experimental data across both the training and testing datasets. These figures demonstrate a very close alignment between the model predictions and experimental points. In this research, graphical analysis complemented statistical methods to provide a more comprehensive evaluation of the models' performance. These visual representations played a vital role in assessing the accuracy and reliability of the models. The percentage distribution of the relative error against the experimental values is presented in Fig. 7. In this type of error evaluation, relative error values are plotted against experimental output values. The closer the data points are to the zero-error line, the model is the more accurate. When the data points are scattered around the zero line, it indicates a significant difference between the predicted values and the experimental data, which proves the high error of the model. As a result, the proximity of the data points to the zero line for the ML + GC model indicates the high accuracy of this model. In Fig. 8, the cross-plot has been depicted. The cross-plot visually represents the comparison between predicted and experimental values. A closer alignment of data points with the unit slope line ($Y = X$) in the cross plot signifies higher accuracy and effectiveness of the model. The ML + GC model shows significant performance with most of the data points lying around the $Y = X$ line. In the next section, a comparison between ANN + GC, ML + GC, and the correlation-based models has been investigated.

Comparison between ANN + GC, ML + GC, and the correlation-based models

The ANN + GC and ML + GC results have been compared to five correlation-based models^{24,70–73}. Singh and Singh proposed a correlation for speed of sound based on the surface tension and density⁷⁰. Hekayati and Esmailzadeh suggested a novel interrelationship between surface tension (σ), density (ρ), and speed of sound (u) of ILs⁷¹. Gardas and Coutinho proposed a relationship between surface tension (σ), density (ρ), and speed of sound (u) for imidazolium based ILs, covering wide ranges of temperature, 278.15–343.15 K⁷³. The aforementioned models are correlation-based. In Table 7 the ARD% of five correlation-based, ML + GC, and ANN + GC models have been reported and compared.

The average ARD% value of Peyrovedin et al.²⁴ model was obtained 5.67%. ARD% values of Haghbakhsh et al.'s model⁷², Hekayati and Esmailzadeh's model⁷¹, and Gardas and Coutinho's model⁷³ for 38 DESs have been obtained 9.52%, 9.38%, and 9.45%, respectively. The average ARD% value of Singh and Singh's model⁷⁰ was obtained about 39%. They correlated the speed of sound of ILs using surface tension and density data. As shown

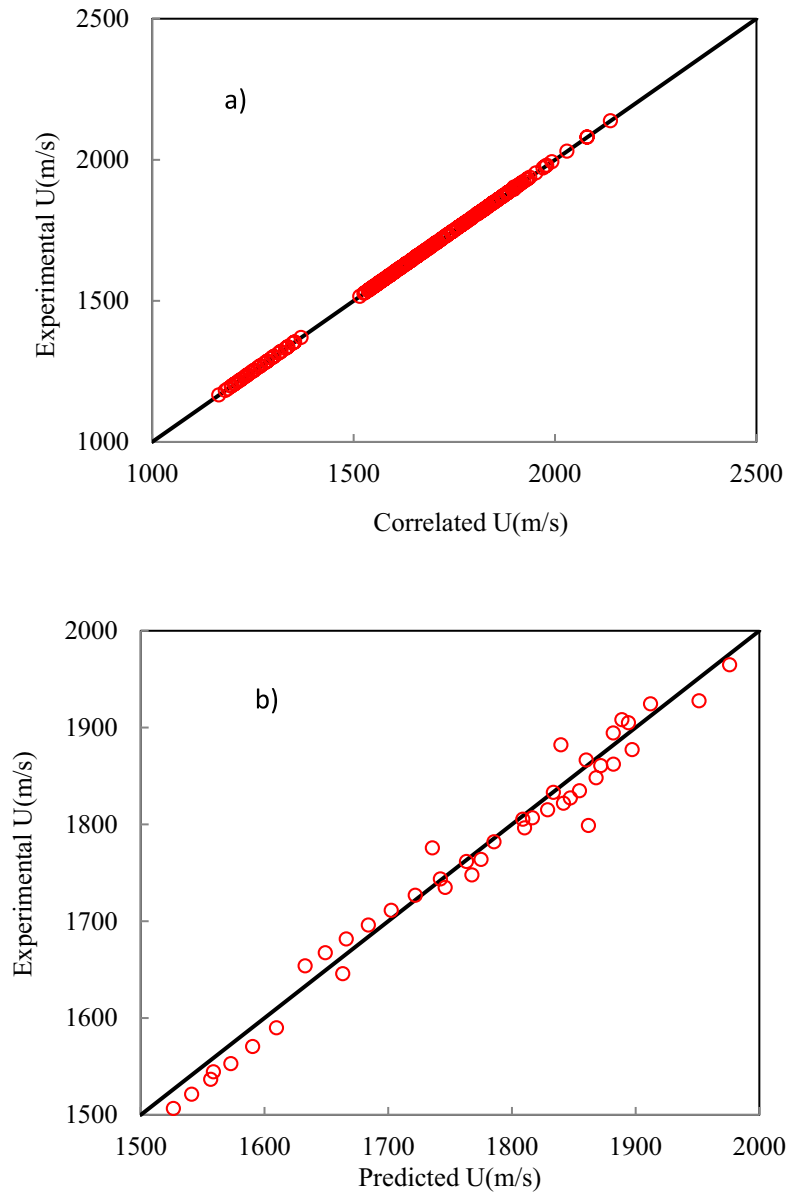


Fig. 4. The results for the correlated (a) and predicted (b) speed of sound using the ANN + GC approach.

in Table 7, the Peyrovedin et al.²⁴ model gives a lower ARD% value compared to other correlation-based models. The ANN + GC and the ML + GC models give lower error values compared to correlation-based models. The ARD% of the ANN + GC model is slightly lower than the ML + GC model. In Fig. 9 the speed of sound of some DESs using the ANN + GC approach have been compared to experimental data.

As shown in Fig. 9, the ANN + GC correlates the speed of sound of DESs satisfactorily. The average ARD% of the ANN + GC model was obtained 0.032%. In Fig. 10, the ANN + GC model results have been compared to experimental data and H. Peyrovedin et al. model.

As depicted in Fig. 10, the ANN + GC approach correlates the speed of sound of four DESs at various temperatures accurately. In the case of DES1, the ARD value of H. Peyrovedin et al. model is higher than ANN + GC, nevertheless, their obtained results are acceptable. Figure 10 shows that, their proposed correlation is accurate at lower temperatures, and the model deviations are increased by increasing temperature. As reported in Table 7, and Figs. 9 and 10, the average ARD% value of the testing and training results of the ANN + GC are acceptable. In Fig. 11, the ANN + GC model has been compared to the ML + GC and five correlation-based models.

As shown in Fig. 11, the average ARD% values of the ANN + GC approach are lower than the correlation-based models. The average error values of ANN + GC and ML + GC models are comparable. In Fig. 12, the error distribution plot for ten DESs has been depicted.

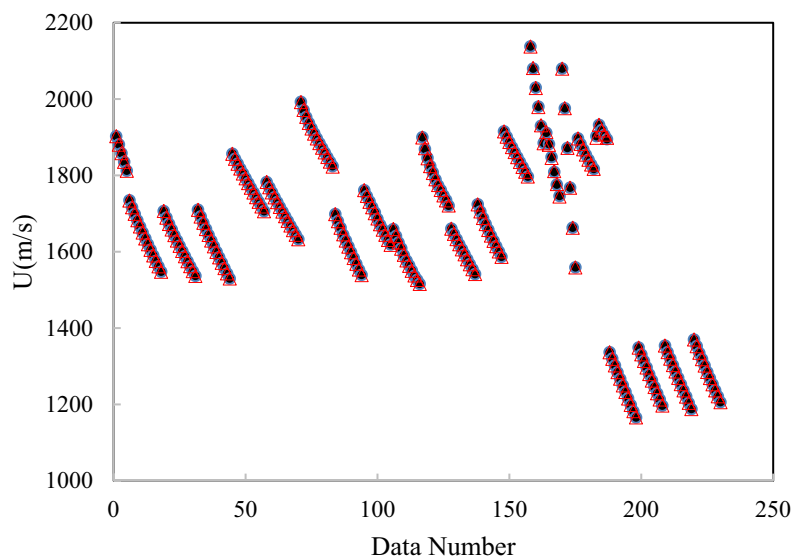


Fig. 5. The speed of sound of DESs obtained using the ANN + GC. (●) Experimental data and (Δ) ANN + GC.

