Heliyon 10 (2024) e28713

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Neural network-based prediction of auto-ignition temperature of ternary mixed liquids

Bingyu Guo [a,*], Zehui Cheng [b], Shuangqi Hu [a,**]

[a] School of Environment and Safety Engineering, North University of China, Taiyuan, 030051, Shanxi, China
[b] School of Software, North University of China, Taiyuan, 030051, Shanxi, China

ARTICLE INFO

Keywords:
Auto-ignition temperature
Quantitative structure-property relationship
Anti-propagation neural networks
One-dimensional convolutional neural network
Ternary mixed liquid

ABSTRACT

Auto-ignition temperature (AIT) is one of the crucial exponents in the design of fire and explosion safety measures. Therefore, in this study, quantitative structure-property relationship approach was used to predict the AIT of ternary hybrid liquids based on molecular structure information. The optimal molecular descriptors were calculated and filtered using Mordred software. Twelve mixing rules were proposed for calculating molecular descriptors of mixtures. A prediction model for the AIT value of binary liquid mixtures was developed, validated and evaluated using a back propagation neural network (BPNN) and a one-dimensional convolutional neural network (1DCNN). The relative contribution and positive and negative correlations between individual molecular descriptors and AIT in the model were interpreted using the shapley additive explanations method. The results show that BPNN and 1DCNN models using mixing rule 1 have the best fitting ability, stability and prediction ability. The determination coefficient of the BPNN and 1DCNN models in the training set were 0.996 and 0.992, the root mean square errors were 3.613 °C and 5.284 °C, the mean absolute errors were 2.483 °C and 4.144 °C, the nash efficiency coefficient was 0.996 and 0.992, respectively, the willmott index was 0.999 and 0.998. and the values of the top three molecular descriptors of relative contribution, $SssCH_2$, SsOH and $SsCH_3$, were negatively correlated with the AIT values. The BPNN and 1DCNN models provide an accurate and reliable method for predicting ternary mixing liquid AIT.

1. Introduction

It is important to have data on the hazardous properties of flammable substances for safe industrial production [1]. Since Moore [2] independently designed a thermal crucible apparatus for the determination of liquid auto-ignition temperature (AIT) values in 1917, a series of experimental devices and methods for determining AIT values have appeared [3,4]. However, the use of experiments to test the AIT values of substances are time-consuming, labor-intensive, and hazardous [5]. Therefore, the establishment of a reliable AIT prediction model has attracted the attention of many scholars.

The QSPR method has been widely used to predict the AIT of substances in recent years. The QSPR has the advantage that it requires only the molecular structure without any experimental properties and kinetic parameters, and has great advantages in modeling [6].

* Corresponding author.
** Corresponding author.
E-mail addresses: sz202114035@st.nuc.edu.cn (B. Guo), hsq@nuc.edu.cn (S. Hu).

https://doi.org/10.1016/j.heliyon.2024.e28713
Received 2 August 2023; Received in revised form 17 March 2024; Accepted 22 March 2024
Available online 27 March 2024

Keshavarz et al. [7] proposed an AIT prediction model based on QSPR theory for the prediction of organic compounds containing energy-containing functional groups nitro, nitrate, nitramine, and peroxide. The model does not require the use of complex computer codes, unusual descriptors, and expert users, but only the molecular structure of organic energy-containing compounds. Nazari et al. [8] proposed a simple model to estimate the AIT values of 109 organic hydroxyl compounds and obtained more satisfactory results. The evaluation showed that the new model has significant reliability.

With the development of artificial intelligence technology, neural network technology has made breakthroughs in computer vision and other fields and has become a hot spot of current research [9]. The application of neural network technology in the modeling study of QSPR can replace manual statistical analysis to improve computational efficiency [6]. Bagheri et al. [10] developed multiple linear regression (MLR) and neural network models for predicting the AIT values of sulfur chemicals and screened the three most suitable molecular descriptors to explain the flammability behavior of this class of combustible substances. Lazzús [11] used a hybrid method of BPNN and particle swarm optimization to predict the AIT values of substances. Borhani et al. [12] predicted AIT values of 813 hydrocarbons from 69 different families by using genetic algorithms-MLR and artificial neural network methods. The resulting model has good predictive ability. He Fan [13] proposed a mixing molecular descriptor to establish a BPNN model for predicting AIT values of mixed liquids. However, domestic and international studies on the AIT of multivariate mixed liquids are still limited. Most studies are still limited to the AIT of pure substances, and the studies on mixed liquids are lagging behind [12–14]. Therefore, it is important to carry out theoretical prediction studies on the AIT of multivariate mixed liquids.

In this paper, the AIT dataset of ternary mixed liquids is expanded on the basis of the above studies, the 1DCNN modelling approach is used for the first time in AIT QSPR studies, and the shapley additive explanations (SHAP) method is applied for the first time to explain the relationship between the molecular descriptors used and AIT. BPNN and 1DCNN modelling algorithms were used to develop a ternary hybrid liquid AIT QSPR model. The main objectives of this study are to develop a new methodology for ternary mixed liquids AIT, including: development of mixing rules applicable to the molecular descriptors of mixtures; development of a QSPR model for predicting ternary mixed liquids AIT; and validation and evaluation of the model, application domain analysis, and interpretation of the relationship between the molecular descriptors used and AIT.

2. Data collection

The AIT data of mixed liquids for this experiment were obtained from the literature [14] and previous work in our laboratory. The data used included 18 pure substances that are widely used in the processing industry, as shown in Table 1. A total of 205 data were measured for the ternary completely miscible mixed liquids, which were divided into 14 groups. The experimental data were obtained according to the ASTM E659-78 test method [15]. Since the reliability of the experimental data directly affects the prediction accuracy of the constructed model, in order to further check the accuracy of this experimental data, the experimental values of the AIT of the pure substances were compared with the values of the international database of ICSCs [16], as shown in Table 1.

Table 1 shows that the MAE of the experimental data and the ICSCs database is 19.89 °C. According to the test standard, the MAE of AIT experiments is allowed to be within 30 °C, and the deviation is mainly due to the differences in experimental instruments and test methods, so the overall experimental data can be considered as real and effective [14]. The "Point out" strategy in QSPR is used to randomly assign the data sets, in which the training and test sets are used for model development and external validation, respectively. In order to assure the validity of the external validation and to prevent the occurrence of chance situations, the test set constitutes 20% of the total data. It is specifically divided into 164 training sets and 41 test sets.

Table 1
Comparison between experimental and ICSCs Database values of AIT for pure substances.

NO.	Pure substances	CAS number	Value of AIT/°C		Absolute error/°C	Relative error/%
			ICSCs database	Experimental value		
1	n-Pentane	109-66-0	260	309	49	18.85
2	n-Octane	111-65-9	220	221	1	0.45
3	Methanol	67-56-1	440	441	1	0.22
4	Ethyl alcohol	64-17-5	400	384	16	4.00
5	n-Propanol	71-23-8	371	383	12	3.23
6	iso-Propyl alcohol	67-63-0	456	437	19	4.17
7	n-Butyl alcohol	71-36-3	345	348	3	0. 72
8	2-Methyl-1-propanol	78-83-1	415	379	36	8.67
9	1-Pentanol	71-41-0	320	302	18	5.63
10	Acetic acid	64-19-7	485	475	10	2.06
11	Propanoic acid	79-09-4	485	460	25	5.15
12	2-Methoxyethanol	109-86-4	285	317	32	11.22
13	2-Ethoxyethanol	110-80-5	235	234	1	0.43
14	2-(2-Methoxyethoxy) ethanol	111-77-3	215	224	9	4.19
15	Methyl acetate	79-20-9	505	465	40	7.92
16	Ethyl acetate	141-78-6	427	448	21	4.92
17	Acetone	67-64-1	465	508	43	9.25
18	Toluene	108-88-3	480	502	22	4.58
	MAE = 19.89 °C			MAPE = 5.31%		

2.1. Molecular descriptor calculation and screening

Molecular descriptor calculation and screening is one of the key steps in the theoretical method of QSPR. The study adopts the software Mordred [17], which can calculate 1824 molecular descriptors for pure compounds, see Table 2. 9 molecular descriptors based on the electronegative topological state index (ETSI) [18] are screened from the 1824 molecular descriptors after screening, see Table 3. ETSI can express both the electronic and topological properties of an atom and the molecular environment to which the atom is subjected, and is now widely used in the prediction studies of material properties [13]. In this paper, the molecular descriptor data of the mixed liquids used are placed in the SI.

