



## OPEN ACCESS

## EDITED BY

Fengqi You,  
Cornell University, United States

## REVIEWED BY

José Ezequiel Santibañez-Aguilar,  
Tecnologico de Monterrey, Mexico  
Brenno Menezes,  
Hamad bin Khalifa University, Qatar  
Chun Deng,  
China University of Petroleum, Beijing,  
China

## \*CORRESPONDENCE

Zuwei Liao,  
liaoZW@zju.edu.cn

## SPECIALTY SECTION

This article was submitted to  
Computational Methods in Chemical  
Engineering,  
a section of the journal  
Frontiers in Chemical Engineering

RECEIVED 30 June 2022

ACCEPTED 18 July 2022

PUBLISHED 08 September 2022

## CITATION

Ren Y, Liao Z, Yang Y, Sun J, Jiang B,  
Wang J and Yang Y (2022), Direct  
prediction of steam cracking products  
from naphtha bulk properties:  
Application of the two sub-  
networks ANN.

*Front. Chem. Eng.* 4:983035.  
doi: 10.3389/fceng.2022.983035

## COPYRIGHT

© 2022 Ren, Liao, Yang, Sun, Jiang,  
Wang and Yang. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Direct prediction of steam cracking products from naphtha bulk properties: Application of the two sub-networks ANN

Yu Ren<sup>1,2</sup>, Zuwei Liao<sup>2\*</sup>, Yao Yang<sup>1,2</sup>, Jingyuan Sun<sup>1</sup>,  
Binbo Jiang<sup>1</sup>, Jingdai Wang<sup>1</sup> and Yongrong Yang<sup>1</sup>

<sup>1</sup>State Key Laboratory of Chemical Engineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, China, <sup>2</sup>ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, China

Steam cracking of naphtha is an important process for the production of olefins. Applying artificial intelligence helps achieve high-frequency real-time optimization strategy and process control. This work employs an artificial neural network (ANN) model with two sub-networks to simulate the naphtha steam cracking process. In the first feedstock composition ANN, the detailed feedstock compositions are determined from the limited naphtha bulk properties. In the second reactor ANN, the cracking product yields are predicted from the feedstock compositions and operating conditions. The combination of these two sub-networks has the ability to accurately and rapidly predict the product yields directly from naphtha bulk properties. Two different feedstock composition ANN strategies are proposed and compared. The results show that with the special design of dividing the output layer into five groups of PIONA, the prediction accuracy of product yields is significantly improved. The mean absolute error of 11 cracking products is 0.53wt% for 472 test sets. The comparison results show that this indirect feedstock composition ANN has lower product prediction errors, not just the reduction of the total error of the feedstock composition. The critical factor is ensuring that PIONA contents are equal to the actual values. The use of an indirect feedstock composition strategy is a means that can effectively improve the prediction accuracy of the whole ANN model.

## KEYWORDS

naphtha, steam cracking, artificial neural network, modeling, product prediction

## 1 Introduction

With the concept of industry 4.0 (BMBF, 2011), intelligent manufacturing has gradually penetrated various industrial fields. Artificial intelligence (AI) is an integral part of this industrial revolution. The term “Artificial intelligence” was coined in 1956 at a math conference at Dartmouth College (Venkatasubramanian, 2019), which means the ability of machines to perform tasks that are generally linked to the behavior of intelligent

beings, such as humans (Dobbelaere et al., 2021). Machine learning, an advanced technique to realize artificial intelligence, is widely used in various aspects of chemical engineering, and it is especially suitable for solving highly complex and nonlinear problems (Plehiens et al., 2019). For example, the current boom in using machine learning to predict quantum chemical properties (Schütt et al., 2017; Yalamanchi et al., 2020). With well-trained machine learning models, bond dissociation enthalpies of organic molecules were predicted at near chemical accuracy with sub-second computational cost (St John et al., 2020). Another example is the application of machine learning for catalyst design and discovery (Goldsmith et al., 2018). Using a combination of machine learning and optimization, a fully automated screening method was constructed to rapidly identify candidate catalyst types (Tran and Ulissi, 2018). In addition, machine learning is also widely used in chemical process engineering, such as plastic pyrolysis (Armenise et al., 2021), Fischer-Tropsch synthesis (Chakkingal et al., 2022), and ethylene thermal cracking (Bi et al., 2021), the topic of this paper.

Steam cracking of naphtha is an important chemical process for olefins production. Naphtha cracking furnace is the core device of this process, where hydrocarbons undergo highly complex cracking reactions through the free radical mechanism to produce olefin products. The schematic diagram of a typical steam cracking furnace is illustrated in Figure 1. The furnace is divided into two sections: the convection section and the radiant section. Naphtha feedstock is first introduced into the convection section for preheating (Zhang et al., 2016). Dilution Steam with a specific ratio of hydrocarbon is added to the feedstock to increase its temperature and lower the hydrocarbon partial pressure (Fakhroleslam and Sadrameli, 2019). The diluted feedstock enters the fired tubular reactor to convert into smaller molecules under controlled residence time, temperature profile, and pressure. The required heat of the radiant reaction is provided via several burners by the fuel source. After leaving the radiant section, the cracked gas is cooled rapidly and indirectly in the transfer line exchanger (TLE) to stop the undesired reactions (Sadrameli, 2015).

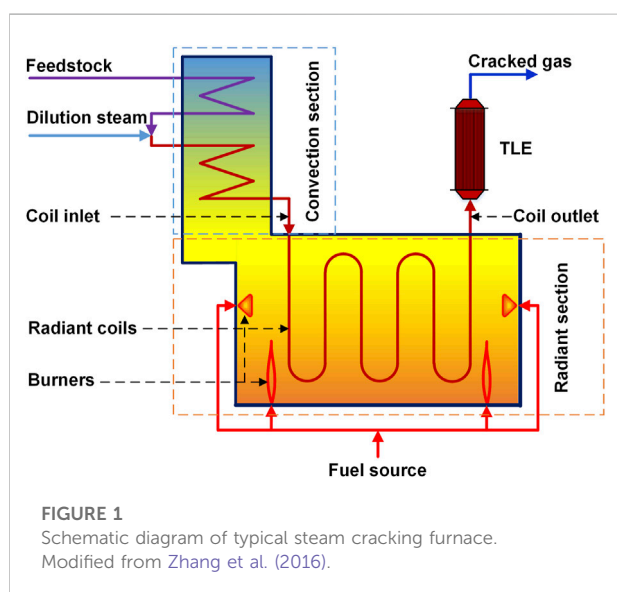
The simulation of steam cracking furnace is significant for production increasing, process control, energy-saving, and emission reduction. The modeling of hydrocarbon pyrolysis could be satisfactorily completed by rigorous calculation of kinetic and thermodynamic parameters, which characterizes many possible reactions. Parmar et al. (2019) established a complete model including mass balance, energy balance, and momentum balance to simulate a naphtha cracker. The kinetic model reaction network used in this work was an extension of the classic Kumar model (Kumar and Kunzru, 1985), which involved one primary reaction and 21 secondary reactions. Fang et al. (2016) established an industrial ethylene cracking furnace model with a free radical mechanism kinetic model, containing 4694 reactions, 93 molecules, and 49 radicals. In order to decrease the

computational scale, a network flow analysis algorithm was proposed to reduce the original network to 2293 reactions. Zhang et al. (2017) directly used the commercial software COILSIM1D (Hillewaert et al., 1988; Willems and Froment, 1988) to simulate a tubular steam cracking reactor. An extensive reaction network consisting of hundreds of species and thousands of elementary reactions had been integrated into the software. Ren et al. (2020) constructed a detailed mechanistic model of naphtha steam cracking with the help of an open-source automatic reaction network generator RMG (Gao et al., 2016; Liu et al., 2021). The final model contains 1947 species and 82130 reactions, and was verified by a set of naphtha steam cracking experiments. However, as demonstrated in the above literature, the traditional complete reactor simulation required a lot of prior knowledge of the pyrolysis reactions, and plenty of time must be spent to solve a large number of differential equations, which was not conducive to real-time prediction and control (Qian et al., 1994). Using machine learning, such as an artificial neural network (ANN) as a surrogate model, could significantly improve the calculation speed of prediction and provide the possibility for high-frequency real-time optimization (RTO) strategy or process control for the naphtha steam cracking process (Fakhroleslam and Sadrameli, 2020).