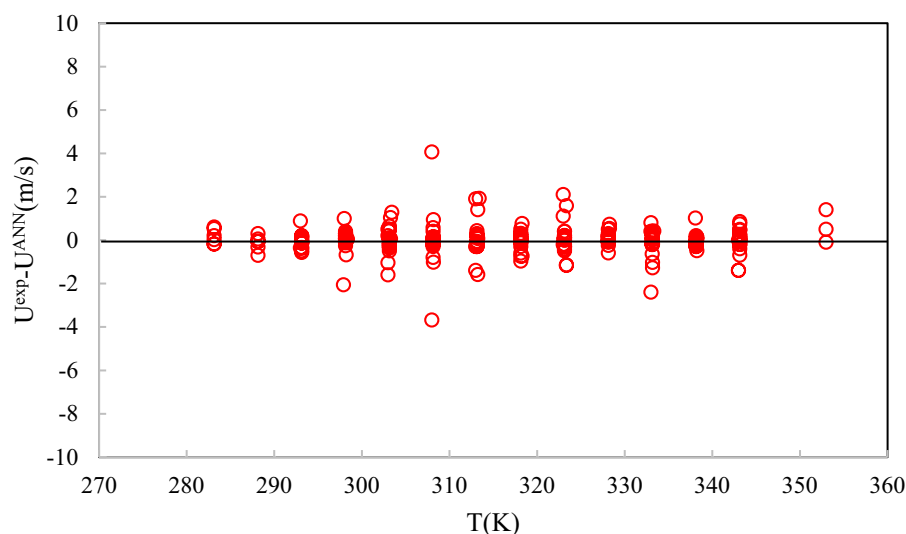


Fig. 6. Deviations between calculations from ANN + GC and experimental speed of sound data at different temperatures.

Model		ARD (%)	MAE	SD	RMSE	R ²
ANN + GC	Train	0.024	1.5625	0.0538	2.1754	0.9981
	Test	0.053	1.4350	0.0524	2.2121	0.9990
	Total	0.032	1.5656	0.0549	2.2270	0.9988

Table 5. Statistical error analysis for the ANN + GC model.

Cumulative frequency diagrams are one of the graphical methods used for evaluating model performance⁷⁴. Figure 13a and b illustrate the cumulative frequency diagrams of the ANN + GC and ML + GC models, along with five correlations (as reported in Table 7).

As shown in Fig. 13a, approximately 90% of the values estimated by the ANN + GC model exhibited an ARD% of less than 0.07%. In the case of ML + GC model, 90% of the ARD% values are less than 0.1%. In Fig. 13b, the cumulative frequency of five correlation-based models has been depicted. The correlation developed by Singh and Singh⁷⁰ demonstrated poor performance. The results show that, the ANN + GC model achieves high precision in forecasting speed of sound of DESs compared to the five correlation-based model.

Model		ARD (%)	MAE	SD	RMSE	R ²
CatBoost	Train	0.0381	0.9523	0.0516	2.1677	0.9991
	Test	0.0590	1.2349	0.0713	3.0208	0.9993
	Total	0.0332	0.9695	0.0556	2.3499	0.9992

Table 6. Statistical error analysis for the ML + GC model using CatBoost approach.

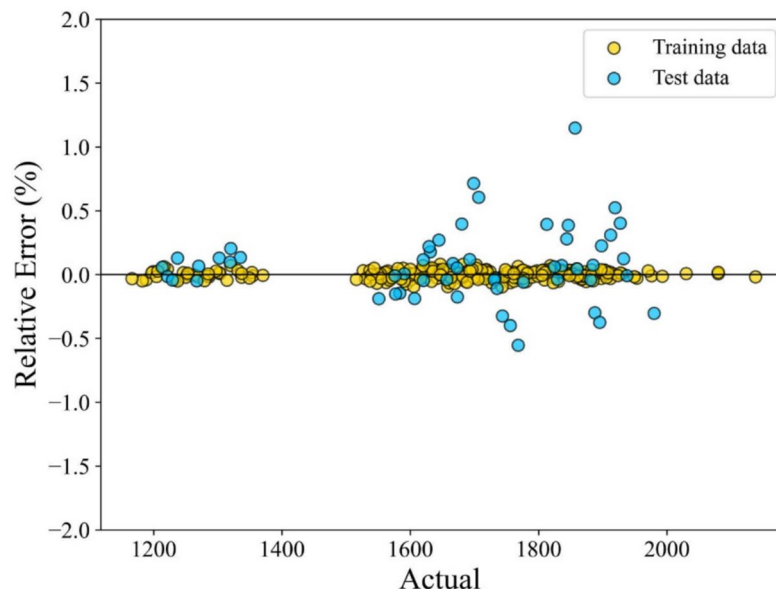


Fig. 7. Error distribution plot of the ML + GC model to predict speed of sound.

The leverage approach for model analysis

The leverage approach is a valuable tool for ensuring the quality and reliability of statistical models. Identifying and addressing high-leverage points, can improve model accuracy, enhance data understanding, and lead to more informed decision-making⁷⁵. Leverage values help identify observations that have a disproportionate influence on the regression coefficients. Points with high leverage and large residuals are particularly problematic, as they can significantly distort the model fit. Leverage diagnostics are used during model validation to assess the stability and generalizability of the model. If the model is highly sensitive to a few high-leverage points, it may not perform well on new data. High-leverage points often indicate data errors or unusual events. Identifying these points allows for a targeted investigation of the data to identify and correct errors or to understand the underlying causes of the unusual observations. High leverage points can sometimes indicate the need to include additional predictor variables in the model. In some cases, transforming the predictor or response variables can reduce the influence of high-leverage points and improve the model fit. As well, investigating high-leverage points can provide valuable insights into the data and the underlying processes that generated it. In this study the leverage approach has been utilized to study the ANN + GC model. In this regard, standardized residuals (SR) and Leverage values, derived from the diagonal elements of the hat matrix have been calculated. The hat matrix was given by:

$$H = X (X^t X)^{-1} X^t \quad (14)$$

where X^t refers to the transpose of matrix X . The critical leverage was calculated as $3(n+1)/m$, where m and n represent the number of data points and model input variables, respectively. The applicability domain of the ANN + GC model can be assessed by plotting standardized residuals against leverage values (Williams plot). The Williams plot is the most common and direct way to do this. The applicability domain (AD) of a model is the region where the model is considered reliable for making predictions. In simpler terms, it's the set of conditions under which you can trust the model's output. Extrapolating beyond the AD can lead to inaccurate or unreliable predictions.

By plotting standardized residuals against leverage values against each other, the Williams plot allows you to identify observations that:

- Are outliers (large standardized residuals)
- Have high leverage (unusual predictor values)
- Are both outliers and have high leverage (potentially very influential and problematic)

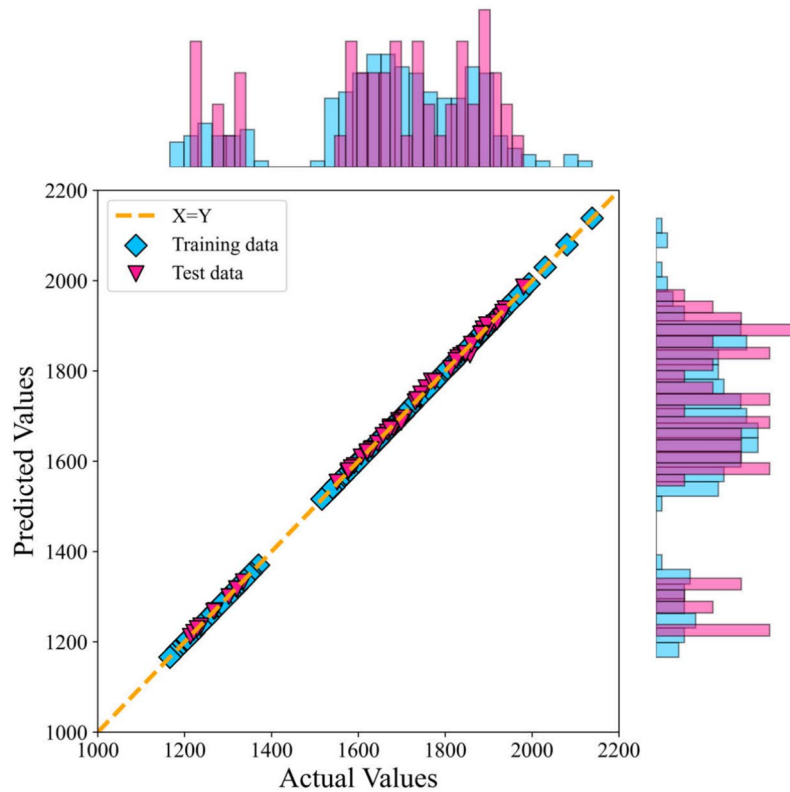


Fig. 8. Cross-plot of the ML+GC model to predict speed of sound.