2.2. Calculation of mixing molecular descriptors

The current QSPR theory lacks effective molecular descriptors that directly represent mixtures. Therefore, based on the computational mixture molecular descriptor studies that have been presented [19,20], 12 mixing rules were selected for building molecular descriptors for ternary mixed liquids. The formulas are shown in Table 4.

Where $(S_i)_{\mathrm{mix}}$ is the molecular descriptor value of the ternary liquid mixture, S_1, S_2 and S_3 indicates the molecular descriptors for each of the three components of the mixture, and x_1, x_2 and x_3 are the volume fractions of the three components. In this study, QSPR models for AIT prediction of ternary mixing liquids were developed using 12 mixing rules. These models are validated and the QSPR model with optimal accuracy is selected.

3. Model development

Based on the QSPR method, a ternary mixed liquid AIT prediction model was developed using BPNN and 1DCNN. The flowchart of the 2 models in this paper is shown in Fig. 1, in which the preparation of the dataset and the dataset alignment have been completed in the first section of the article Data collection, this chapter is mainly for the establishment of the BPNN and 1DCNN models and other processes afterwards.

In the experiment, the activation function is ReLU [21], the loss function adopts mean square error (MSE), and the optimizer is Adam [22]. The input of the network is 205 groups of mixed molecular descriptors, and the output is the measured value of ternary mixed liquid AIT.

The computational conditions are: computer system: WIN 10; development platform: TensorFlow-gpu2.3.0; programming language: Python 3.7; Pycharm 2021; artificial neural network library: Keras 2.3.1.

The packages used in modelling the BPNN and 1DCNN model were numpy 1.19.2; pandas 1.1.5; sklearn 1.0.2 (scikit-learn); tensorflow 2.3.0; matplotlib 3.4.3; SkillMetrics 1.2.4; shap 0.41.

The hyperparameters of the BPNN and 1DCNN model are shown in Table 5.

Table 2
List of Mordred descriptors.

	Descriptor name	Number	Descriptor name	Number
2D	ABC Index	2	Kappa Shape Index	3
	Acid-Base	2	Lipinski	2
	Adjacency Matrix	13	Mc Gowan Volume	1
	Aromatic	2	Moe Typea	53
	Atom Count	16	Molecular Distance Edge	19
	Autocorrelation	606	MolecularId	12
	BCUTa	24	Path Count	21
	Balaban Ja	1	Polarizability	2
	Barysz Matrixa	104	Ring Count	138
	Bertz CT	1	Rotatable Bonda	2
	Bond Count	9	Slog Pa	2
	Carbon Types	10	Topo PSAa	2
	Chi	56	Topological Charge	21
	Constitutional	16	Topological Index	4
	Detour Matrix	14	Vdw Volume ABC	1
	Distance Matrix	13	Vertex Adjacency Information	1
	EState	316	Walk Count	21
	Eccentric Connectivity Index	1	Weight	2
	Extended Topochemical Atom	45	Wiener Index	2
	Fragment Complexity	1	Zagreb Index	4
	Framework	1	Information Content	42
	Hydrogen Bonda	2		
3D	CPSA	43	MoRSE	160
	Geometrical Index	4	Moment OfInertia	3
	Gravitational Index	4		

Table 3
ETSI symbols for 9 atomic types.

Atomic type	ETSI symbol	Intrinsic state value
O =	SdO	7.000
—OH	SsOH	6.000
—O—	SssO	3.500
—CH_3	SsCH_3	2.000
>C =	SdssC	1.667
—CH_2—	SssCH_2	1.500
>CH—	SsssCH	1.333
aCHa	SaaCH	2.000
asCa	SaasC	1.667

s: single bond; d: double bond; a: a bond in an aromatic ring.

Table 4
12 Mixing rules used for calculation of mixture descriptors.

Number	Mixing rule	Number	Mixing rule
1	$(S_i)_{mix} = x_1S_1 + x_2S_2 + x_3S_3$	7	$(S_i)_{mix} = \sqrt[3]{x_1^3S_1 + x_2^3S_2 + x_3^3S_3}$
2	$(S_i)_{mix} = \sqrt{(x_1S_1)^2 + (x_2S_2)^2 + (x_3S_3)^2}$	8	$(S_i)_{mix} = (x_1S_1 + x_2S_2 + x_3S_3)^2$
3	$(S_i)_{mix} = \sqrt{x_1S_1} + \sqrt{x_2S_2} + \sqrt{x_3S_3}$	9	$(S_i)_{mix} = (S_1 - S_2 - S_3)^2$
4	$(S_i)_{mix} = \sqrt{\lvert S_1\rvert} + \sqrt{\lvert S_2\rvert} + \sqrt{\lvert S_3\rvert}$	10	$(S_i)_{mix} = \lvert S_1 - S_2 - S_3\rvert$
5	$(S_i)_{mix} = x_1^2S_1 + x_2^2S_2 + x_3^2S_3$	11	$(S_i)_{mix} = (S_1 + S_2 + S_3)/2$
6	$(S_i)_{mix} = x_1S_1^2 + x_2S_2^2 + x_3S_3^2$	12	$(S_i)_{mix} = \lvert x_1S_1 - x_2S_2 - x_3S_3\rvert$

3.1. BPNN model

The BP algorithm is a gradient descent algorithm that makes learning of feedforward networks easier [21,22]. Fig. 2 illustrates the framework of the BPNN model and Table 6 shows the structure of BPNN, the input is the 9 molecular descriptor data (Table 3) computed by 12 mixing rules (Table 4). And finally, the predicted value of AIT is obtained after the hidden layer.

The pseudo-code for the BPNN model is shown below:

```
# Importing the required libraries
import numpy as np
import tensorflow as tf
# Preparing the data set (In our experiment the feature dimension = 9)
# Feature data — Values of 9 molecular descriptors
# Target data — Values of AIT
X_train = … # Feature data for training set, shape is (number of training set, feature dimension)
y_train = … # Target data for training set, shape is (number of training set)
X_test = … # Feature data for the test set, shape is (number of test set, feature dimension)
# Creating a BPNN model
model = tf.keras.Sequential([
tf.keras.layers.Dense(units = 128, activation = 'relu', input_shape = 9),
tf.keras.layers.Dense(units = 256, activation = 'relu'),
tf.keras.layers.Dense(units = 128, activation = 'relu'),
tf.keras.layers.Dense(units = 1)
# compilation model
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
# Training Models
model.fit(X_train, y_train, epochs = num1, batch_size = num2)
# Prediction using models
y_pred = model.predict(X_test)
```

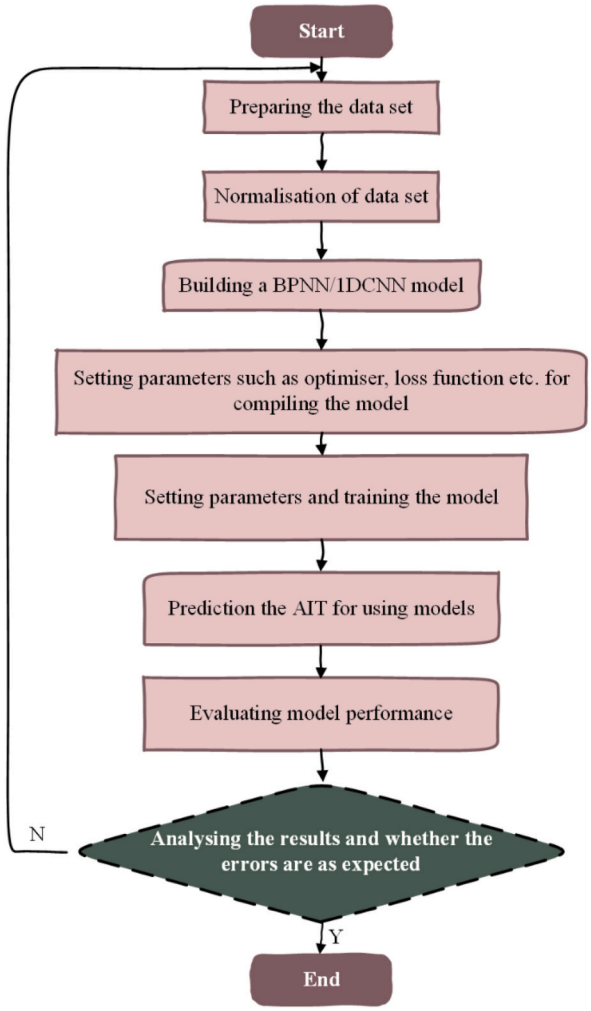

Fig. 1. Flowchart of the model.