Nowadays, ANN has been widely used to predict the products of the naphtha steam cracking furnace. Niaei et al. (2007) used both ANN and mathematical models to predict main product yields in the thermal cracking of naphtha and found that the performance of the former was better than the latter. Sedighi et al. (2011) compared several mechanistic and empirical models, including ANN, to predict the main product yields of heavy liquid hydrocarbon thermal cracking. The results showed that ANN had better results than the kinetic and polynomial models. Jin et al. (2015), Jin et al. (2016) used a feed-forward neural network to replace the complex free radical pyrolysis model of naphtha. Good prediction accuracy of the developed ANN was validated. Li et al. (2021) used two-layer neural networks as the predict model of the ethylene yield in their proposed DPC-ANN model. Hua et al. (2018a), Hua et al. (2018b) found that, similar to images, the architecture of the naphtha pyrolysis network has modular features after a graph theoretical analysis. The authors introduced convolutional neural networks (CNN) to extract the features of its topology to build a novel model of naphtha pyrolysis. The proposed model was considered to be well generalized and predict the yields of key products with high accuracy. Based on Hua's CNN model, Bi et al. (2020) introduced transfer learning to create the source data from numerical simulation in order to deal with conflicts between the high demands of data quantity and low supplies from practical cases. Recently, this research group developed an innovative graph neural network model to extract effective features for product prediction, aiming to further improve the prediction accuracy (Zhang et al., 2021). In their works, neural network inputs are usually feed compositions and operating conditions, and outputs are product yields. The feed compositions are converted from feedstock properties using the molecular reconstruction model in an ethylene cracker simulation and

TABLE 1 Summary of the literature using ANN as a surrogate model.

Type of ANN	Input variables of ANN	Output variables of ANN	Size of training set/test set	References
Back-propagation ANN	COT, reactor pressure, dilution ratio, and residence time	Yields of CH <sub>4</sub> , C <sub>2</sub> H <sub>4</sub> , C <sub>2</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>4</sub> H <sub>6</sub> , C <sub>4</sub> , C <sub>5</sub> +, and aromatics	48/15	Niaei et al. (2007)
Unspecified	Dilution ratio, COT, and COP	Yields of C <sub>2</sub> H <sub>4</sub> and C <sub>3</sub> H <sub>6</sub>	200/100	Li et al. (2007)
Radical basis functions neural network	Feed flow rate, dilution ratio, resident time, COT, and COP	Yields of C <sub>2</sub> H <sub>4</sub> , C <sub>3</sub> H <sub>6</sub> , and CH <sub>4</sub>	160/160	Geng et al. (2010)
Back-propagation ANN	COT, dilution ratio, and feed flow rate	Yields of C <sub>2</sub> H <sub>4</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>2</sub> H <sub>6</sub> , CH <sub>4</sub> , and C <sub>5</sub> +	106/14	Sedighi et al. (2011)
Feed-forward neural network	Feed flow rate and COT	Yields of H <sub>2</sub> , C <sub>2</sub> H <sub>4</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>4</sub> H <sub>6</sub> , benzene, and TMT	1024/462	Jin et al. (2015)
Convolutional neural networks	Feed compositions, CIT, COT, CIP, feed flow rate, and dilution ratio	Yields of H <sub>2</sub> , CH <sub>4</sub> , C <sub>2</sub> H <sub>4</sub> , C <sub>2</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>8</sub> , C <sub>4</sub> H <sub>6</sub> , NC <sub>4</sub> H <sub>8</sub> and IC <sub>4</sub> H <sub>8</sub>	2000/500	Hua et al. (2018a)
Deep learning back-propagation ANN	Naphtha bulk properties including PIONA fractions, density, vapor pressure; COT, COP, E/E, P/E, and M/P	Yields of 28 (pseudo-)components including C <sub>2</sub> H <sub>4</sub> , C <sub>3</sub> H <sub>6</sub> , benzene, C7 iso-paraffins, C10+ aromatics and so on	239/33 for naphtha; 13600/1587 for reactor	Plehiers et al. (2019)
Convolutional neural networks with transfer learning	Feed compositions, COT shape Variables, CIP, CIP, COP, feed flow rate, dilution ratio, and coke thickness	Yields of H <sub>2</sub> , CH <sub>4</sub> , C <sub>2</sub> H <sub>4</sub> , C <sub>4</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>2</sub> H <sub>6</sub> , C <sub>6</sub> H <sub>6</sub> , C <sub>7</sub> H <sub>8</sub> , C <sub>8</sub> H <sub>10</sub> , and TMT	4593/1968	Bi et al. (2020)
Graph neural network	Feed compositions, CIT, COT, CIP, feed flow rate, and dilution ratio	Yields of H <sub>2</sub> , CH <sub>4</sub> , C <sub>2</sub> H <sub>4</sub> , C <sub>2</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>8</sub> and C <sub>4</sub> H <sub>6</sub>	1500/500	Zhang et al. (2021)
Back-propagation ANN	Naphtha bulk properties including molecular weight, density, PIONA fractions, and distillation curve; residence time, dilution ratio, pressure, and COT	Yields of H <sub>2</sub> , CH <sub>4</sub> , C <sub>2</sub> H <sub>6</sub> , C <sub>2</sub> H <sub>4</sub> , C <sub>3</sub> H <sub>8</sub> , C <sub>3</sub> H <sub>6</sub> , C <sub>2</sub> H <sub>2</sub> , 1-C <sub>4</sub> H <sub>6</sub> , iso-C <sub>4</sub> H <sub>6</sub> , 1,3-C <sub>4</sub> H <sub>6</sub> and 2-C <sub>4</sub> H <sub>6</sub>	45/5 for naphtha; 4520/472 for reactor	This work

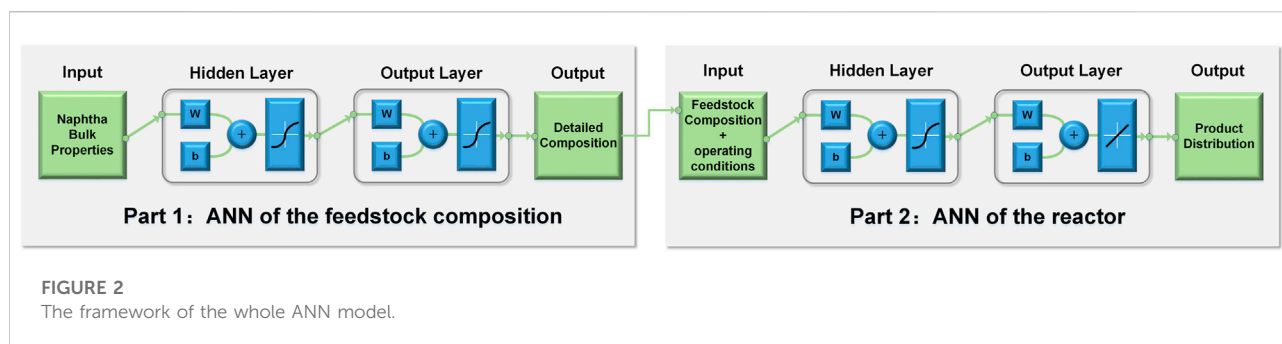


optimization system EcSOS (Fang et al., 2017; Bi and Qiu, 2019). Plehiers et al. (2019) considered both ANN of feedstock composition and reactor modeling. A four interacting deep learning ANNs (DL ANNs) was developed to make the model deal with different types of naphtha feedstock. Among the DL ANNs framework, the second network used the PIONA composition of the naphtha and the boiling points to reconstruct the detailed composition of the feedstock instead of molecular reconstruction algorithms. The initial input

was limited bulk properties of the naphtha feedstock and process-related variables. The final goal was to predict steam cracker product compositions.