If the majority of data points were situated within the boundaries of the $0 \leq H \leq$ critical leverage, and $-3 \leq SR \leq 3$, the established model is deemed reliable, and its predictions are confined within the applicability domain⁷⁵. In Fig. 14, the William’s plot is illustrated.

As shown in Fig. 14, the critical leverage value has been obtained about 0.0545. As depicted in Fig. 14, most of the data point falls between $0 \leq H \leq 0.0545$, and $-3 \leq SR \leq 3$. The results indicated that, the ANN+GC model is highly reliable. There are some suspicious data ($SR > 3$ or $SR < -3$). Figure 14 shows that, only five data points have an SR-value outside the range of -3 to 3 , classifying them as questionable data. On the other hand, all data points have H values lower than 0.0545. This result indicated that all data points have satisfactory leverage. The Leverage approach confirms the accuracy of databank and the high reliability of ANN+GC model in estimating speed of sound of DESs.

In the next section, the sensitivity analysis (SA) of input variables in the ANN+GC model has been studied.

Sensitivity analysis

Sensitivity analysis in ANNs involves determining how much each input variable influences the network’s output. It helps you understand which inputs are most important and how changes in those inputs affect the model’s predictions. Sensitivity analysis using weight-based methods involves evaluating the influence of input variables on the output by analyzing the weights within the network. These methods are generally more straightforward and computationally less expensive than perturbation-based methods. Garson suggested an equation based on partitioning of connection weights for sensitivity analysis of input variables as follows⁷⁶:

$$IF_j = \frac{\sum_{m=1}^{N_h} \left(\left(\frac{|w_{jm}^{ih}|}{\sum_{k=1}^{N_i} |w_{km}^{ih}|} \right) \cdot w_{mn}^{ho} \right)}{\sum_{k=1}^{N_h} \left\{ \sum_{m=1}^{N_h} \left(\left(\frac{|w_{km}^{ih}|}{\sum_{k=1}^{N_i} |w_{km}^{ih}|} \right) \cdot w_{mn}^{ho} \right) \right\}} \tag{15}$$

where IF_j is the relative importance of the j^{th} input variable on output variable; N_i and N_h refer to the number of input and hidden neurons, respectively. The superscripts i , h and o refer to input, hidden and output layers, respectively. The subscripts k , m and n refer to input, hidden and output layers, respectively. w is connection weights. The relative importance of input variables (IF_i) were calculated by Eq. (15). This approach expands on Garson’s method by considering the direct and indirect paths from inputs to outputs. It involves calculating the influence of each input across the network layers into the final output. In Fig. 15 the importance of input variables based on normalized percentage has been depicted.

DESs	ML and ANN approaches			Correlations				
	Data set	ANN + GC	ML + GC	Peyrovedin et al. ²⁴	Haghighbakhsh et al.'s model ⁷²	Hekayati and Esmailzadeh's model ⁷¹	Gardas and Coutinho's model ⁷³	Singh and Singh's model ⁷⁰
DES1	Test	0.074	0.084	13.9	14.5	8.6	15.0	4.9
DES2	Train	0.008	0.020	7.5	16.6	3.6	9.0	11.8
DES3	Test	3.85	0.062	1.3	17.6	2.6	1.3	31.6
DES4	Train	0.012	0.105	5.8	2.4	8.1	5.2	33.0
DES5	Train	0.007	0.022	1.8	3.3	6.6	7.1	36.5
DES6	Train	0.006	0.047	0.9	8.3	5.0	9.3	43.8
DES7	Test	0.063	0.021	5.4	9.0	4.7	14.2	84.5
DES8	Train	0.006	0.043	0.7	68.8	2.8	1.9	27.2
DES9	Train	0.005	0.068	1.1	1.4	3.6	1.1	28.0
DES10	Test	0.042	0.042	2.8	7.0	2.1	2.9	32.1
DES11	Train	0.005	0.052	4.6	7.6	0.9	6.8	76.5
DES12	Train	0.023	0.189	1.9	2.7	3.6	3.6	4.4
DES13	Train	0.041	0.065	2.5	1.8	4.8	1.2	26.1
DES14	Train	0.055	0.036	8.9	22.0	11.9	13.7	54.0
DES15	Train	0.023	0.030	1.3	9.0	2.1	10.9	77.4
DES16	Train	0.012	0.124	1.6	1.3	4.6	2.8	37.4
DES17	Train	0.010	0.036	12.5	7.6	16.7	20.1	82.7
DES18	Train	0.003	0.059	2.8	2.0	4.9	1.0	15.8
DES19	Train	0.046	0.064	1.1	1.4	0.8	3.8	29.6
DES20	Test	0.092	0.053	2.4	1.5	1.8	6.2	79.5
DES21	Train	0.061	0.065	12.6	25.2	10.4	15.5	7.8
DES22	Test	0.030	0.026	4.7	10.4	2.6	7.9	13.8
DES23	Test	0.026	0.069	8.1	5.2	4.5	4.1	66.2
DES24	Train	0.029	0.032	1	4.7	4.6	3.6	55.2
DES25	Test	0.052	0.092	21.5	25.5	25.7	19.5	19.1
DES26	Train	0.007	0.054	2.5	11.6	8.8	4.0	36.6
DES27	Train	0.013	0.059	4.5	4.6	2.7	1.3	18.9
DES28	Train	0.104	0.029	5.9	3.6	3.7	4.2	30.9
DES29	Test	0.023	0.062	8.7	1.3	4.0	4.5	26.4
DES30	Train	0.018	0.105	3.9	6.8	2.6	2.7	39.3
DES31	Train	0.011	0.147	3.3	15.4	31.7	25.5	36.0
DES32	Train	0.017	0.0246	7.1	6.6	27.3	25.3	39.1
DES33	Train	0.018	0.216	7.7	13.1	29.0	19.5	22.7
DES34	Train	0.021	0.0459	2.1	11.6	26.1	18.9	23.5
DES35	Train	0.016	0.0313	9.2	3.1	24.3	17.8	36.8
DES36	Train	0.042	0.050	21.2	2.6	28.0	20.8	53.8
DES37	Test	0.05	0.075	6	2.2	12.8	13.7	56.8
DES38	Train	0.062	0.035	4.8	2.6	7.8	13.1	75.8
Average error	Test	0.053	0.059	–	–	–	–	–
	Train	0.024	0.038	–	–	–	–	–
	Correlated	–	–	5.67	9.52	9.38	9.45	38.83

Table 7. ARD% values of ANN + GC, ML + GC, and five correlation-based models.

It is evident that all selected input variables have a strong influence on the speed of sound values, with importance levels ranging from 21 to 29%. However, it is important to note that highly nonlinear models or coupled input variables can complicate sensitivity analysis. These results highlight which inputs have the most significant impact on the output, aiding in model refinement, feature selection, or providing insight into the underlying process. As shown in Fig. 15, the contributions are typically normalized to sum to 1 (or 100%) to facilitate easier interpretation of the results.