Table 5
Hyperparameters of the model.

Hyperparameters	Model	
	BPNN	1DCNN
learning rate	0.002	0.002
batch size	32	32
epochs	5000	5000
kernel size	–	4
filters	–	16
strides	–	1
padding	–	valid

3.2. 1DCNN model

The 1DCNN is usually used for analyzing a fixed-length segment of signal data [23–25], and one-dimensional means that the convolution kernel moves on the input matrix in one dimension only, i.e., it moves from top to bottom. Fig. 3 illustrates the framework of the 1DCNN model and the structure of the 1DCNN, as shown in Table 7, the input is the 9 molecular descriptor data (Table 3) computed by 12 mixing rules (Table 4), and finally, the predicted value of AIT is obtained through the dense layer.

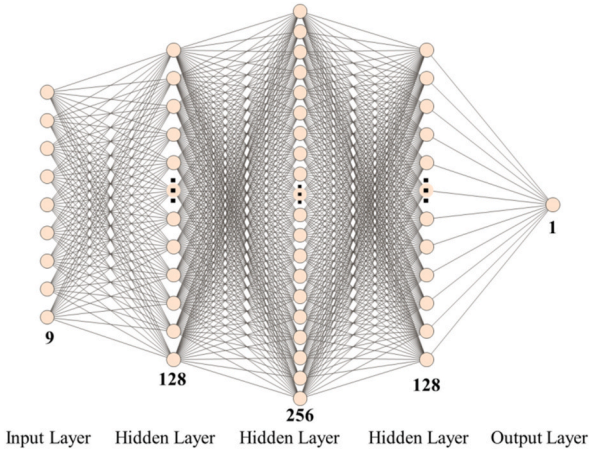

Fig. 2. BPNN model framework diagram.

Table 6
Structure of BPNN.

Layer(type)	Output Shape	Param
Hidden Layer	(None,128)	1280
Hidden Layer	(None,256)	33024
Hidden Layer	(None,128)	32896
Output Layer	(None,1)	129

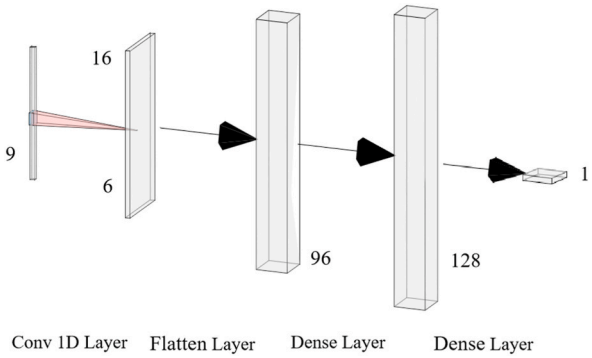

Fig. 3. 1DCNN model framework diagram.

Table 7
Structure of 1DCNN.

Layer(type)	Output Shape	Param
Conv 1D	(None,6,16)	80
Flatten	(None,96)	0
Dense	(None,128)	12416
Output Layer	(None,1)	129

The pseudo-code for the 1DCNN model is shown below:

```
# Importing the required libraries
import numpy as np
from tensorflow. keras. models import Sequential
from tensorflow. keras. layers import Conv1D, Flatten, Dense
# Preparing the data set (In our experiment the time step = 9)
# Feature data — Values of 9 molecular descriptors
# Target data — Values of AIT
X_train = … # Feature data for training set, shape is (number of training set, time step, feature dimension)
y_train = … # Target data for training set, shape is (number of training set)
X_test = … # Feature data for the test set, shape is (number of test set, time step, feature dimension)
# Creating 1DCNN model
model = Sequential(-)
model.add(Conv1D(filters = 16, kernel_size = 4, activation = 'relu', input_shape=(9, 1)))
model.add(Flatten(-))
model.add(Dense(units = 128, activation = 'relu'))
model.add(Dense(units = 1))
# compilation model
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
# Training Models
model.fit(X_train, y_train, epochs = num1, batch_size = num2)
# Prediction using models
y_pred = model.predict(X_test)
```

4. Analysis and discussion of results

4.1. Validation of the model

According to the principles proposed by the Organization for Economic Cooperation and Development (OECD), appropriate methods should be used to measure the model's performance, including goodness-of-fit, robustness, and predictive ability [5]. This study discusses the performance of the model based on the above principles.

4.1.1. Fitting ability of the model

The following equations (Eqs. (1)–(6)) were used to validate the model's fitting ability: determination coefficient (R^2) [26], root mean square errors (RMSE) [27], mean absolute errors (MAE) [28], mean absolute percentage errors were (MAPE) [29], nash efficiency coefficient (NSE) [30], and willmott index (WI) [31]. The specific formula is:

Table 8
Statistical parameters for the models based on 12 mixing rules.

Mixing rule	Models	Training set				Test set			
		R^2	RMSE/°C	MAE/°C	MAPE/%	R^2	RMSE/°C	MAE/°C	MAPE/%
1	BPNN	0.996	3.613	2.483	18.878	0.909	16.092	10.540	15.150
	1DCNN	0.992	5.284	4.144	19.038	0.911	15.906	10.830	15.544
2	BPNN	0.995	4.159	3.050	19.082	0.832	21.871	14.551	15.199
	1DCNN	0.986	6.833	5.392	18.802	0.805	23.571	18.639	15.064
3	BPNN	0.996	4.345	2.484	19.011	0.938	16.264	9.087	15.296
	1DCNN	0.993	5.973	3.619	18.959	0.903	16.589	11.661	15.515
4	BPNN	0.695	32.870	24.511	17.382	0.654	31.404	21.811	15.208
	1DCNN	0.703	32.464	23.643	17.977	0.640	32.023	21.182	15.793
5	BPNN	0.989	6.020	4.033	18.945	0.678	30.293	20.389	15.628
	1DCNN	0.985	7.172	5.518	18.951	0.708	28.832	20.891	15.422
6	BPNN	0.994	4.277	3.238	18.969	0.894	17.341	9.953	15.352
	1DCNN	0.989	6.067	3.967	18.662	0.877	18.700	11.309	15.442
7	BPNN	0.989	6.017	4.537	19.067	0.581	34.540	21.375	15.055
	1DCNN	0.992	4.992	3.617	18.961	0.611	33.288	23.190	15.978
8	BPNN	0.994	4.300	2.937	18.879	0.781	24.973	16.144	15.820
	1DCNN	0.994	4.384	2.755	18.967	0.816	22.887	16.568	15.835
9	BPNN	0.702	32.526	23.873	18.074	0.649	31.599	21.845	15.892
	1DCNN	0.700	32.627	24.032	17.663	0.659	31.153	21.545	15.481
10	BPNN	0.700	32.633	23.920	17.488	0.669	30.715	20.889	15.333
	1DCNN	0.700	32.624	24.007	17.867	0.652	31.500	21.711	15.767
11	BPNN	0.698	32.721	24.323	17.741	0.646	31.768	21.891	15.559
	1DCNN	0.687	33.349	25.219	18.319	0.632	32.372	23.493	16.097
12	BPNN	0.993	4.728	3.619	18.958	0.874	18.942	13.123	15.186
	1DCNN	0.989	6.152	4.861	18.712	0.832	21.824	13.943	14.886

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_{i,exp} - \hat{y}_{i,pre}\right)^2}{\sum_{i=1}^{n} \left(y_{i,exp} - \overline{y}_{exp}\right)^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(y_{i,exp} - y_{i,pre}\right)^2}{n}} \tag{2}$$

$$MAE = \frac{\sum_{i=1}^{n} \left|y_{i,exp} - y_{i,pre}\right|}{n} \tag{3}$$

$$MAPE = \frac{\sum_{i=1}^{n} \left|\left(y_{i,exp} - y_{i,pre}\right)/y_{i,exp}\right|}{n} * 100\% \tag{4}$$

$$NSE = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} \left(y_{i,exp} - y_{i,pre}\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n} \left(y_{i,exp} - \overline{y}_{exp}\right)^2} \tag{5}$$

$$d_{WI} = 1 - \frac{\sum_{i=1}^{n} \left(y_{i,exp} - y_{i,pre}\right)^2}{\sum_{i=1}^{n} \left(\left|y_{i,pre} - \overline{y}_{exp}\right| + \left|y_{i,exp} - \overline{y}_{exp}\right|\right)^2} \tag{6}$$

where $y_{i,exp}$ and $y_{i,pre}$ are the experimental and predicted values for mixture i. $\hat{y}_{i,pre}$ is the simulated value of mixture i of the model. $\overline{y}_{exp}$ is the average of the experimental values. n is the amount of data in the dataset. $y_{i,pre}$ is the predicted value of the variable by the predictive model. The closer the values of R^2, NSE, and WI of the model are to 1, and the smaller the values of RMSE, MAE, and MAPE are, the better the fit of the model is.