In addition to using ANN for prediction, many articles use ANN as a surrogate model for process optimization. Li et al. (2007) optimized the operation of a naphtha industrial cracking furnace by combining the ANN hybrid model with a multi-objective particle swarm optimization procedure. The steam-to-naphtha ratio (also called dilution ratio), the coil outlet temperature (COT), and the outlet pressure were considered to be the decision variables of multi-objective problems. Geng et al. (2010) searched the optimal operating conditions to maximize the total yields of ethylene and propylene and linked them to cracking depth control. The prediction model was radical basis functions neural network (RBFNN). Jin et al. (2015) also applied the ANN model to investigate the multi-objective optimization of the mean yields of key products and the day mean profits. The feed flow rate and COT were chosen as the direct input variables, and the coke thickness inside the coil was taken as the state variable.

Table 1 summarizes the above literature based on ANN as a surrogate model. In the above literature, the naphtha feedstock was fixed or converted from molecular reconstruction models (except the work of Plehiers et al., 2019), whether for prediction or optimization. In modern refineries, timeliness and accuracy are both important to meet the demands of RTO systems. It is necessary to make a rapid



prediction when the compositions of naphtha change. However, the experimental methods for composition analysis are often costly and time-consuming, such as gas chromatography (GC). In the contract, it is relatively easy to obtain the common bulk properties of naphtha. If the naphtha feedstock was fixed, the ANN prediction performance and the process optimization were limited by the type of naphtha. Even though molecular reconstruction models are able to convert properties into feed compositions, it is also limited by the prediction accuracy and relatively long solution times. Therefore, in order to make artificial intelligence better serve the steam cracking industry, it is reasonable to take into account the model of bulk properties to the detailed product compositions, which not only improves the applicability of different naphtha feedstock but also better meets the needs of industrial RTO.

In this work, an artificial neural network model composed of two sub-networks is constructed for the naphtha steam cracking process, where the naphtha feedstock compositions are variables. The first sub-network is the prediction from the bulk properties of the naphtha to the detailed compositions, and the second sub-network is the prediction from feedstock compositions and operating conditions to cracking product yields. The two sub-networks combined ANN realizes the prediction of the steam cracking process directly from limited naphtha bulk properties. Among the first sub-network, direct and indirect feedstock composition ANN are proposed, and the effects of these two strategies on the prediction accuracy of the final product yields are compared.

## 2 Construction of the artificial neural network

The artificial neural network model of the naphtha steam cracking process is generally divided into two parts, as shown in Figure 2. The first part is the ANN of the feedstock composition, which predicts the detailed compositions from bulk properties of naphtha; the second part is the ANN of the tubular reactor, which predicts the product yields from the feedstock

compositions and operating conditions. The type of ANN used in these two parts is both the back-propagation network (BP ANN), with an input layer, a hidden layer, and an output layer.

MAE (mean absolute error), RMSE (root mean square error), and MAPE (mean absolute percentage error) are selected as the evaluation indicators to evaluate the performance of ANN. MAE is the average value of absolute errors, which can better reflect the actual situation of prediction error.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i^{\text{cal}} - x_i^{\text{exp}}| \quad (1)$$

RMSE measures the deviation between the predicted and actual values, which is more sensitive to outlier data and can highlight the error values with greater influence.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^{\text{cal}} - x_i^{\text{exp}})^2} \quad (2)$$

MAPE considers the error between the predicted value and the actual value and the ratio between the error and the actual value. A MAPE of 0% indicates a perfect model, and a MAPE greater than 100% indicates an inferior model.

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{x_i^{\text{cal}} - x_i^{\text{exp}}}{x_i^{\text{exp}}} \right| \quad (3)$$

### 2.1 Acquisition of data sets

Fifty naphtha samples in the literature (Mei et al., 2017; Ren et al., 2019) were used in this work. The input variables of the feedstock composition ANN are the bulk properties of naphtha, including molecular weight, density, PIONA fractions, and distillation curve. The scopes of the bulk properties are shown in Supplementary Table S1. The output variables are 35 (pseudo-) components to characterize the detailed compositions of naphtha, including C<sub>4</sub>–C<sub>12</sub> n-paraffins, C<sub>4</sub>–C<sub>12</sub> iso-paraffins, C<sub>4</sub>–C<sub>7</sub> olefins, C<sub>5</sub>–C<sub>11</sub> naphthenes, and C<sub>6</sub>–C<sub>11</sub> aromatics (Supplementary Table S2).

TABLE 2 Ranges for process operation-related input variables of reactor ANN.

Variable	Minimum value	Maximum value
Coil output temperature (°C)	753	900
Reactor pressure (bar)	1.7	2.3
Dilution ratio (kg <sub>Steam</sub> /kg <sub>Hydrocarbon</sub> )	0.40	0.65
Residence time (s)	0.1	0.3

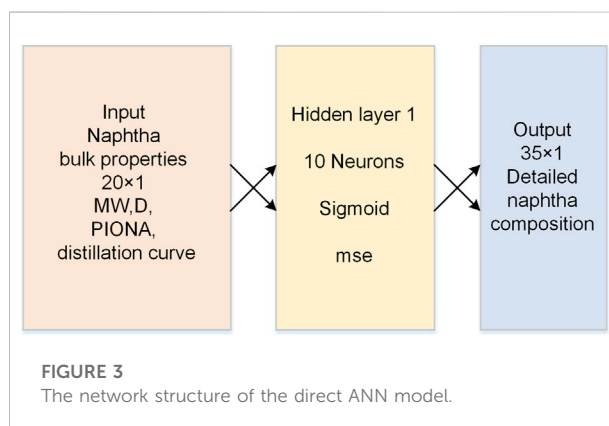
The reactor ANN network contains 39 input variables, which are 35 (pseudo-)components and 4 operating variables including residence time, dilution ratio, reactor pressure, and coil output temperature. The output variables are the mass fractions of 11 representative products: hydrogen, methane, ethane, ethylene, propane, propylene, acetylene, 1-butene, isobutene, 1,3-butadiene, 2-butene. The data sets of reactor ANN were obtained through the mechanism kinetic model in the existing literature (Ren et al., 2020) combined with the reactor simulation performed on CHEMKIN-PRO (2013) software.

Considering the limited amount of available naphtha samples, random variations of 0%~±20% were introduced into the 35 (pseudo-) component concentrations of each naphtha feedstock. The value of ±20% was determined by the relative error of the feedstock composition ANN, as shown in Figure 6 in Section 3.1. The purpose was to make the naphtha composition of each training set different, which could enrich the coverage of the feedstock composition of the network.