In summary, ANN methods have several key advantages and disadvantages⁷⁷. They can model complex nonlinear relationships by selecting an appropriate architecture through trial and error. Once the input layer, the number of neurons, and hidden layers are established, ANNs can predict values beyond those considered during training without reprogramming. However, acquiring large datasets is often challenging and time-consuming.

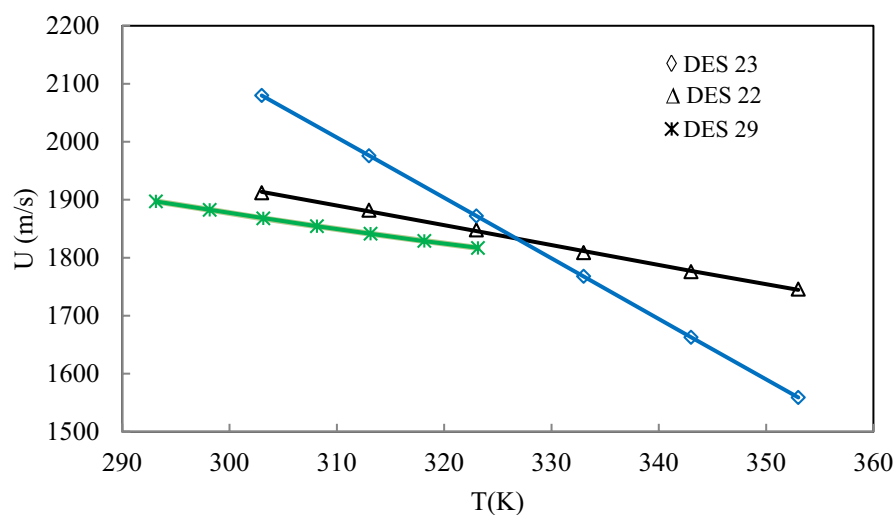


Fig. 9. Prediction of speed of sound of DESs using ANN+GC approach. Lines are model prediction and symbols are experimental data. Standard uncertainty of DES speed of sound is 1.0 m/s.

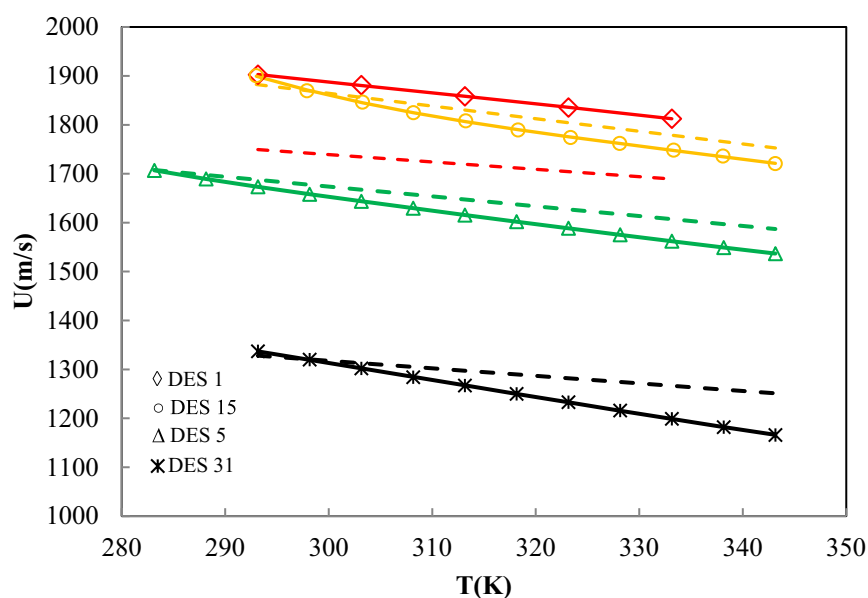


Fig. 10. Correlation of speed of sound of DESs using ANN+GC approach (lines) and Peyrovedin et al. model (dashed-lines). Symbols are experimental data. Standard uncertainty of DES speed of sound is 1.0 m/s.

Additionally, the complexity of ANNs can make their implementation difficult. Another drawback is that ANNs require robust central processing units (CPUs) or hardware, which can be resource-intensive. This study demonstrates the strong performance of ANN models in predicting second-order derivative thermodynamic properties, such as the speed of sound in DESs, despite the aforementioned limitations. Traditionally, equations of state (EoS) models have been widely used to estimate the thermo-physical properties of complex systems like ILs and DESs. However, predicting the speed of sound using EoS-based models requires the ideal gas heat capacity of the pure components. Estimating this property using GC models often results in significant deviations in some cases. Consequently, researchers are seeking alternative approaches to predict the speed of sound and specific heat capacity without relying on ideal gas heat capacity estimations. This work shows that the ANN+GC method can be considered a robust and efficient alternative, particularly for predicting second-order derivative thermodynamic properties, such as the speed of sound.

Conclusions

In this work, the Group Contribution (GC) approach was employed to estimate the input variables for the ANN model. Critical properties and the acentric factor were determined based on the molecular structure of DESs, utilizing the Lydersen and Joback–Reid GC methods. The combined ANN+GC model was developed to predict

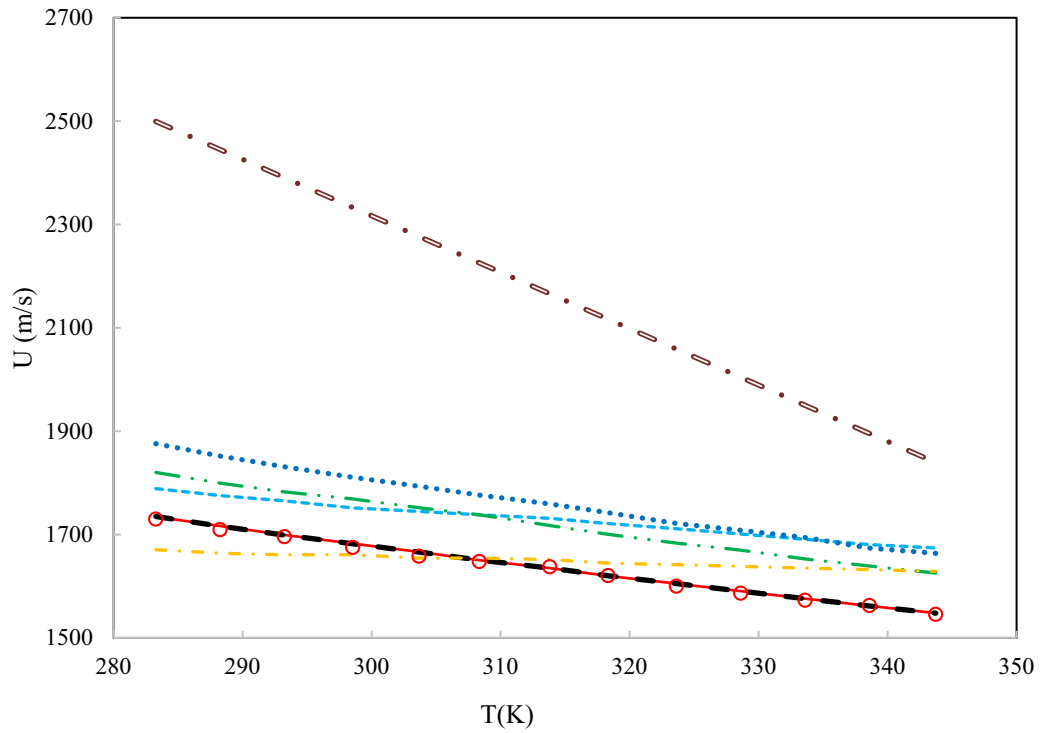


Fig. 11. Comparison of the behavior of the speed of sound of DES 4 versus the temperature for the ANN+GC model, ML+GC model, and literature models. (-) ANN+GC, (--) ML+GC, (- - -) H. Peyrovedin et al.²⁴, (-.-) Haghbakhsh et al.'s model⁷², (...) Hekayati and Esmaeilzadeh's model⁷¹, (-.-) Gardas and Coutinho's model⁷³, and (=.=) Singh and Singh's model⁷⁰. Symbols are experimental data.