The statistical parameters of the model under each mixing rule are calculated from Eqs. (1)–(4). As known from Table 8, the R^2 in the training set of BPNN models under mixing rules 1, 2, 3, 6, 8, and 12 are all higher than 0.99 and have low RMSE, MAE, and MAPE, and among these models, the training set of BPNN models under mixing rule 1 has the highest R^2 as well as the lowest RMSE, MAE, and MAPE. In the test set, only mixing rules 1 and 3 have an R^2 exceeds 0.9, although the R^2 of the test set of mixing rule 3 is higher than that of mixing rule 1, its RMSE and MAPE are higher than those of the test set of rule 1, which indicates that the fitting accuracy of the BPNN model under mixing rule 3 is not as good as that of the BPNN model under mixing rule 1.

As we know from Table 8, the R^2 in the training set of 1DCNN models under mixing rules 1, 3, 7, and 8 are all higher than 0.99 and have low RMSE, MAE, and MAPE, and among these models, the 1DCNN model under mixing rule 1 has higher R^2 as well as the lowest RMSE. In the test set, only the R^2 of mixing rules 1 and 3 exceeds 0.9 and the R^2 of mixing rule 1 is higher than that of mixing rule 3, the RMSE and MAE of mixing rule 3 are higher than those of rule 1, which indicates that the fitting accuracy of the 1DCNN model under mixing rule 3 is not as good as that of the 1DCNN model under mixing rule 1.

In summary, mixing rule 1 was chosen as the best mixing descriptor computation rule and was used as the input parameter for both models.

Comparing the AIT predictions obtained from the BPNN model with the experimental values (Fig. 4), it is found that most of the

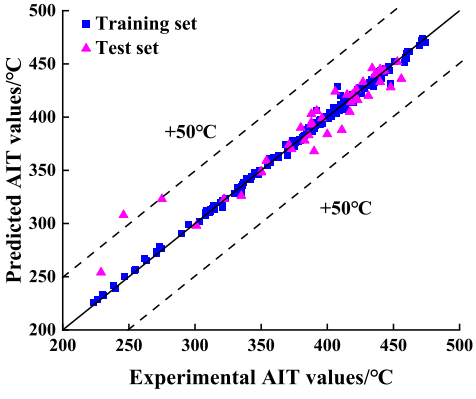

Fig. 4. Plot of predicted versus experimental values of AIT (BPNN).

Table 9
Validation results of fitting ability of models.

Model	Training set						Test set					
	R^2	RMSE/°C	MAE/°C	MAPE/%	NSE	WI	R^2	RMSE/°C	MAE/°C	MAPE/%	NSE	WI
BPNN	0.996	3.613	2.483	18.878	0.996	0.999	0.909	16.092	10.540	15.150	0.876	0.973
1DCNN	0.992	5.284	4.144	19.038	0.992	0.998	0.911	15.906	10.830	15.544	0.890	0.975
MLR [34]	0.879	22.011	13.001	–	–	–	–	16.336	11.227	–	–	–
MNR [34]	0.880	21.912	12.984	–	–	–	–	16.502	11.045	–	–	–

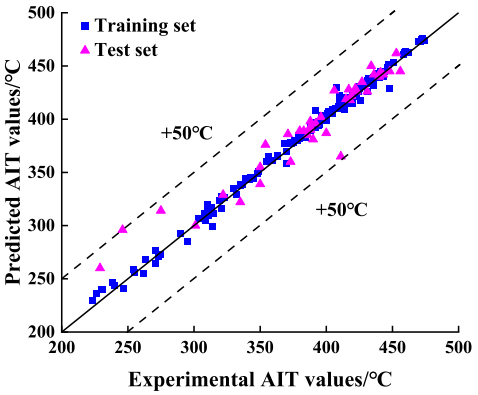

Fig. 5. Plot of predicted versus experimental values of AIT (1DCNN).

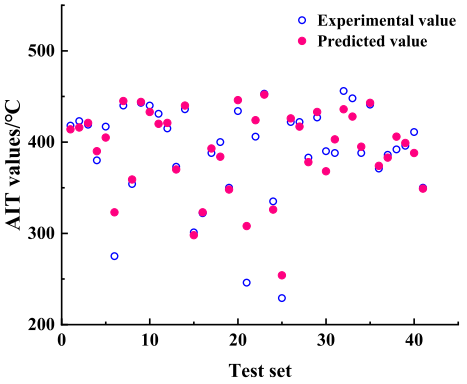

Fig. 6. Comparison of experimental and predicted values for the test set of the BPNN model plots.

data points fall on the diagonal line in the graph, and only individual data points are seriously off the diagonal line(**>** 50 °C), which indicates that the prediction accuracy of the BPNN model is good. The statistical parameters of the BPNN and 1DCNN models are calculated from Eqs. (1)–(6). The results of the calculations are shown in Table 9. In Table 9, the MAE, RMSE, and MAPE of the training and test sets of the BPNN model are less than the experimental standard error (±30 °C), indicating that the BPNN model in this paper has a good fitting ability [17,26].

Fig. 5 shows that the comparative values of 1DCNN model are essentially the same as for the BPNN model, indicating that the overall prediction of the two models is satisfactory. The validation results of the 1DCNN model are shown in Table 9. The MAE, RMSE, and MAPE of the model's training and test sets are all less than the experimental standard error (±30 °C).

On the test set, the results of the comparison between the predicted and experimental values of the two models AIT are shown in Figs. 6 and 7 The experimental values are represented by hollow circles and the predicted values are represented by solid circles, and the higher the overlap indicates that the predicted values are closer to the experimental values. From the figure, it is found that the

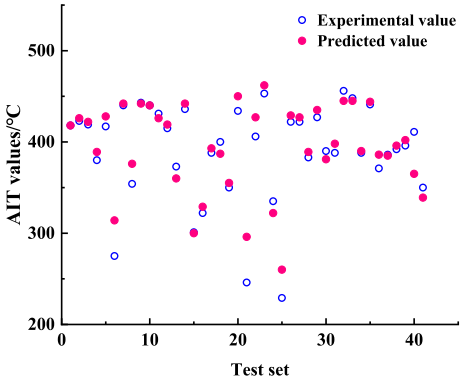

Fig. 7. Comparison of experimental and predicted values for the test set of the 1DCNN model plots.

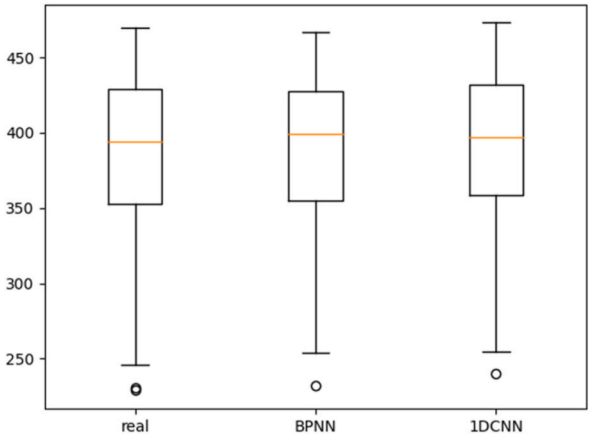

Fig. 8. Box-line diagram of the model.