The determination of the ranges of operating variables is detailed in Supplementary Section S2. The final ranges of the operating variables are shown in Table 2. The input variables of reactor ANN were sampled using Latin hypercube sampling (uniform distribution) (Mckay and Beckman, 1979) within their respective sampling ranges. Finally, there was a total of 4520 simulation data sets in the reactor ANN data sets. Note that only the naphtha samples used to train the network needed to introduce random variations into the concentrations (corresponding to the 45 training sets below), and the naphtha samples used as the test sets remained the same as the real concentrations (corresponding to the 5 test naphtha below).

## 2.2 Feedstock composition artificial neural network

The first sub-network, feedstock composition ANN, was to predict the contents of 35 components from 20 bulk properties. The 50 sets of naphtha bulk properties with detailed composition data were divided into two parts: 45 sets of oil samples were randomly selected to generate ANN, and the remaining 5 sets were used for testing. It should be emphasized that the 5 selected



test sets do not involve any process in ANN generation, which are only used to test the prediction accuracy of ANN. The hidden layer and the output layer of the feedstock composition ANN were both sigmoid functions. The number of hidden layer neurons was determined by selecting the minimum network error after multiple adjustments. The process was to perform 50 repetitions of training for each network with various hidden layer neurons (in the range of 5–20) through an automatic program. The network with the minimum error was screened out, and the number of neurons in its hidden layer was used. The ANN mentioned below also implemented this unified adjustment process.

Predicting the detailed composition directly from bulk properties of naphtha is called direct ANN in this article, i.e., realizing prediction only through an ANN with 20 inputs and 35 outputs. The network structure of direct ANN is shown in Figure 3. The number of neurons of the hidden layer was determined to be 10 after adjustments. In addition to the direct ANN model, we developed another indirect ANN model inspired by the literature. Plehiers et al. (2019) mentioned that splitting the output layer into five separate PIONA categories could reduce the impact of wide ranges in the absolute concentrations of the components in the different categories. However, the authors just considered that directly predicting all fractions at once would result in a network that was difficult to train. After our analysis, it was found that the greater significance of dividing the output layer into five groups of PIONA lay in making the PIONA contents in the output



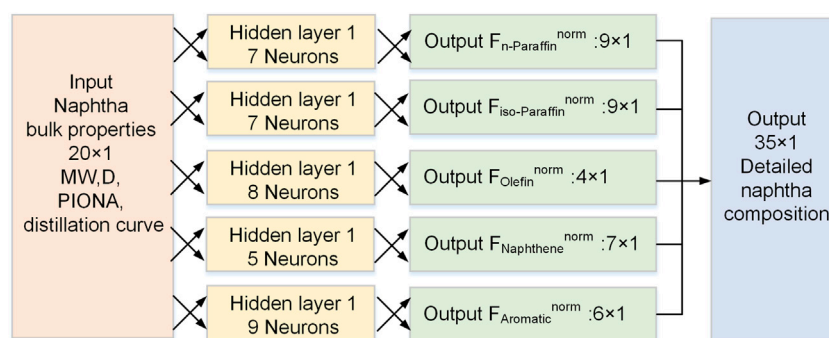


FIGURE 4  
The network structure of the indirect ANN model.

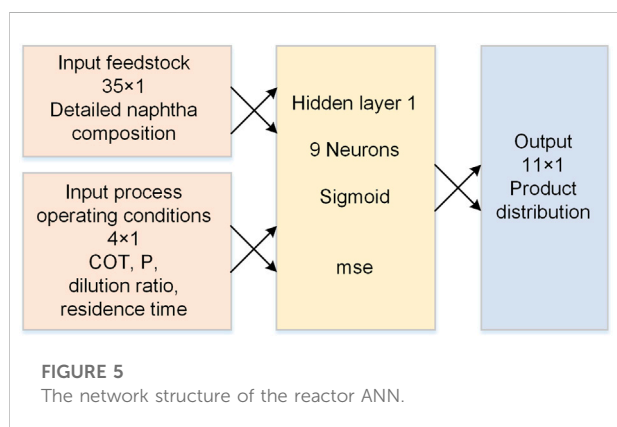


FIGURE 5  
The network structure of the reactor ANN.

exactly equal to the actual values of each homologues series. The differences in PIONA contents had greater impacts on the cracking products compared to the content error of each component, as shown in Section 3.3. Therefore, we referred to the proposal of literature, while in the of modeling process, there were differences:

- Preprocessing: The detailed composition data of the naphtha samples was normalized in each homologous series according to PIONA. The normalized data were used for network training.

$$y_{i,j}^{actual} = \frac{F_{i,j}^{actual}}{\sum_{i \in I} F_{i,j}^{actual}} \quad \forall j \in J \quad (4)$$

where  $F_{i,j}^{actual}$  was the actual mass fraction of component  $i$  in the whole oil,  $y_{i,j}^{actual}$  was the fraction of component  $i$  in homologous series  $j$  through normalizing  $F_{i,j}^{actual}$ ,  $I$  was the set of components in homologous series  $j$ .  $J$  was the set of homologous series.

- Network division: The indirect ANN was divided into 5 sub-networks. The inputs of each network were the bulk

properties, and the outputs were the fraction of each component in the corresponding homologous series.

$$Y = f_2(W_2 \cdot f_1(W_1 \cdot X + b_1) + b_2) \quad (5)$$

where  $Y$  was the output vector,  $X$  was the input vector,  $f_1$  and  $f_2$  were the activation functions for the hidden layer and output layer, respectively;  $W_1$  and  $W_2$  was the weight vectors for hidden layer and output layer, respectively;  $b_1$  and  $b_2$  were the bias vectors for the hidden layer and output layer, respectively.