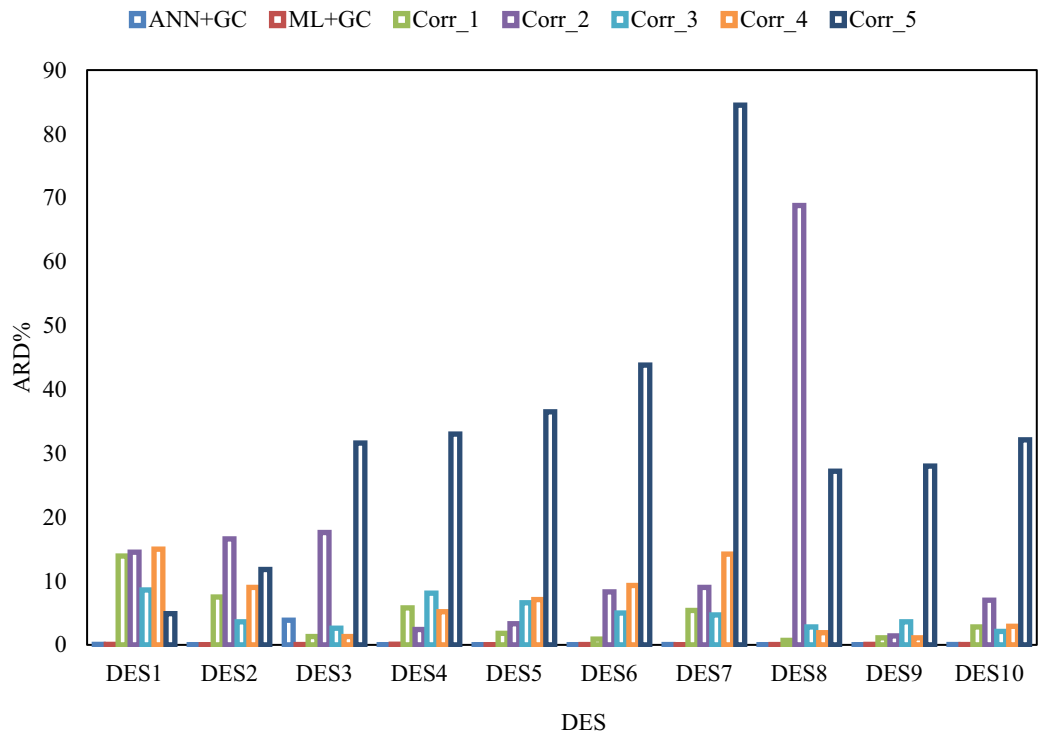


Fig. 12. Error distribution plot for ten DESs. Corr_1, Corr_2, Corr_3, Corr_4, and Corr_5 refer to H. Peyrovedin et al.²⁴, Haghbakhsh et al.'s model⁷², Hekayati and Esmaeilzadeh's model⁷¹, Gardas and Coutinho's model⁷³, and Singh and Singh's model⁷⁰.

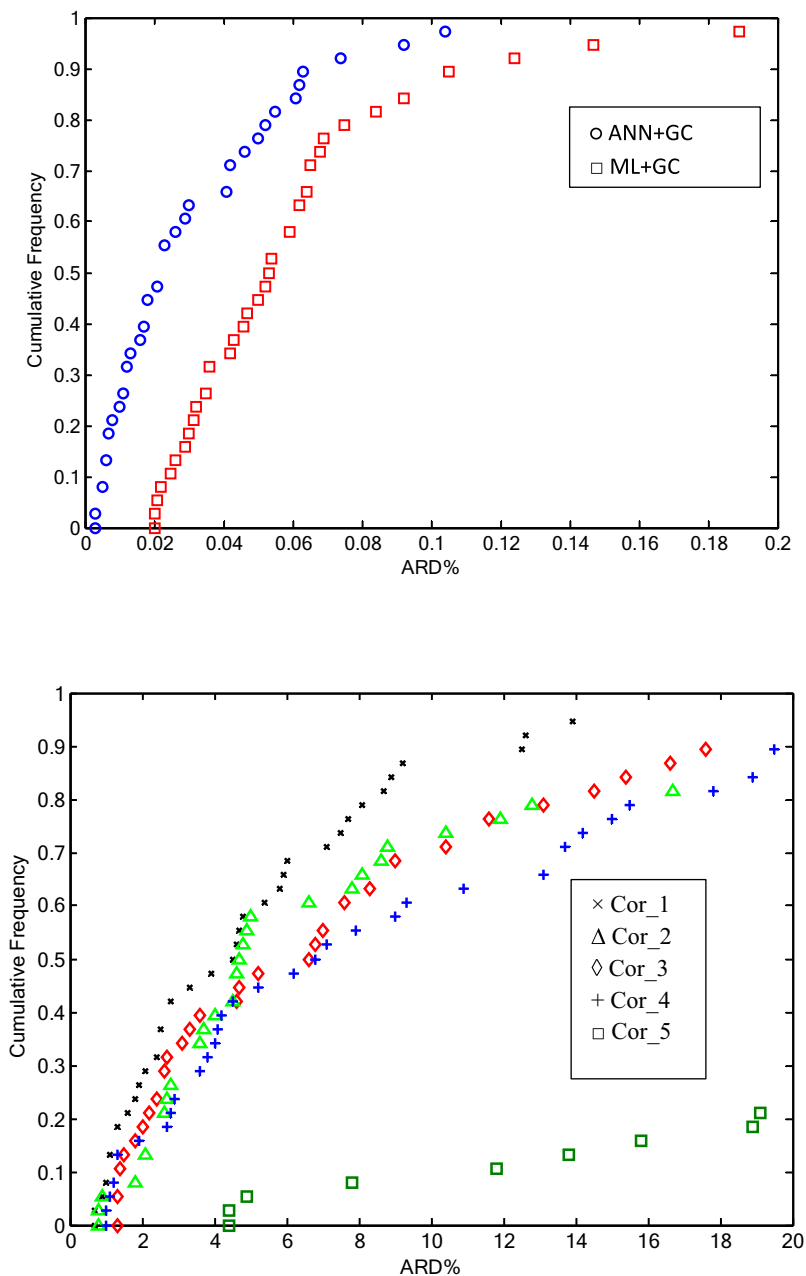


Fig. 13. Cumulative frequency plot for all studied DESs. (a) ANN + GC and ML + GC methods, (b) Corr_1, Corr_2, Corr_3, Corr_4, and Corr_5 refer to H. Peyrovedin et al.²⁴, Haghbakhsh et al.'s model⁷², Hekayati and Esmailzadeh's model⁷¹, Gardas and Coutinho's model⁷³, and Singh and Singh's model⁷⁰, respectively.

the speed of sound in DESs across a wide temperature range. For model development, 415 data points from 38 DESs were selected. The results indicate that a single hidden layer with sixteen neurons provides optimal values for ARD% and R^2 . The model's performance was evaluated using several metrics: average relative deviation percent (ARD%), standard deviation (SD), mean absolute error (MAE), root mean square deviation (RMSE), and R^2 . The findings demonstrate that the ANN + GC model can accurately estimate the speed of sound in DESs over a wide temperature range. The obtained values for the metrics were ARD% = 0.032%, SD = 0.0549, MAE = 1.5656, RMSE = 2.227, and R^2 = 0.9988. The model results were compared to five correlation-based models and a Machine Learning (ML) model. Similar to the ANN + GC approach, the GC method was combined with the ML model. The results indicated that the ML + GC and ANN + GC models were comparable; however, the ANN + GC model showed slightly lower error values. Overall, the error metrics for the ANN + GC approach were lower than those of the five correlation-based models. The cumulative frequency diagrams and the leverage approach were implemented to validate the quality and reliability of the proposed model. The leverage analysis confirmed the accuracy of the data used and the high reliability of the ANN + GC model for estimating the speed of sound in DESs. The primary goal of this study was to predict the speed of sound in DESs based solely on their