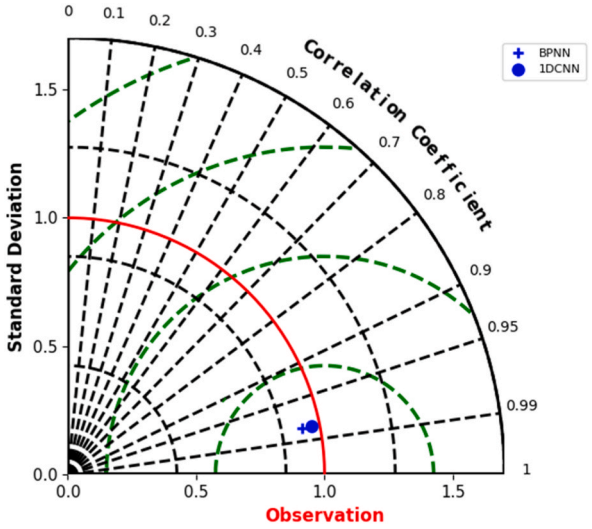

Fig. 9. Taylor diagram of the model.

experimental and predicted values of the two models have a high degree of overlap, although there is a small part of no overlap but the deviation is not large, which indicates that the prediction effect of the two models is good.

In order to validate the prediction results, a box-and-line diagram [32] is plotted as shown in Fig. 8. From the figure, it can be seen that the BPNN and 1DCNN models are roughly the same as the experimental values, but the maximum value (upper limit) predicted by the BPNN model is slightly lower than the experimental value, and the maximum value (upper limit) predicted by the 1DCNN model is slightly higher than the experimental value. Secondly the median and lower edges of the BPNN and 1DCNN models are a bit higher than the real situation, indicating that the AIT values predicted by both are mostly higher than the experimental values.

The Taylor diagrams [33] of the predicted and observed AIT data for the BPNN and 1DCNN models are shown in Fig. 9. In the figure, the round centre corner is the correlation coefficient, the green arc represents the root mean square error, the black arc represents the rms error, and Observation represents the location of the observation (i.e., the true AIT value in the test set). The higher the correlation coefficient, the closer the standard deviation is to 1, and the closer the root mean square error is to 0, the better the prediction of the model. In layman's terms the closer to the red line and the lower down it is the better the model's prediction. It can be found that the BPNN and 1DCNN models are roughly at the same height, but the 1DCNN model is closer to the red line, indicating that the prediction effect of the 1DCNN model is better than that of the BPNN model.

The results of the two models for predicting the AIT are approximately the same, but the combined R^2, RMSE, MAE and MAPE results show that the RMSE and MAE in the BPNN model training set are smaller than those in the training set of the 1DCNN model, which indicates that the BPNN model has better fitting ability. In order to further verify the superiority of the two models in this paper, a comparison is made with the MLR and multiple nonlinear regression (MNR) pairs in the literature [34], and the performance parameters are shown in Table 9. At the same time, the four models are applied to predict the same experimental values respectively, and the results are shown in Table 10. According to Tables 9 and 10, it can be seen that the R^2 of both MLR and MNR models on the training set is lower than that of BPNN and 1DCNN models, and the values of RMSE and MAE are higher than those of BPNN and 1DCNN models. Comparing the predicted values of the models with the experimental values in Table 10 shows that the difference between the predicted and experimental values of MLR and MNR is significantly higher than that of the BPNN and 1DCNN models.

In summary, the prediction performance of the BPNN model and the 1DCNN model in this paper is significantly better than the MNR and MLR models.

4.1.2. Stability of the model

The statistical indicators used in this study to test the stability of the model are: K-fold cross-validation, Y-randomization, and residual analysis.

(1) K-fold cross-validation method

The stability of the model can be verified by applying cross-validated correlation coefficients(Q^2_{cv})and mean square error (MSEK-f) [35]. Q^2_{cv} calculated as shown in Eq. (7):

$$Q^2_{cv} = \frac{1}{k}\sum_{j=1}^{k}\left(1 - \frac{\sum_{i=1}^{n}\left(y_{i,exp} - \widehat{y}_{i,pre}\right)^2}{\sum_{i=1}^{n}\left(y_{i,exp} - \overline{y}_{exp}\right)^2}\right) \tag{7}$$

If the difference between the results of Q^2_{cv} and R^2 is > 0.3, it indicates the possibility of overfitting the model [36]. The models 5 and 10fold cross-validated correlation coefficients are calculated from Eq. (7). The results of the cross-validation analysis of the two models with 5 and 10 folds are shown in Table 11. The difference between Q^2_{cv} and R^2 did not exceed 0.3, indicating that the models were not overfitted.

(2) Y-randomization

The y-randomization method is to build a new prediction model using new independent and dependent variable correspondences [37]. The modeling was repeated 50 times to compare the prediction results with those of the original model. The maximum R^2_{y-rand} of the new model is 0.343 and 0.238 which are significantly lower than the R^2 of the original model, and the experimental results show that the two models are stable.

(3) Residual analysis

Figs. 10 and 11 show the residual plots of AIT values prediction for the BPNN and 1DCNN models, respectively. The scatter points in both plots are randomly dispersed on both sides of the zero line, which indicates that there is no systematic error in the modeling process of the two models [38].

The above statistical indicators show that both models have good stability. Comparing the difference between R^2 and 5-fold and 10-fold Q^2_{cv}, it can be inferred that the BPNN model (0.093, 0.142) has better stability than the 1DCNN model (0.126, 0.182).

4.1.3. External predictive ability of the model

The external predictive ability of the model is measured by the coefficient of cross-validation (Q^2_{ext}) between the predicted and

Table 10
Predicted value of AIT of the model.

NO.	Mixtures composition(A + B + C)	volume fraction			Experimental value/°C	Predicted value/°C			
		A	B	C		BPNN	1DCNN	MNR [34]	MLR [34]
1	Acetone + n-Butyl alcohol + Ethyl alcohol	0.1	0.7	0.2	374	378	378	371	371
		0.5	0.4	0.1	394	393	394	403	401
		0.7	0.1	0.2	435	429	436	431	429
2	Methanol + Ethyl alcohol + iso-Propyl alcohol	0.1	0.7	0.2	423	424	423	418	417
		0.3	0.5	0.2	427	433	435	418	417
		0.5	0.4	0.1	436	437	437	431	426
		0.7	0.1	0.2	444	446	445	429	429
3	Ethyl alcohol + Acetone + Ethyl acetate	0.3	0.3	0.4	434	430	437	437	437
		0.5	0.1	0.4	422	420	425	431	431
		0.7	0.1	0.2	417	417	418	410	409
4	n-Pentane + Toluene + iso-Propyl alcohol	0.1	0.7	0.2	321	315	316	268	271
		0.3	0.5	0.2	315	313	311	327	326
		0.5	0.4	0.1	311	310	310	290	293
		0.7	0.1	0.2	335	326	322	337	339
5	Acetone + 2-Methoxyethanol + n-Octane	0.3	0.5	0.2	336	336	339	338	338
		0.5	0.4	0.1	371	374	386	373	376
		0.7	0.1	0.2	396	394	400	390	388
6	Toluene + Methanol + 2-Ethoxyethanol	0.1	0.7	0.2	345	348	344	344	342
		0.3	0.3	0.4	309	312	305	310	312
		0.5	0.1	0.4	262	267	255	266	269
		0.7	0.1	0.2	246	308	296	251	252
7	Acetic acid + iso-Propyl alcohol + 2-Methoxyethanol	0.1	0.7	0.2	392	406	396	404	402
		0.3	0.3	0.4	335	329	339	338	339
		0.5	0.1	0.4	380	378	382	438	439
8	Acetic acid + Methyl acetate + 2-Methoxyethanol	0.3	0.5	0.2	407	403	410	443	444
		0.5	0.1	0.4	381	381	383	380	377
		0.7	0.1	0.2	432	433	438	446	448
9	Propanoic acid + 2-(2-methoxyethoxy) ethanol + n-Octane	0.3	0.3	0.4	263	265	268	276	280
		0.5	0.1	0.4	303	302	307	323	324
		0.7	0.1	0.2	354	359	376	353	351
10	Acetic acid + Methyl acetate + n-Propanol	0.1	0.7	0.2	456	436	445	452	454
		0.5	0.4	0.1	470	467	473	464	468
		0.7	0.1	0.2	462	462	463	461	460
11	Propanoic acid + 1-pentanol + 2-Ethoxyethanol	0.3	0.5	0.2	363	364	366	361	359
		0.5	0.4	0.1	383	378	389	380	379
		0.7	0.1	0.2	390	390	395	389	386

Table 11
Stability and predictive ability verification results of BPNN and 1DCNN models.