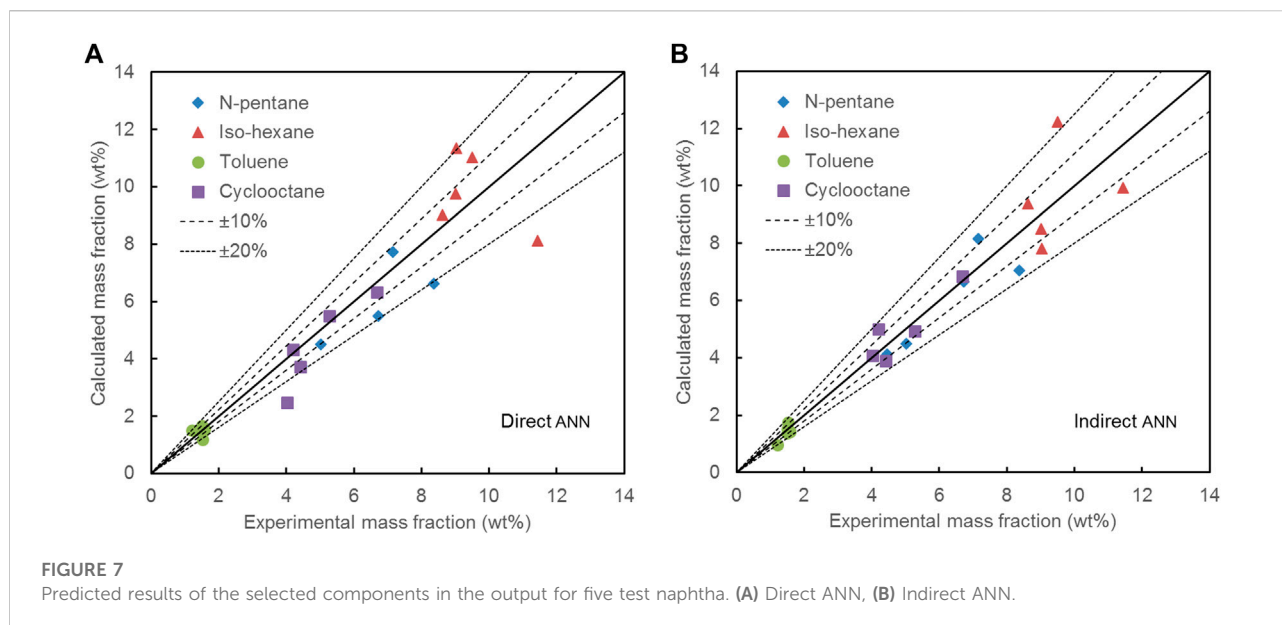
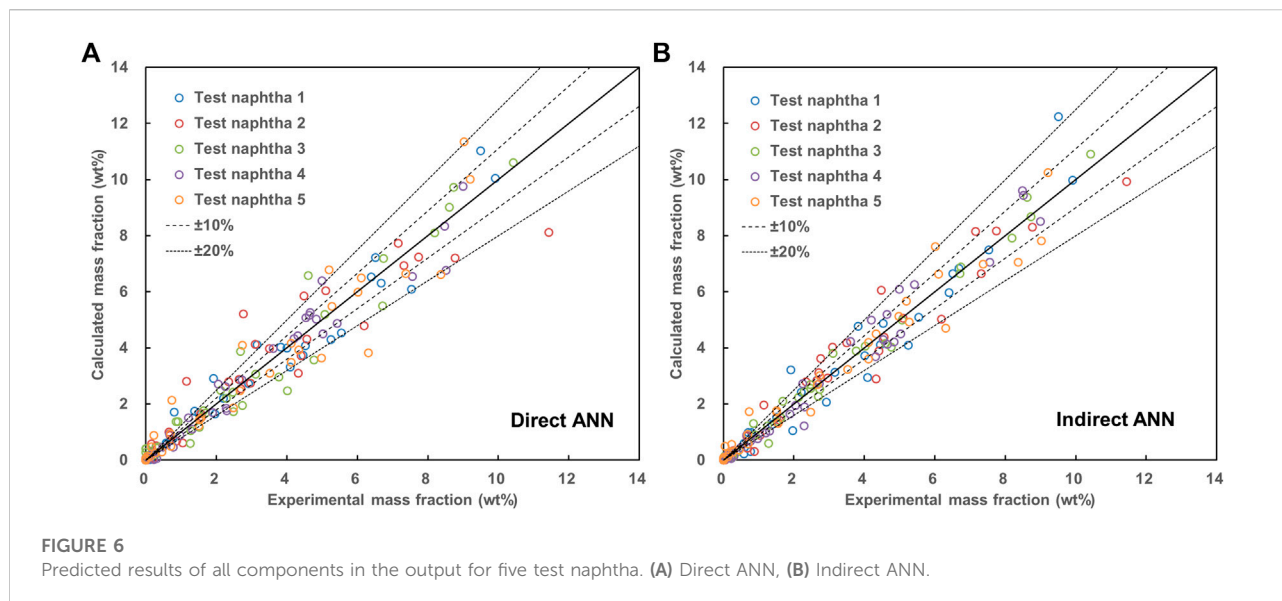
- Post-processing: After obtaining the network outputs, considering the difference between the predicted results and the actual fractions, although the difference was small, a step of normalization was performed to make the total fractions to be 100%. Subsequently, the component fractions in each sub-network were multiplied by the total content of the corresponding homologous series to ensure that the PIONA content was exactly the same as the actual value.

$$y_{i,j}^{predict} = \frac{Y_{i,j}^{predict}}{\sum_{i \in I} Y_{i,j}^{predict}} \quad \forall j \in J \quad (6)$$

$$F_{i,j}^{predict} = y_{i,j}^{predict} \cdot F_j^{actual} \quad (7)$$

where  $Y_{i,j}^{predict}$  was the predicted fraction of component  $i$  in homologous series  $j$ ,  $y_{i,j}^{predict}$  was the predicted fraction of component  $i$  in homologous series  $j$  through normalizing  $Y_{i,j}^{predict}$ ,  $F_{i,j}^{predict}$  was the predicted mass fraction of component  $i$  in the whole oil,  $F_j^{actual}$  was the actual content of homologous series  $j$  available in the bulk properties.

The number of neurons in the five sub-networks of the indirect method ANN was determined to be 7, 7, 8, 5, and 9 after adjustments. Its network structure is shown in Figure 4.



## 2.3 Reactor artificial neural network

The second sub-network, reactor ANN, was to predict the yields of 11 representative cracking products by 35 feedstock composition variables and 4 reactor operation variables. Among the 4520 sets of reactor simulation data, 472 sets of data related to the 5 test naphtha were used to verify the reactor ANN further, and the other 4048 sets of data were used to train the reactor

ANN. The test sets were used to verify the reactor ANN, and their compositions were the same as the real values. Similarly, the 472 test sets still did not participate in any network training process. The reactor ANN used a network structure similar to the feed composition ANN. The hidden layer activation function was the sigmoid function, and the output layer was the purelin function. After adjustments, the number of neurons in the hidden layer was determined to be 9. Its network structure is shown in [Figure 5](#).

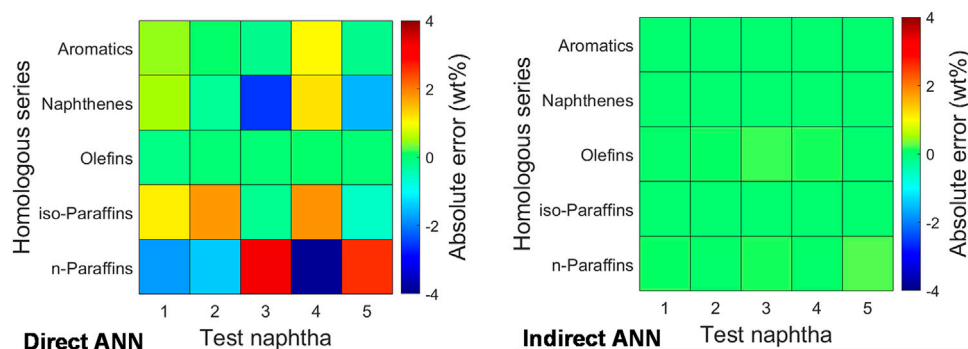


FIGURE 8

Contents deviations of each homologues series for five test naphtha of the direct and indirect ANN.

### 3 Results and discussion

#### 3.1 Predictive performance of feedstock composition artificial neural network

Five test naphtha were used to test the direct and indirect ANN, respectively. Figure 6 shows the parity plots for all components in the output for five test naphtha. Figure 7 shows the selected components with high contents in each homologous series (the olefins are too low to be ignored). It can be seen from Figure 6 that although the errors of the predicted mass fractions of most components fell within  $\pm 20\%$  for both two methods, the indirect ANN had higher accuracy than the direct ANN. For the indirect ANN, the number of components whose errors were greater than  $\pm 20\%$  was significantly reduced, and the errors were more concentrated within  $\pm 10\%$ . Figure 7 more clearly shows the prediction accuracy of the two methods for representative components. Obviously, the errors of the four representative components of the indirect method were smaller, which proved that the indirect method effectively improved the prediction accuracy.