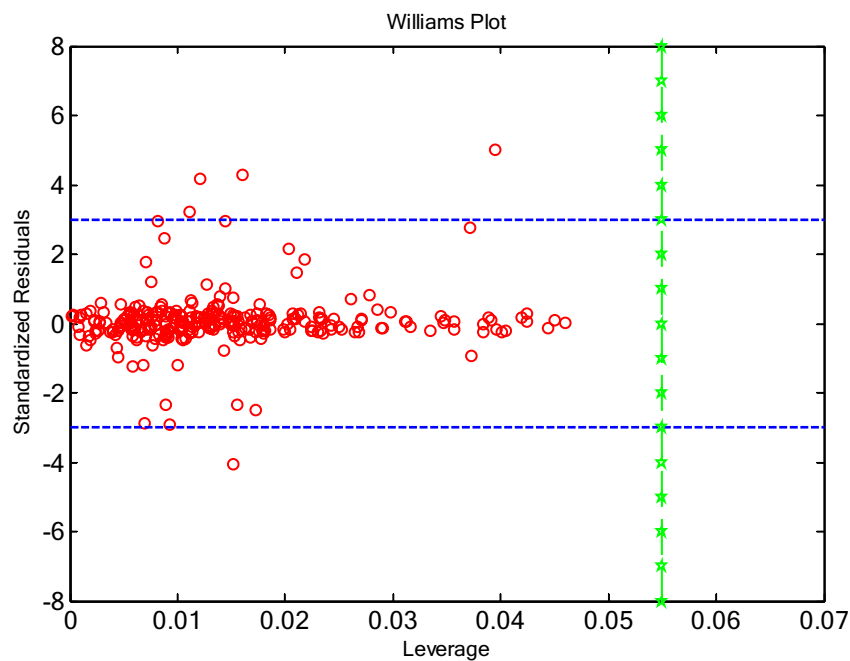


Fig. 14. The Williams plots for outlier detection using the ANN + GC model.

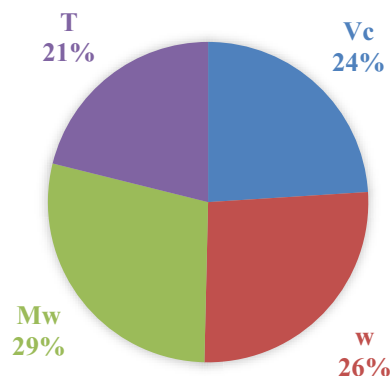


Fig. 15. Relative importance (%) of input variables on the value of the speed of sound of DESs.

molecular structure, without relying on any experimental data. This work demonstrates that the ANN + GC method can serve as a robust model for predicting second-order derivative thermodynamic properties of DESs in the absence of experimental measurements. The proposed model could be employed in future studies, particularly in the pre-design of new solvents.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 2 December 2024; Accepted: 29 July 2025

Published online: 10 August 2025

References

- Alhadid, A., Mokrushina, L. & Minceva, M. Design of deep eutectic systems: A simple approach for preselecting eutectic mixture constituents. *Molecules* **25**, 1077 (2020).
- Lomba, L. et al. Deep eutectic solvents: Are they safe?. *Appl. Sci.* **11**, 10061 (2021).
- Emami, S. & Shayanfar, A. Deep eutectic solvents for pharmaceutical formulation and drug delivery applications. *Pharm. Dev. Technol.* **25**, 779–796. <https://doi.org/10.1080/10837450.2020.1735414> (2020).
- Chen, Y. et al. Capture of toxic gases by deep eutectic solvents. *ACS sustainable chemistry & engineering* **8**, 5410–5430 (2020).
- Maleki, B., Reza, T., Ali, K. & Ahmadpoor, F. Facile protocol for the synthesis of 2-Amino-4H-chromene derivatives using choline chloride/urea. *Org. Prep. Proced. Int.* **53**, 34–41. <https://doi.org/10.1080/00304948.2020.1833623> (2020).