Model	Q^2_{5-FCV}	Q^2_{10-FCV}	$MSE_{5\text{-}f}$	$MSE_{10\text{-}f}$	R^2_{y-rand}	Q^2_{ext}
BPNN	0.903	0.854	1.837E-4	1.717E-4	0.343	0.998
1DCNN	0.866	0.810	2.492E-4	1.817E-4	0.238	0.998

experimental values of the test set data [39], as shown in Eq. (8):

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}} \left(y_{i,exp} - y_{i,pre}\right)^2}{\sum_{i=1}^{n_{ext}} \left(y_{i,exp} - \bar{y}_{tr}\right)^2} \tag{8}$$

where n_{ext} is the number of mixtures in the test set and $\bar{y}_{tr}$ is the average of the experimental values in the training set. The Q^2_{ext} for the BPNN model and the 1DCNN model is calculated from Eq. (8) The results of the prediction ability of the two models are shown in Table 11, which indicate that both models have reliable external prediction ability.

4.2. Application domain analysis of the model

Leverage method was applied to determine the model application domain [5]. Figs. 12 and 13 show the application domain of the two models, most of the data points in the dataset fall on the application domain (±3 standardized residuals and to the left of the standardized leverage value h*). Although there are some points outside the application domain, they are all located to the left of h*. The reason these mixtures are outliers may be due to anomalies in the experimental data rather than structural anomalies. In summary, the predicted AIT values of these mixed combustible liquids are reliable.

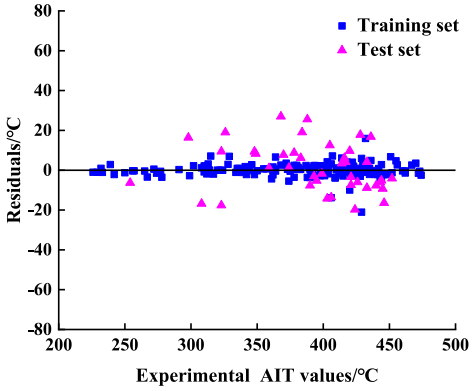

Fig. 10. Residual plot of BPNN model for predicting AIT.

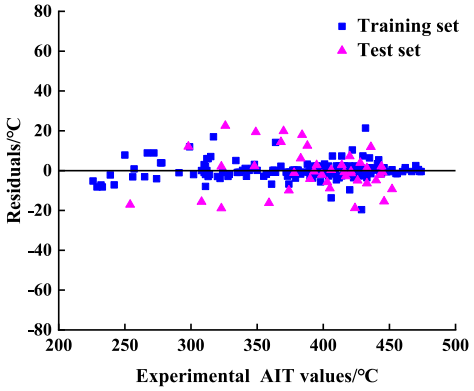

Fig. 11. Residual plot of 1DCNN model for predicting AIT.

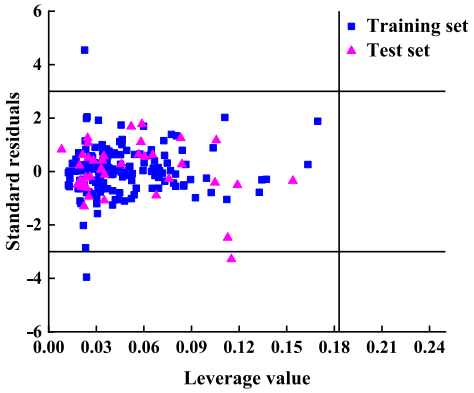

Fig. 12. Analysis diagram of the application domain of BPNN model($h^* = 0.183$).

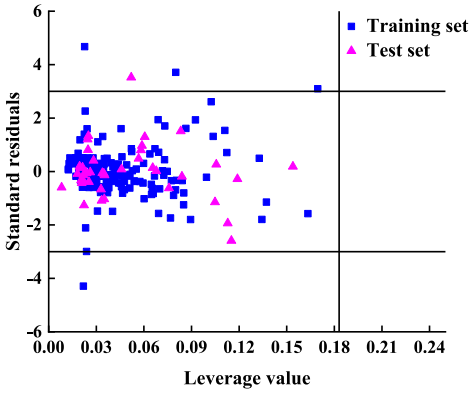

Fig. 13. Analysis diagram of the application domain of 1DCNN model(h* = 0.183).

4.3. Comparison of the models obtained

From the above section 3.1, it can be seen that the performance parameters of the BPNN model and the 1DCNN model constructed in this experiment are basically the same on the training set and the test set, and the 2 models show very strong performance in terms of fitting ability, stability and external prediction. However, in terms of fitting ability, the R^2 of the training set of the BPNN model is higher than that of the 1DCNN model, and the RMSE, MAE and MAPE of the training set of the BPNN model are lower than those of the 1DCNN model, which indicates that the fitting ability of the BPNN model is comparatively better than that of the 1DCNN model; and from the perspective of the error, in terms of RMSE, the BPNN model is relatively better than the 1DCNN model, indicating that the BPNN model has better external prediction ability; from the perspective of model stability, the $Q^2_{5\text{-FCV}}$ and $Q^2_{10\text{-FCV}}$ of the BPNN model are higher than that of the 1DCNN model, indicating that the stability of the BPNN model is relatively better than that of the 1DCNN model.

In summary, both 2 models are able to successfully complete the work of predicting the spontaneous combustion temperature of mixed combustible liquids. However, in terms of the predictive ability and stability of the models, the BPNN model is stronger than the 1DCNN model, which is due to the stronger nonlinear modelling ability and adaptability of BPNN, for different types of data and problems, BPNN is able to adaptively adjust the weights and biases, and improve the generalization ability of the model [21]. And in terms of the architecture of the model, the architecture of the BPNN model is relatively simple, and the architecture of the 1DCNN model is more complex. However, in the future, with the increase of the mixed-liquid spontaneous combustion temperature dataset, the training process of BPNN may be more time-consuming when dealing with such a large-scale dataset, especially in the deep network structure, and its ability is slightly insufficient, which is due to the fact that it needs to traverse the whole dataset to update the weights several times. The 1DCNN, on the other hand, is more capable of parallel computing and will appear more efficient in handling large-scale datasets.

4.4. Mechanistic explanation of the model

This study uses the SHAP method, which can take into account not only the effects on individual variables but also the possible

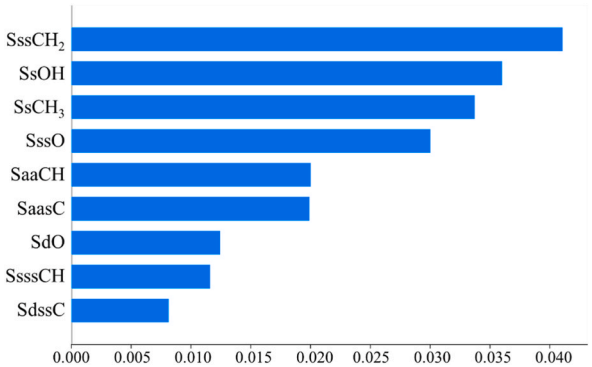

Fig. 14. Average importance plot based on SHAP values.

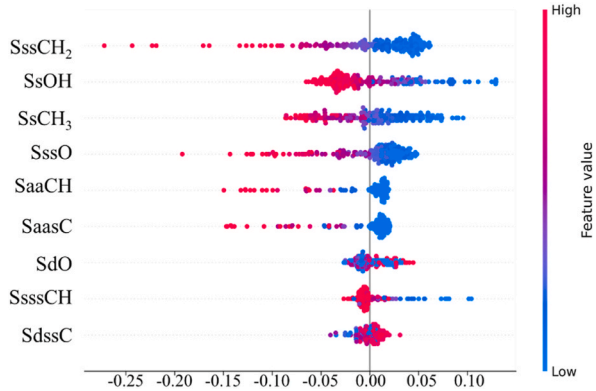

Fig. 15. SHAP summary plot of input variables.

synergistic effects between the influencing variables of a group of variables [40]. The core idea of SHAP is to compute the marginal contribution of the features to the output of the model, and then to explain the "black-box model" at both the global and local levels [41]. Fig. 14 demonstrate the relative importance of each molecular descriptor of the test set on the AIT.$SssCH_2$ and SsOH are commonly found in alkanes and alcohols, which are the basic skeleton structures of alkanes and alcohols organic compounds, and their high importance indicates that $SssCH_2$ and SsOH have a large influence on the AIT of the ternary mixing system of alkanes and alcohols, and the model has the best prediction accuracy for the ternary mixing system of alkanes and alcohols; $SsCH_3$ is commonly found in various organic compounds with high importance, but it appears in all compounds in the dataset, so it is not used as the main reference data; SaaCH and SaasC are commonly found in benzene, with low importance, which indicates that the influence of SaaCH and SaasC on the AIT of ternary mixed system of benzene is relatively small, and the prediction effect of this model on the ternary mixed system of benzene needs to be improved; SsssCH is commonly found in the isomers of alcohols, and its low importance indicates that it has little effect on the AIT of ternary mixed systems containing alcohol isomers, and the predictive effect of this model on the AIT of ternary mixed systems containing alcohol isomers needs to be improved; SdO and SdssC are commonly found in ketones and acids, and the molecular descriptors are of low importance, indicating that SdO and SdssC have little effect on the AIT of the ternary mixed system of ketones and acids, respectively, and that the predictive effect of this model on the ternary mixed system containing ketones and acids needs to be further strengthened.