Figure 8 shows the contents deviations of each homologues series (PIONA) for five test naphtha. Obviously, the direct ANN had significant errors in predicting PIONA contents. Among them, for n-paraffin, the average absolute error of the five test naphtha reached 2.60wt%. The predicted PIONA contents of indirect ANN were almost equal to the actual values. The slight errors of the indirect method were due to the difference between the PIONA contents obtained from the naphtha bulk properties and that obtained by adding the detailed compositions. Table 3 shows the MAE and RMSE of the 35 component contents of the direct and indirect methods. The MAE and RMSE of the direct ANN were 0.47wt% and 0.55wt%, respectively, while these of the indirect ANN were reduced to 0.36wt% and 0.42wt%, respectively. The above results showed that the strategy of indirect ANN, i.e., the compositions were divided into five sub-networks of PIONA according to the actual physical

meaning, significantly improved the prediction accuracy of the feedstock compositions. Moreover, the accuracy of indirect ANN was also higher than the results obtained by our previous molecular reconstruction model (Ren et al., 2019). The MAE and RMSE of the five test-naphtha obtained by molecular reconstruction were 1.00wt% and 1.20wt%, respectively. It was worth noting that the average calculation time was shortened from 215s by molecular reconstruction to 0.104s by indirect ANN per naphtha. The computations were carried out on the same PC constituted by Intel Xeon E31230 processor 3.20 GHz.

#### 3.2 Predictive performance of reactor artificial neural network

Figure 9 is the parity plots of four important products selected from the reactor ANN output. Ethylene and propylene are the most concerning products of naphtha steam cracking. Methane is a by-product with high yield, and 1,3-butadiene is a by-product with high added value. Note that the yields of all reactor outlet products in this article are dry yields after removing water. It can be seen from Figure 9 that the errors of 472 sets of test data were almost all within  $\pm 20\%$ , except for very few points (4%) exceeding the margin of error. There were 71%, 63%, 58%, 61% of the data within  $\pm 10\%$  for ethylene, propylene, methane, 1,3-butadiene, respectively. The results showed that reactor ANN had good prediction accuracy for these four important products. Table 4 shows the MAE, RMSE, and MAPE of 11 predicted products. The average MAE of all predicted products was 0.24wt%, and the RMSE was 0.28wt%. The MAPE of all products except ethane was less than 10%, indicating that the calculated product yields of the reactor ANN had an acceptable deviation from the actual values. It should be emphasized that the single calculation of reactor ANN on the same PC took only 0.247s. In contrast, the calculation results obtained by the mechanism model through CHEMKIN-PRO reactor simulation usually took 180s.



TABLE 3 Statistics on the performance of the 35 component contents of the direct and indirect ANN.

Component		Direct ANN		Indirect ANN		
		MAE	RMSE	MAE	RMSE	
N-paraffin	C4	0.75	0.92	0.49	0.56	
	C5	0.96	1.06	0.65	0.79	
	C6	0.78	0.83	0.39	0.43	
	C7	1.31	1.42	0.48	0.55	
	C8	0.86	1.13	0.86	0.96	
	C9	0.79	0.94	0.67	0.78	
	C10	0.83	0.96	0.89	1.02	
	C11	0.14	0.15	0.36	0.39	
	C12	0.12	0.15	0.11	0.13	
	Iso-paraffin	C4	0.15	0.21	0.07	0.08
		C5	1.37	1.67	0.82	0.93
		C6	1.66	1.97	1.34	1.55
C7		0.31	0.41	0.64	0.76	
C8		0.74	0.93	0.78	0.95	
C9		0.93	0.97	0.54	0.68	
C10		0.23	0.30	0.51	0.54	
C11		0.27	0.35	0.22	0.23	
C12		0.14	0.20	0.08	0.09	
Olefin		C4	0.00	0.00	0.00	0.00
		C5	0.02	0.02	0.05	0.06
		C6	0.03	0.04	0.05	0.06
	C7	0.00	0.00	0.00	0.00	
Naphthenic	C5	0.32	0.39	0.19	0.21	
	C6	0.42	0.51	0.10	0.12	
	C7	0.61	0.64	0.27	0.32	
	C8	0.58	0.78	0.37	0.46	
	C9	0.43	0.48	0.36	0.39	
	C10	0.32	0.33	0.61	0.71	
Aromatics	C11	0.08	0.13	0.02	0.03	
	C6	0.12	0.15	0.07	0.08	
	C7	0.19	0.22	0.16	0.18	
	C8	0.41	0.48	0.09	0.11	
	C9	0.13	0.19	0.31	0.33	
Average	C10	0.26	0.30	0.18	0.24	
	C11	0.12	0.19	0.02	0.03	
Average		0.47	0.55	0.36	0.42	

### 3.3 Predictive performance of two sub-networks combined artificial neural network

The feedstock composition ANN and reactor ANN were combined to form a complete prediction framework from naphtha bulk properties to pyrolysis product yields. The values of the 35 composition variables in the 472 sets of reactor test data were replaced with the composition results of

the direct and indirect feedstock composition ANN respectively. The reactor operating conditions remained unchanged to investigate the influence of composition prediction accuracy on the prediction of the final product yields.

Table 5 shows the statistics for 11 predicted products with the reactor model combined with direct and indirect ANN. Compared with the single reactor ANN in Section 3.2, the errors of the two sub-networks combined ANN were obviously increased. For the combination with direct ANN,

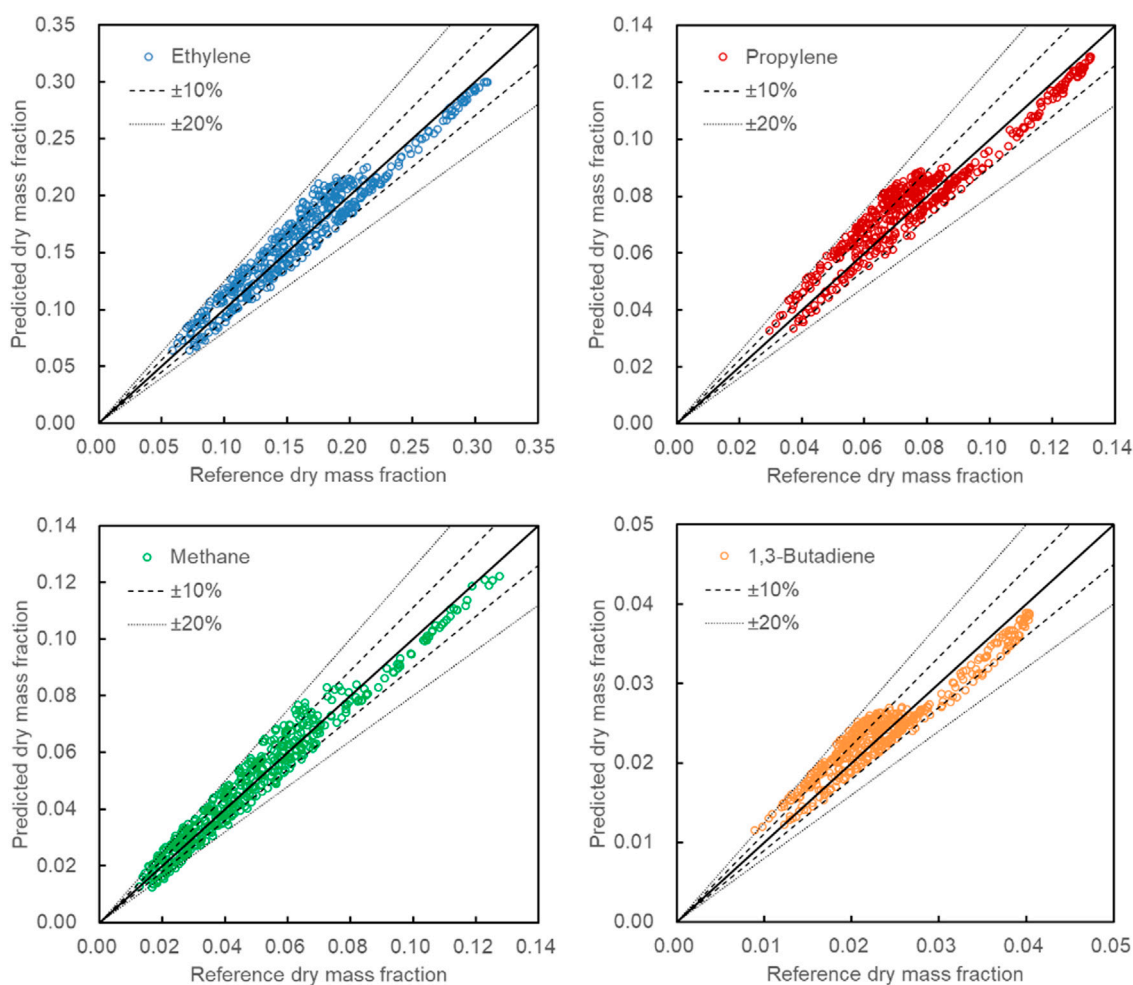


FIGURE 9

Predicted results of four selected products for 472 test sets: ethylene, propylene, methane, 1,3-butadiene.