6. Esfahani, H. S., Khoshsima, A., Pazuki, G. & Hosseini, A. Separation of methanol and ethanol from azeotropic MTBE mixtures using choline chloride-based deep eutectic solvents. *J. Mol. Liq.* **381**, 121641. <https://doi.org/10.1016/j.molliq.2023.121641> (2023).
7. Esfahani, H. S., Khoshsima, A. & Pazuki, G. Choline chloride-based deep eutectic solvents as green extractant for the efficient extraction of 1-butanol or 2-butanol from azeotropic n-heptane + butanol mixtures. *J. Mol. Liq.* **313**, 113524. <https://doi.org/10.1016/j.molliq.2020.113524> (2020).
8. Florindo, C., Monteiro, N. V., Ribeiro, B. D., Branco, L. C. & Marrucho, I. M. Hydrophobic deep eutectic solvents for purification of water contaminated with Bisphenol-A. *J. Mol. Liq.* **297**, 111841. <https://doi.org/10.1016/j.molliq.2019.111841> (2020).
9. Atilhan, M. & Aparicio, S. A review on the thermal conductivity of deep eutectic solvents. *J. Therm. Anal. Calorim.* **148**, 8765–8776. <https://doi.org/10.1007/s10973-023-12280-4> (2023).
10. Bowen, H. et al. Application of deep eutectic solvents in protein extraction and purification. *Front. Chem.* <https://doi.org/10.3389/fchem.2022.912411> (2022).
11. Yu, Y. et al. Green recycling of end-of-life photovoltaic modules via deep-eutectic solvents. *Chem. Eng. J.* **499**, 155933. <https://doi.org/10.1016/j.cej.2024.155933> (2024).
12. Jahanbakhsh-Bonab, P., Khoshnazar, Z., Sardroodi, J. J. & Heidaryan, E. A computational probe into the physicochemical properties of cyclodextrin-based deep eutectic solvents for extraction processes. *Carbohydr. Polym. Technol. Appl.* **8**, 100596. <https://doi.org/10.1016/j.carpta.2024.100596> (2024).
13. Jahanbakhsh-Bonab, P., Pazuki, G., Sardroodi, J. J. & Dehnavi, S. M. Assessment of the properties of natural-based chiral deep eutectic solvents for chiral drug separation: Insights from molecular dynamics simulation. *Phys. Chem. Chem. Phys.* **25**, 17547–17557. <https://doi.org/10.1039/D3CP00875D> (2023).
14. Jahanbakhsh-Bonab, P., Esrafil, M. D., Rastkar Ebrahimzadeh, A. & Jahanbin Sardroodi, J. Exploring the structural and transport properties of glycine DES-Based boron nitride nanotube Nanofluid: The effects of nanotube diameter. *J. Mol. Liquids* **341**, 117277. <https://doi.org/10.1016/j.molliq.2021.117277> (2021).
15. Jahanbakhsh-Bonab, P., Sardroodi, J. J. & Avestan, M. S. Electric field effects on the structural and dynamical properties of a glycine deep eutectic solvent. *J. Chem. Eng. Data* **67**, 2077–2087. <https://doi.org/10.1021/acs.jced.2c00066> (2022).
16. Jahanbakhsh-Bonab, P., Esrafil, M. D., Rastkar Ebrahimzadeh, A. & Jahanbin Sardroodi, J. Are choline chloride-based deep eutectic solvents better than methyl diethanolamine solvents for natural gas Sweetening? Theoretical insights from molecular dynamics simulations. *J. Mol. Liquids* **338**, 116716. <https://doi.org/10.1016/j.molliq.2021.116716> (2021).
17. Jahanbakhsh-Bonab, P., Jahanbin Sardroodi, J. & Sadegh Avestan, M. The pressure effects on the amine-based DES performance in NG sweetening: Insights from molecular dynamics simulation. *Fuel* **323**, 124249. <https://doi.org/10.1016/j.fuel.2022.124249> (2022).
18. Jahanbakhsh-Bonab, P. & Sardroodi, J. J. Potential of amine-based DES for separation of CO₂ and H₂S from NG: Study of temperature effect. *J. Environ. Chem. Eng.* **11**, 110517. <https://doi.org/10.1016/j.jece.2023.110517> (2023).
19. Jahanbakhsh-Bonab, P., Jahanbin Sardroodi, J. & Heidaryan, E. Understanding the performance of amine-based DESs for acidic gases capture from biogas. *Renewable Energy* **223**, 120069. <https://doi.org/10.1016/j.renene.2024.120069> (2024).
20. Taherzadeh, M., Haghbakhsh, R., Duarte, A. R. C. & Raeissi, S. Estimation of the heat capacities of deep eutectic solvents. *J. Mol. Liq.* **307**, 112940. <https://doi.org/10.1016/j.molliq.2020.112940> (2020).
21. Leron, R. B. & Li, M.-H. Molar heat capacities of choline chloride-based deep eutectic solvents and their binary mixtures with water. *Thermochim. Acta* **530**, 52–57. <https://doi.org/10.1016/j.tca.2011.11.036> (2012).
22. Naser, J., Mjalli, F. S. & Gano, Z. S. Molar heat capacity of selected type III deep eutectic solvents. *J. Chem. Eng. Data* **61**, 1608–1615. <https://doi.org/10.1021/acs.jced.5b00989> (2016).
23. Zhang, K., Li, H., Ren, S., Wu, W. & Bao, Y. Specific heat capacities of two functional ionic liquids and two functional deep eutectic solvents for the absorption of SO₂. *J. Chem. Eng. Data* **62**, 2708–2712. <https://doi.org/10.1021/acs.jced.7b00102> (2017).
24. Peyrovedin, H., Haghbakhsh, R., Duarte, A. R. C. & Raeissi, S. A global model for the estimation of speeds of sound in deep eutectic solvents. *Molecules* **25**, 1626 (2020).
25. Lapeña, D., Lomba, L., Artal, M., Lafuente, C. & Giner, B. Thermophysical characterization of the deep eutectic solvent choline chloride: Ethylene glycol and one of its mixtures with water. *Fluid Phase Equil.* **492**, 1–9. <https://doi.org/10.1016/j.fluid.2019.03.018> (2019).
26. Peng, D. & Minceva, M. Predicting the density and viscosity of deep eutectic solvents at atmospheric and elevated pressures. *Fluid Phase Equilib.* **582**, 114086. <https://doi.org/10.1016/j.fluid.2024.114086> (2024).
27. Lashkarbolooki, M., Hezave, A. Z. & Ayatollahi, S. Artificial neural network as an applicable tool to predict the binary heat capacity of mixtures containing ionic liquids. *Fluid Phase Equilib.* **324**, 102–107. <https://doi.org/10.1016/j.fluid.2012.03.015> (2012).
28. Valavi, M., Dehghani, M. R. & Shahriari, R. Application of modified PHSC model in prediction of phase behavior of single and mixed electrolyte solutions. *Fluid Phase Equilib.* **344**, 92–100. <https://doi.org/10.1016/j.fluid.2013.01.007> (2013).
29. Sun, R. & Dubessy, J. Prediction of vapor–liquid equilibrium and PVTx properties of geological fluid system with SAFT-LJ EOS including multi-polar contribution. Part I: Application to H₂O–CO₂ system. *Geochim. Cosmochim. Acta* **74**, 1982–1998. <https://doi.org/10.1016/j.gca.2010.01.011> (2010).
30. Shahriari, R., Dehghani, M. R. & Behzadi, B. A modified polar PHSC model for thermodynamic modeling of gas solubility in ionic liquids. *Fluid Phase Equilib.* **313**, 60–72. <https://doi.org/10.1016/j.fluid.2011.09.029> (2012).
31. Shahriari, R., Dehghani, M. R. & Behzadi, B. Thermodynamic modeling of aqueous ionic liquid solutions using PC-SAFT equation of state. *Ind. Eng. Chem. Res.* **51**, 10274–10282. <https://doi.org/10.1021/ie3012984> (2012).
32. Shahriari, R. & Dehghani, M. R. New electrolyte SAFT-VR Morse EOS for prediction of solid-liquid equilibrium in aqueous electrolyte solutions. *Fluid Phase Equilib.* **463**, 128–141. <https://doi.org/10.1016/j.fluid.2018.02.006> (2018).
33. Ramdan, D. et al. Prediction of CO₂ solubility in electrolyte solutions using the e-PHSC equation of state. *J. Supercrit. Fluids* <https://doi.org/10.1016/j.supflu.2021.105454> (2021).
34. Khoshsima, A. & Shahriari, R. Molecular modeling of systems related to the biodiesel production using the PHSC equation of state. *Fluid Phase Equilib.* **458**, 58–83. <https://doi.org/10.1016/j.fluid.2017.10.029> (2018).
35. Khoshsima, A. & Dehghani, M. R. Vapor–liquid and liquid–liquid equilibrium calculations in mixtures containing non-ionic glycol ether surfactant using PHSC equation of state. *Fluid Phase Equilib.* **377**, 16–26. <https://doi.org/10.1016/j.fluid.2014.05.041> (2014).
36. Coimbra, P., Duarte, C. M. M. & de Sousa, H. C. Cubic equation-of-state correlation of the solubility of some anti-inflammatory drugs in supercritical carbon dioxide. *Fluid Phase Equilib.* **239**, 188–199. <https://doi.org/10.1016/j.fluid.2005.11.028> (2006).
37. Chapman, W. G., Gubbins, K. E., Jackson, G. & Radosz, M. New reference equation of state for associating liquids. *Ind. Eng. Chem. Res.* **29**, 1709–1721. <https://doi.org/10.1021/ie00104a021> (1990).
38. Chapman, W. G., Gubbins, K. E., Jackson, G. & Radosz, M. New reference equation of state for associating liquids. *Ind. Eng. Chem. Res.* **29**, 1709–1721 (1990).
39. Bülow, M., Ascami, M. & Held, C. ePC-SAFT advanced—Part I: Physical meaning of including a concentration-dependent dielectric constant in the born term and in the Debye–Hückel theory. *Fluid Phase Equilib.* **535**, 112967. <https://doi.org/10.1016/j.fluid.2021.112967> (2021).
40. Joback, K. G. & Reid, R. C. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* **57**, 233–243. <https://doi.org/10.1080/00986448708960487> (1987).
41. Liang, P. & Bose, N. *Neural Network Fundamentals with Graphs, Algorithms, and Applications* (Mac Graw-Hill, 1996).