Fig. 15 shows the summary plot of the SHAP features of the BPNN model test set. Taking the first three molecular descriptors with greater relative importance as an example, the lower molecular descriptor values (blue dots in the figure) are mostly distributed on the side with positive SHAP values, and the higher molecular descriptor values (red dots in the figure) are mostly distributed on the side with negative SHAP values, which indicates the molecular descriptor $SssCH_2$, SsOH and $SsCH_3$ are negatively correlated with the AIT values, and as the values of the molecular descriptors $SssCH_2$, SsOH and $SsCH_3$ decrease in the ternary mixed combustible liquid system, the corresponding AIT values will increase. Similarly, the three molecular descriptors SssO, SaaCH and SaasC, which are ranked in the middle of the relative importance list, are negatively correlated with the AIT values. On the other hand, the relative importance of SdO, SsssCH and SdssC are small and have no decisive influence on the prediction results of the final model, so they are not analysed.

5. Conclusions

Two QSPR prediction models, BPNN and 1DCNN, were established based on neural network techniques. Among the 12 mixing rules selected for building ternary hybrid liquid molecular descriptors, mixing rule 1 outperforms the other mixing rules. The validation results show that the R^2 in the training set of the BPNN model and the 1DCNN model are 0.996 and 0.992, the RMSE is 3.613 °C and 5.284 °C, the $Q^2_{10\text{-}FCV}$ is 0.854 and 0.810, and the Q^2_{ext} of the two models is 0.998, respectively. both models have good fitting ability, stability, and external prediction ability. Combining the parameter criteria, the fitting ability and stability of the BPNN model are better than that of the 1DCNN model, and the performance of the two models is significantly better than that of the existing MLR and MNR models. The BPNN model has better predictive ability compared to the 1DCNN model, but the 1DCNN model will be more efficient when dealing with large-scale AIT datasets in the future. The range of applicability of the two models was determined based on the Leverage method. The SHAP tool was used to explain the mechanism of the BPNN model with better prediction performance, and the effects of nine atom types on the AIT of ternary mixtures of liquids were summarized, and the values of the top three molecular descriptors, $SssCH_2$, SsOH and $SsCH_3$, were negatively correlated with the value of AIT. This study can not only provide a new method for the safety design of chemical enterprises by predicting the AIT values of ternary mixed liquids, but also provide some ideas for the subsequent prediction of the AIT of quaternary and quintuple mixed flammable liquids.

This study needs to do more exploration in the future: (1) The current study mainly focuses on the ternary mixed system, and in the future, the ternary mixed system should be further improved and gradually extended to the study of the quaternary and above multivariate mixed system. This will this contribute to a more comprehensive understanding of the changing law of AIT. (2) In order to better understand the interaction mechanism between single components within a mixed system, the screening and calculation

methods of the hybrid molecular descriptor should be further optimized, so as to better understand and explain the results of the model. (3) The lack of data sources also hinders the AIT study of mixed liquids to a certain extent, and in the future, more AIT experimental data of mixed liquids should be measured to support the in-depth development of the QSPR study of mixed liquids.

Data availability statement

Data available on request from the authors.

The data that support the findings of this study are available from the corresponding author, [Bingyu Guo], upon reasonable request.

CRediT authorship contribution statement

Bingyu Guo: Writing – review & editing, Writing – original draft, Visualization, Validation, Data curation. **Zehui Cheng:** Validation, Software, Methodology, Formal analysis. **Shuangqi Hu:** Writing – review & editing, Data curation.

Declaration of competing interest

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

References

[1] M.E. Redd, C.J. Guffey, E.L. Gustafson, E.H. Hart, K.S. McQuade, N.F. Giles, T.A. Knotts IV, W.V. Wilding, Autoignition temperature trends for various chemical families, J. Fuel. 335 (2024) 1–9, https://doi.org/10.1016/j.fuel.2023.129321.
[2] H. Moore, Sponmneous ignition temperature of liquid fuels for internal combustion engines, J. Journal of the Chemical Society of London 36 (1917) 109–112.
[3] B. Nazari, M.H. Keshavarz, M.H. Mobarhan, The simplest method for reliable prediction of autoignition temperature of organic hydroxyl compounds to assess their process safety in industrial applications, J. Process Safety and Environmental Protection 148 (2021) 283–290, https://doi.org/10.1016/j.psep.2020.10.017.
[4] D. Liu, W. Hu, Y. Zhang, H. Liu, Z. Zheng, M. Yao, On the entropy generation and exergy loss of laminar premixed flame under engine-relevant conditions, J. Fuel. 283 (2021) 11–24, https://doi.org/10.1016/j.fuel.2020.119245.
[5] Y. Pan, J. Jiang, R. Wang, H. Cao, Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds, J. Chemometrics and Intelligent Laboratory Systems 92 (2) (2008) 169–178, https://doi.org/10.1016/j.chemolab.2008.03.002.
[6] Y.T. Jin, J.C. Jiang, Y. Pan, L. Ni, Prediction of the auto-ignition temperature of binary liquid mixtures based on the quantitative structure-property relationship approach, J. Journal of thermal analysis and calorimetry 140 (1) (2020) 397–409, https://doi.org/10.1007/s10973-019-08774-9.
[7] M.H. Keshavarz, M. Jafari, K. Esmaeilpour, M. Samiee, New and reliable model for prediction of autoignition temperature of organic compounds containing energetic groups, J. Process Safety and Environmental Protection 113 (2017) 491–497, https://doi.org/10.1016/j.psep.2017.12.001.
[8] B. Nazari, M.H. Keshavarz, M.H. Mobarhan, The simplest method for reliable prediction of autoignition temperature of organic hydroxyl compounds to assess their process safety in industrial applications, J . Process Safety and Environmental Protection 148 (2021) 283–290, https://doi.org/10.1016/j.psep.2020.10.017.
[9] J.R. Zhang, C.Q. Li, T.T. Zhang, An assessment of the mobility of toxic elements in coal fly ash using the featured BPNN model, J . Sustainability 15 (23) (2023) 1–18, https://doi.org/10.3390/su152316389.
[10] M. Bagheri, T. Borhani, G. Zahedi, Estimation of flash point and autoignition temperature of organic sulfur chemicals, J. Energy Conversion and Management 58 (9) (2012) 185–196, https://doi.org/10.1016/j.enconman.2012.01.014.
[11] J. Lazzús, Neural network-particle swarm modeling to predict thermal properties, J . Mathematical and Computer Modelling 57 (s9–10) (2013) 2408–2418, https://doi.org/10.1016/j.mcm.2012.01.003.
[12] T.N.G. Borhani, A. Afzali, M. Bagheri, QSPR estimation of the auto-ignition temperature for pure hydrocarbons, J. Process Safety and Environmental Protection 103 (2016) 115–125, https://doi.org/10.1016/j.psep.2016.07.004.
[13] F. He, J.C. Jiang, Y. Pan, H.L. Chen, Forecasting the auto-ignition tempera-tures of binary liquid mixtures based on BP neural network, J. Journal of Safety and Environment 17 (4) (2017) 1343–1348. CNKI:SUN:AQHJ.0.2017-04-029.
[14] L.T. Ye, Y. Pan, J.C. Jiang, Experimental determination and calculation of auto-ignition temperature of binary flammable liquid mixtures, J. Acta petrolei sinica (Petroleum processing section) 31 (3) (2017) 753–759, https://doi.org/10.1016/j.psep.2017.09.012.
[15] J. Lan, J.C. Jiang, Y. Pan, D. Zhang, Q.S. Wang, Experimental measurements and numerical calculation of auto-ignition temperatures for binary miscible liquid mixtures, J. Process Safety & Environmental Protection 113 (2018) 22–29, https://doi.org/10.1016/j.psep.2017.09.012.
[16] K. Qu, S. Garamszegi, F. Wu, H. Thorvaldsdottir, T. Liefeld, M. Ocana, D.B. Rivera, N. Pochet, J.T. Robinson, B. Demchak, T. Hull, G.B. Artzi, D. Blankenberg, G. P. Barber, B.T. Lee, R.M. Kuhn, A. Nekrutenko, E. Segal, T. Ideker, M. Reich, A. Regev, H.Y. Chang, J.P. Mesirov, Integrative genomic analysis by interoperation of bioinformatics tools in genomespace, J. Nature Methods 13 (2016) 245–247, https://doi.org/10.1038/nmeth.3732.
[17] H. Moriwaki, Y.S. Tian, N. Kawashita, T. Takagi, Mordred: a molecular descriptor calculator, J. Journal of Cheminformatics 10 (1) (2018) 4–17, https://doi.org/10.1186/s13321-018-0258-y.
[18] L.H. Hall, L.B. Kier, Electro-topological state indices for atom types: a novel combination of electro-topological and valence state information, J. Journal of Chemical Information and Computer Sciences 35 (6) (1995) 1039–1045, https://doi.org/10.1021/ci00028a014.
[19] T. Gaudin, P. Rotureau, G. Fayet, Mixture descriptors toward the development of quantitative structure-property relationship models for the flash points of organic mixtures, J. Industrial and Engineering Chemistry Research 54 (2015) 6596–6604, https://doi.org/10.1021/acs.iecr.5b01457.
[20] L.T. Qin, Y.H. Chen, X. Zhang, L.Y. Mo, H.H. Zeng, Y.P. Liang, QSAR prediction of additive and non-additive mixture toxicities of antibiotics and pesticide, J. Chemosphere. 198 (2018) 122–129, https://doi.org/10.1016/j.chemosphere.2018.01.142.
[21] F.Y. Wang, X.Y. Wang, D.L. Liu, H. Liu, Comprehensive safety risk evaluation of fireworks production enterprises using the frequency-based ANP and BPNN, J. Heliyon 9 (11) (2023) e21724, https://doi.org/10.1016/j.heliyon.2023.e21724.
[22] J.H. Shu, Y.L. Zhao, Y.H. Zhou, F.F. Lin, J.M. Song, X.H. Li, Optimization of tetrastigma hemsleyanum extraction process based on GA-BPNN model and analysis of its antioxidant effect, J. Heliyon 9 (10) (2023) e20200, https://doi.org/10.1016/j.heliyon.2023.e20200.
[23] S.Z. Gao, T.C. Li, Y.M. Zhang, Z.M. Pei, Fault diagnosis method of rolling bearings based on adaptive modified CEEMD and 1DCNN model, J . ISA transactions 140 (2023) 309–330, https://doi.org/10.1016/j.isatra.2023.05.014.
[24] F. Asadi, T. Angsuwatanakul, J.A. O'Reilly, Evaluating synthetic neuroimaging data augmentation for automatic brain tumour segmentation with a deep fully-convolutional network, J. IBRO Neuroscience Reports 16 (6) (2024) 57–66, https://doi.org/10.1016/j.ibneur.2023.12.002, 2024.