42. Taskinen, J. & Yliruusi, J. Prediction of physicochemical properties based on neural network modelling. *Adv. Drug Deliv. Rev.* **55**, 1163–1183. [https://doi.org/10.1016/S0169-409X\(03\)00117-0](https://doi.org/10.1016/S0169-409X(03)00117-0) (2003).
43. Ketabchi, S., Ghanadzadeh, H., Ghanadzadeh, A., Fallahi, S. & Ganji, M. Estimation of VLE of binary systems (tert-butanol+ 2-ethyl-1-hexanol) and (n-butanol+ 2-ethyl-1-hexanol) using GMDH-type neural network. *J. Chem. Thermodyn.* **42**, 1352–1355 (2010).
44. Vrigazova, B. The proportion for splitting data into training and test set for the bootstrap in classification problems. *Bus. Syst. Res. J.* **12**, 228–242 (2021).
45. Valderrama, J. O., Sanga, W. W. & Lazzús, J. A. Critical properties, normal boiling temperature, and acentric factor of another 200 ionic liquids. *Ind. Eng. Chem. Res.* **47**, 1318–1330. <https://doi.org/10.1021/ie071055d> (2008).
46. Valderrama, J. O. & Robles, P. A. Critical properties, normal boiling temperatures, and acentric factors of fifty ionic liquids. *Ind. Eng. Chem. Res.* **46**, 1338–1344. <https://doi.org/10.1021/ie0603058> (2007).
47. Zhu, J., Xu, Y., Feng, X. & Zhu, X. A detailed study of physicochemical properties and microstructure of EmimCl-EG deep eutectic solvents: Their influence on SO₂ absorption behavior. *J. Ind. Eng. Chem.* **67**, 148–155. <https://doi.org/10.1016/j.jiec.2018.06.025> (2018).
48. Basaiahgari, A., Panda, S. & Gardas, R. L. Effect of ethylene, diethylene, and triethylene glycols and glycerol on the physicochemical properties and phase behavior of benzyltrimethyl and benzyltributylammonium chloride based deep eutectic solvents at 283.15–343.15 K. *J. Chem. Eng. Data* **63**, 2613–2627. <https://doi.org/10.1021/acs.jced.8b00213> (2018).
49. Basaiahgari, A., Panda, S. & Gardas, R. L. Acoustic, volumetric, transport, optical and rheological properties of Benzyltriethylammonium based deep eutectic solvents. *Fluid Phase Equilib.* **448**, 41–49. <https://doi.org/10.1016/j.fluid.2017.03.011> (2017).
50. Sánchez, P. B., González, B., Salgado, J., José Parajó, J. & Domínguez, Á. Physical properties of seven deep eutectic solvents based on l-proline or betaine. *J. Chem. Thermodyn.* **131**, 517–523. <https://doi.org/10.1016/j.jct.2018.12.017> (2019).
51. Abdel Jabbar, N. M. & Mjalli, F. S. Ultrasonic study of binary aqueous mixtures of three common eutectic solvents. *Phys. Chem. Liquids* **57**, 1–18. <https://doi.org/10.1080/00319104.2017.1385075> (2019).
52. Vuksanović, J., Kijevčanin, M. L. & Radović, I. R. Effect of water addition on extraction ability of eutectic solvent choline chloride+ 1,2-propanediol for separation of hexane/heptane+ethanol systems. *Korean J. Chem. Eng.* **35**, 1477–1487. <https://doi.org/10.1007/s11814-018-0030-z> (2018).
53. Sas, O. G., Fidalgo, R., Domínguez, I., Macedo, E. A. & González, B. Physical properties of the pure deep eutectic solvent, [ChCl]:[Lev] (1:2) DES, and its binary mixtures with alcohols. *J. Chem. Eng. Data* **61**, 4191–4202. <https://doi.org/10.1021/acs.jced.6b00563> (2016).
54. Kuddushi, M., Nangala, G. S., Rajput, S., Tjardar, S. P. & Malek, N. I. Understanding the peculiar effect of water on the physicochemical properties of choline chloride based deep eutectic solvents theoretically and experimentally. *J. Mol. Liquids* **278**, 607–615. <https://doi.org/10.1016/j.molliq.2019.01.053> (2019).
55. Shekaari, H., Zafarani-Moattar, M. T., Mokhtarpour, M. & Faraji, S. Volumetric and compressibility properties for aqueous solutions of choline chloride based deep eutectic solvents and Prigogine–Flory–Patterson theory to correlate of excess molar volumes at T = (293.15 to 308.15) K. *J. Mol. Liquids* **289**, 111077. <https://doi.org/10.1016/j.molliq.2019.11.1077> (2019).
56. Sas, O. G., Castro, M., Domínguez, Á. & González, B. Removing phenolic pollutants using deep eutectic solvents. *Separ. Purif. Technol.* **227**, 115703. <https://doi.org/10.1016/j.seppur.2019.115703> (2019).
57. Abri, A., Babajani, N., Zonouz, A. M. & Shekaari, H. Spectral and thermophysical properties of some novel deep eutectic solvent based on l-menthol and their mixtures with ethanol. *J. Mol. Liquids* **285**, 477–487. <https://doi.org/10.1016/j.molliq.2019.04.001> (2019).
58. Dabiri, M.-S., Hadavimoghaddam, F., Ashoorian, S., Schaffie, M. & Hemmati-Sarapardeh, A. Modeling liquid rate through wellhead chokes using machine learning techniques. *Sci. Rep.* **14**, 6945. <https://doi.org/10.1038/s41598-024-54010-2> (2024).
59. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulín, A. CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems* **31** (2018).
60. Bandpey, A. F., Abdi, J. & Firozjaee, T. T. Improved estimation of dark fermentation biohydrogen production utilizing a robust categorical boosting machine-learning algorithm. *Int. J. Hydrogen Energy* **52**, 190–199 (2024).
61. Sheikhsheaei, A. H., Khoshshima, A. & Zabihzadeh, D. Predicting the heat capacity of strontium-praseodymium oxysilicate SrPr₄(SiO₄)₃O using machine learning, deep learning, and hybrid models. *Chem. Thermodyn. Therm. Anal.* **17**, 100154. <https://doi.org/10.1016/j.ctta.2024.100154> (2025).
62. Meng, Q. et al. A communication-efficient parallel algorithm for decision tree. *Advances in Neural Information Processing Systems* **29** (2016).
63. Abiodun, O. I. et al. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **4** (2018).
64. Zou, J., Han, Y. & So, S.-S. Overview of artificial neural networks. *Artificial neural networks: methods and applications*, 14–22 (2009).
65. Zhang, Z. & Zhang, Z. Artificial neural network. *Multivariate time series analysis in climate and environmental research*, 1–35 (2018).
66. Wu, Y.-C. & Feng, J.-W. Development and application of artificial neural network. *Wirel. Pers. Commun.* **102**, 1645–1656 (2018).
67. Asadollahfardi, G. in *Water Quality Management: Assessment and Interpretation* 77–91 (Springer, 2014).
68. Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11**, 431–441. <https://doi.org/10.1137/0111030> (1963).
69. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **2**, 164–168 (1944).
70. Singh, M. P. & Singh, R. K. Correlation between ultrasonic velocity, surface tension, density and viscosity of ionic liquids. *Fluid Phase Equilib.* **304**, 1–6. <https://doi.org/10.1016/j.fluid.2011.01.029> (2011).
71. Hekayati, J. & Esmaeilzadeh, F. Predictive correlation between surface tension, density, and speed of sound of ionic liquids: Auerbach model revisited. *J. Mol. Liquids* **274**, 193–203. <https://doi.org/10.1016/j.molliq.2018.10.099> (2019).
72. Haghbakhsh, R., Keshtkari, S. & Raeissi, S. Simple estimations of the speed of sound in ionic liquids, with and without any physical property data available. *Fluid Phase Equilib.* **503**, 112291. <https://doi.org/10.1016/j.fluid.2019.112291> (2020).
73. Gardas, R. L. & Coutinho, J. A. P. Estimation of speed of sound of ionic liquids using surface tensions and densities: A volume based approach. *Fluid Phase Equilib.* **267**, 188–192. <https://doi.org/10.1016/j.fluid.2008.03.008> (2008).
74. Dabiri, M.-S. et al. Artificial Intelligence approaches to modeling equivalent circulating density for improved drilling mud management. *ACS Omega* <https://doi.org/10.1021/acsomega.5c02050> (2025).
75. Dabiri, M.-S. et al. Modeling drilling fluid density at high-pressure high-temperature conditions using advanced machine-learning techniques. *Geoenergy Sci. Eng.* **244**, 213369. <https://doi.org/10.1016/j.geoen.2024.213369> (2025).
76. GD, G. *Interpreting neural network connections weights*. *AI Expert*. Vol. 6 46–51 (Miller Freeman, 1991).
77. Valderrama, J. O., Faúndez, C. A. & Vicencio, V. J. Artificial neural networks and the melting temperature of ionic liquids. *Ind. Eng. Chem. Res.* **53**, 10504–10511. <https://doi.org/10.1021/ie5010459> (2014).

Author contributions

Ayat Hussein Adhab: Writing—Original draft and review, Validation, Investigation, Methodology. Morug Salih Mahdi: Writing—Original draft and review, Validation, Investigation, Methodology. Hardik Doshi: Supervi-

sion, Administration, Methodology, Conceptualization Anupam Yadav: Software (Programming, software development; designing computer programs, implementation of the computer code and supporting algorithms). Manjunatha R: Data Curation, Investigation (data collection), writing –review, Validation Sushil Kumar: Data Curation, Investigation (data collection), writing –review, Validation Debasish Shit: Data Curation, Investigation (data collection), writing –review, Validation Gargi Sangwan: Writing—Original draft and review, data collection Aseel Salah Mansoor: Writing—Original draft and review, data collection Usama Kadem Radi: Writing—Original draft and review, data collection Nasr Saadoun Abd: computational, and formal techniques to analyze or synthesize study data.

Declaration

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-14094-w>.

Correspondence and requests for materials should be addressed to H.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025