[25] K. Yao, A. Ortiz, F. Bonnin-Pascual, A DCNN-based arbitrarily-oriented object detector with application to quality control and inspection, J .Computers in Industry 142 (2022), https://doi.org/10.1016/j.compind.2022.103737, 103737-103737.
[26] C.Y. Lin, Y.C.E. Yang, S. Wi, HydroCNHS: a Python package of hydrological model for coupled natural–human systems, J . J of Water Resources Plan and Management 148 (12) (2022), https://doi.org/10.1061/(ASCE)WR.1943-5452.0001630, 06022005-06022005.
[27] E. Towler, D. Woodson, S. Baker, M. Ge, J. Prairie, B. Rajagopalan, S. Shanahan, R. Smith, Incorporating mid-term temperature predictions into streamflow forecasts and operational reservoir projections in the Colorado river basin, J . J of Water Resources Plan and Management 148 (4) (2022), https://doi.org/10.1061/(ASCE)WR.1943-5452.0001534.
[28] P. Pandey, N.D. Bokde, S. Dongre, R. Gupta, Hybrid models for water demand forecasting, J . J of Water Resources Plan and Management 147 (2) (2020), https://doi.org/10.1061/(ASCE)WR.1943-5452.0001331, 2020.
[29] J.H. Li, D. Zhu, C.X. Li, Comparative analysis of BPNN, SVR, LSTM, Random Forest, and LSTM-SVR for conditional simulation of non-Gaussian measured fluctuating wind pressures, J. Mechanical Systems & Signal Processing 178 (2022), https://doi.org/10.1016/j.ymssp.2022.109285, 109285-109285.
[30] J.R. Lamontagne, C.A. Barber, R.M. Vogel, Improved estimators of model performance efficiency for skewed hydrologic data, J . Water Resources Res 56 (9) (2020), https://doi.org/10.1029/2020WR027101.
[31] W.G. Sun, D. Fleisher, D. Timlin, S.A. Li, Z.J. Wang, S. Beegum, V. Reddy, Evaluation of models for simulating soybean growth and climate sensitivity in the U.S. Mississippi Delta, J. Eur J of Agronomy 140 (2022), https://doi.org/10.1016/j.eja.2022.126610, 126610-126610.
[32] Y. Song, Q. Jin, J. Qiu, D. Ye, A systematic review and meta-analysis on the correlation between HIV infection and multidrug-resistance tuberculosis, Heliyon 9 (11) (2023) e21956, https://doi.org/10.1016/j.heliyon.2023.e21956.
[33] F. Cheng, Q. Yang, C.w. Liu, B. Han, S.J. Peng, G.H. Hao, Evaluating parameterizations for turbulent fluxes over the landfast sea-ice surface in prydz bay, Antarctica, J, Adv. Atmos. Sci. 40 (10) (2023) 1816–1832, https://doi.org/10.1017/jog.2022.8.
[34] Y. Zhang, in: D. Jiangsu (Ed.), Experimental Research and Theoretical Prediction of Auto-Ignition Temperature of Polybasic Combustible Mixed Solution, Nanjing Tech University, 2016.
[35] T.S. Wiens, B.C. Dale, M.S. Boyce, G.P. Kershawet, Three way k-fold cross-validation of resource selection functions, J. Ecological Modelling 212 (3–4) (2008) 244–255, https://doi.org/10.1016/j.ecolmodel.2007.10.005.
[36] R. Kiralj, M. Ferreira, Basic validation procedures for regression models in QSAR and QSPR studies: theory and application, J. Journal of the Brazilian Chemical Society 20 (4) (2009) 770–787, https://doi.org/10.1590/S0103-50532009000400021.
[37] C. Rucker, G. Rucker, M. Meringer, y-Randomization and its variants in QSPR/QSAR, J. Journal of Chemical Information and Modeling 47 (6) (2007) 2345–2357, https://doi.org/10.1021/CI700157B.
[38] A. Tropsha, P. Cramatica, V.K. Comhar, The importance of being earnest: validation is the absolule essential for suceesf/ul application andinterpretation of QSPR models, J. Qsar and Combinatorial Science 22 (1) (2003) 69–77, https://doi.org/10.1002/qsar.200390007.
[39] V. Consonni, D. Ballabio, R. Todeschini, Comments on the defi-nition of the Q2 parameter for QSAR validation, J. Journal of Chemical Information and Modeling 49 (7) (2009) 1669–1678, https://doi.org/10.1021/ci900115y.
[40] R. Mitchell, E. Frank, G. Holmes, GPUTreeShap: massively parallel exact calculation of SHAP scores for tree ensembles, J. PeerJ Computer Science 8 (5) (2022) 880–898, https://doi.org/10.7717/peerj-cs.880.
[41] J.F. Jia, Z.X. Chen, Y.L. Bai, Y.L. Yu, An interpretable ensemble learning method to predict the compressive strength of concrete, J. Structures 46 (2022) 201–213, https://doi.org/10.1016/j.istruc.2022.10.056.