TABLE 4 Statistics on the performance of the 11 cracking products of reactor ANN.

Products	MAE (wt%)	RMSE (wt%)	MAPE (%)
ethylene	1.12	1.28	7.66
propylene	0.60	0.68	8.75
methane	0.39	0.47	9.52
1,3-butadiene	0.19	0.22	9.00
hydrogen	0.02	0.02	7.10
acetylene	0.03	0.04	8.19
ethane	0.06	0.07	10.49
propane	0.01	0.01	7.13
1-butene	0.08	0.10	8.07
2-butene	0.03	0.04	9.40
isobutene	0.09	0.11	9.88
Average of all above products	0.24	0.28	8.65

TABLE 5 Statistics for 11 predicted products with the reactor model combined with direct and indirect ANN.

Products	Combined with direct ANN			Combined with indirect ANN		
	MAE (wt%)	RMSE (wt%)	MAPE (%)	MAE (wt%)	RMSE (wt%)	MAPE (%)
ethylene	4.58	5.46	30.56	2.40	2.90	14.96
propylene	2.60	3.07	37.35	1.32	1.56	17.61
methane	1.82	2.24	42.19	0.92	1.13	19.05
1,3-butadiene	0.83	0.97	38.90	0.43	0.50	18.37
hydrogen	0.06	0.07	25.29	0.03	0.04	13.16
acetylene	0.05	0.08	11.37	0.05	0.07	12.28
ethane	0.38	0.45	65.20	0.19	0.23	28.55
propane	0.05	0.05	34.04	0.02	0.03	17.11
1-butene	0.31	0.40	24.56	0.19	0.24	16.22
2-butene	0.13	0.15	37.41	0.07	0.08	18.88
isobutene	0.40	0.48	42.36	0.20	0.23	19.19
Average of all above products	1.02	1.22	35.38	0.53	0.64	17.76

TABLE 6 Statistics for 11 predicted products with the compared ANN.

Products	MAE (wt%)	RMSE (wt%)	MAPE (%)
ethylene	3.56	4.33	21.87
propylene	1.95	2.35	25.82
methane	1.42	1.76	29.80
1,3-butadiene	0.63	0.75	26.63
hydrogen	0.05	0.06	19.34
acetylene	0.06	0.08	14.61
ethane	0.29	0.35	45.87
propane	0.04	0.04	24.43
1-butene	0.26	0.34	21.38
2-butene	0.09	0.12	26.26
isobutene	0.30	0.36	29.02
Average of all above products	0.79	0.96	25.91

the average MAE, RMSE, and MAPE of 11 products were more than 4 times higher than those of single reactor ANN. For the combination with indirect ANN, they were more than twice that of single reactor ANN. It was indicated that the prediction accuracy of the feedstock composition had a significant impact on the final product yields. However, improving the predictive ability of the feedstock composition ANN was limited by the size of the naphtha data set, which was more difficult and expensive to obtain than reactor data. Therefore, based on the limited naphtha data set, improving the prediction accuracy of feedstock composition ANN was the key point.

It was worth noting that the MAE, RMSE, and MAPE of all products obtained by the indirect ANN were smaller than those of the direct ANN. The average MAE of 11 products of the two

methods were 0.53wt% and 1.02wt%, respectively. The difference between them was almost double, while the errors of single direct feedstock composition ANN were only 1.3 times higher than those of single indirect feedstock composition ANN, as shown in Section 3.1. Note that the used reactor ANN in the combination process was the same. The difference between these two combined ANN implied that the composition errors of the direct ANN were further amplified when they were transferred to the product errors, while the indirect ANN could control the product errors to an acceptable level.

An indirect feedstock composition ANN was designed for further comparison, called: compared ANN. The model framework of this compared ANN was consistent with the

aforementioned indirect ANN, and the error of the model was the same as that of the direct ANN by adjusting the parameters of the neural network. As shown in [Supplementary Table S3](#), the MAE and RMSE of 35 components of compared ANN were equal to those of the direct ANN, which were 0.47wt% and 0.55wt%, respectively. The composition results of the compared ANN were brought into the reactor ANN to predict the product yields. As shown in [Table 6](#), the following results were obtained: the average MAE, RMSE, and MAPE of 11 products with the compared ANN were 0.79wt%, 0.96wt%, and 25.91%, respectively. The errors were larger than the aforementioned indirect ANN (0.53wt%, 0.64wt%, 17.76%) but smaller than the direct ANN (1.02wt%, 1.22wt%, 35.38%).

The above results showed that even if the feedstock composition errors were the same, the indirect ANN reduced the errors of product prediction and the error transmission due to the accurate PIONA contents. The indirect ANN transmitting fewer errors benefited not only from reducing errors of each (pseudo-)component, but also from the ability to ensure PIONA contents equal to the actual values of each homologues series, which may be the most important advantage of the indirect ANN. In connection with the actual physical background, the contents of naphthenes and aromatics greatly impact the product yields. If the contents of these two parts in naphtha feedstock are predicted to be higher or lower, it will seriously cause the olefin yield to decrease or increase.

## 4 Conclusion

An artificial neural network model of the naphtha steam cracking process was constructed. It was composed of two sub-networks: the feedstock composition ANN and the reactor ANN. The whole model could directly predict the yields of cracking products from the bulk properties of the naphtha feedstock. In the first part of the feedstock composition ANN, two different ANN construction methods were compared: directly predicting the detailed composition from bulk properties of naphtha; indirectly predicting the respective contents of PIONA first, and then obtaining the total detailed composition. The results showed that the indirect method had higher prediction accuracy, and the MAE of 35 components was 0.36wt%. In the second part of the reactor ANN, the MAE of 11 cracking products was 0.24wt%. Both the feedstock composition ANN and the reactor ANN are computationally much faster than the molecular reconstruction algorithms and mechanistic models compared to them, respectively. Combining the two parts of the network, i.e., directly predicted the product yields from the bulk properties, for the direct method and the indirect method, the MAE of the product yields were 0.53wt% and 1.02wt%, respectively. Through the design of the compared case, there is evidence to believe that the indirect method had lower product prediction errors, not just

the reduction of the total error of the feedstock composition. The critical factor was ensuring that PIONA contents were equal to the actual values. Overall, it should be emphasized that the prediction accuracy of feedstock composition ANN is the key point. In addition to the indirect ANN strategy supported in this paper, for the conflicts between “data-hungry” and “data-lacking,” more advanced machine learning technologies such as transfer learning, are needed to deal with naphtha feedstock’s few-shot learning.

## Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## Author contributions

YR: conceptualization, methodology, writing—original draft. ZL: conceptualization, methodology, writing—review and editing. YY: acquisition, supervision. JS: acquisition, supervision. BJ: acquisition, supervision. JW: conceptualization, funding acquisition. YY: conceptualization, funding acquisition.

## Funding

The financial supports provided by the Project of the National Natural Science Foundation of China (21822809 and 21978256) are gratefully acknowledged.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fceng.2022.983035/full#supplementary-material>

## References

- Armenise, S., Wong, S., Ramírez-Velásquez, S., Launay, F., Wuebben, D., Nyakuma, B. B., et al. (2021). Application of computational approach in plastic pyrolysis kinetic modelling: A review. *Reac. Kinet. Mech. Cat.* 134 (2), 591–614. doi:10.1007/s11144-021-02093-7
- Bi, K., Beykal, B., Avraamidou, S., Pappas, I., Pistikopoulos, E. N., and Qiu, T. (2020). Integrated modeling of transfer learning and intelligent heuristic optimization for a steam cracking process. *Ind. Eng. Chem. Res.* 59 (37), 16357–16367. doi:10.1021/acs.iecr.0c02657
- Bi, K., and Qiu, T. (2019). Novel naphtha molecular reconstruction process using a self-adaptive cloud model and hybrid genetic algorithm–particle swarm optimization algorithm. *Ind. Eng. Chem. Res.* 58 (36), 16753–16760. doi:10.1021/acs.iecr.9b02605
- Bi, K., Zhang, S., Zhang, C., Li, H., Huang, X., Liu, H., et al. (2021). Knowledge expression, numerical modeling and optimization application of ethylene thermal cracking: From the perspective of intelligent manufacturing. *Chin. J. Chem. Eng.* 38, 1–7. doi:10.1016/j.cjche.2021.03.033
- BMBF (2011). Industrie 4.0 - BMBF. Available at: <https://www.bmbf.de/de/zukunftspjunkt-industrie-4-0-848.html> (Accessed July 29, 2022).
- Chakkingal, A., Janssens, P., Poissonnier, J., Barrios, A. J., Virginie, M., Khodakov, A. Y., et al. (2022). Machine learning based interpretation of microkinetic data: A Fischer-Tropsch synthesis case study. *Reac. Chem. Eng.* 7 (1), 101–110. doi:10.1039/d1re00351h
- CHEMKIN-PRO (2013). *CHEMKIN-PRO 15131: Reaction design*. San Diego.
- Dobbelaere, M. R., Plehiers, P. P., Van de Vijver, R., Stevens, C. V., and Van Geem, K. M. (2021). Machine learning in chemical engineering: strengths, weaknesses, opportunities, and threats. *Eng.* 7 (9), 1201–1211. doi:10.1016/j.eng.2021.03.019
- Fakhroleslam, M., and Sadrameli, S. M. (2019). Thermal/catalytic cracking of hydrocarbons for the production of olefins; a state-of-the-art review III: process modeling and simulation. *Fuel* 252, 553–566. doi:10.1016/j.fuel.2019.04.127
- Fakhroleslam, M., and Sadrameli, S. M. (2020). Thermal cracking of hydrocarbons for the production of light olefins; a review on optimal process design, operation, and control. *Ind. Eng. Chem. Res.* 59 (27), 12288–12303. doi:10.1021/acs.iecr.0c00923
- Fang, Z., Qiu, T., and Chen, B. (2016). Improvement of ethylene cracking reaction network with network flow analysis algorithm. *Comput. Chem. Eng.* 91, 182–194. doi:10.1016/j.compchemeng.2016.04.020
- Fang, Z., Qiu, T., and Zhou, W. (2017). Coupled simulation of recirculation zonal firebox model and detailed kinetic reactor model in an industrial ethylene cracking furnace. *Chin. J. Chem. Eng.* 25 (8), 1091–1100. doi:10.1016/j.cjche.2017.03.020
- Gao, C. W., Allen, J. W., Green, W. H., and West, R. H. (2016). Reaction mechanism generator: Automatic construction of chemical kinetic mechanisms. *Comp. Phys. Comm.* 203, 212–225. doi:10.1016/j.cpc.2016.02.013
- Geng, Z., Shang, T., and Li, F. (2010). "Integrated MMPSO and RBFNN for optimal control of cracking depth in ethylene cracking furnace," in 2010 Sixth International Conference on Natural Computation, Yantai, China, 10–12 Aug. 2010. doi:10.1109/ICNC.2010.5584856
- Goldsmith, B. R., Esterhuizen, J., Liu, J. -X., Bartel, C. J., and Sutton, C. (2018). Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* 64 (7), 2311–2323. doi:10.1002/aic.16198
- Hillewaert, L. P., Dierickx, J. L., and Froment, G. F. (1988). Computer generation of reaction schemes and rate equations for thermal cracking. *AIChE J.* 34 (1), 17–24. doi:10.1002/aic.690340104
- Hua, F., Fang, Z., and Qiu, T. (2018a). Application of convolutional neural networks to large-scale naphtha pyrolysis kinetic modeling. *Chin. J. Chem. Eng.* 26 (12), 2562–2572. doi:10.1016/j.cjche.2018.09.021
- Hua, F., Fang, Z., and Qiu, T. (2018b). Modeling ethylene cracking process by learning convolutional neural networks. *Comput. Aided Chem. Eng.* 44, 841–846.
- Jin, Y., Li, J., Du, W., and Qian, F. (2015). Multi-objective optimization of pseudo-dynamic operation of naphtha pyrolysis by a surrogate model. *Chem. Eng. Technol.* 38 (5), 900–906. doi:10.1002/ceat.201400162
- Jin, Y., Li, J., Du, W., and Qian, F. (2016). Adaptive sampling for surrogate modelling with artificial neural network and its application in an industrial cracking furnace. *Can. J. Chem. Eng.* 94 (2), 262–272. doi:10.1002/cjce.22384
- Kumar, P., and Kunzru, D. (1985). Modeling of naphtha pyrolysis. *Ind. Eng. Chem. Proc. Des. Dev.* 24 (3), 774–782. doi:10.1021/i200030a043
- Li, C., Zhu, Q., and Geng, Z. (2007). Multi-objective particle swarm optimization hybrid algorithm: an application on industrial cracking furnace. *Ind. Eng. Chem. Res.* 46 (11), 3602–3609. doi:10.1021/ie051084t
- Li, Q., Zhang, M., Shi, X., Lan, X., Guo, X., and Guan, Y. (2021). An intelligent hybrid feature subset selection and production pattern recognition method for modeling ethylene plant. *J. Anal. Appl. Pyrolysis* 160, 105352. doi:10.1016/j.jaap.2021.105352
- Liu, M. J., Dana, A. G., Johnson, M. S., Goldman, M. J., Jocher, A., Payne, A. M., et al. (2021). Reaction Mechanism generator v3.0: Advances in automatic mechanism generation. *J. Chem. Inf. Model.* 61 (6), 2686–2696. doi:10.1021/acs.jcim.0c01480
- Mckay, M. D., and Beckman, R. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer Code. *Technometrics* 21, 239. doi:10.1080/00401706.1979.10489755
- Mei, H., Wang, Z., and Huang, B. (2017). Molecular-based Bayesian regression model of petroleum fractions. *Ind. Eng. Chem. Res.* 56 (50), 14865–14872. doi:10.1021/acs.iecr.7b02905
- Niaei, A., Towfighi, J., Khataee, A. R., and Rostamizadeh, K. (2007). The use of ANN and the mathematical model for prediction of the main product yields in the thermal cracking of naphtha. *Petroleum Sci. Technol.* 25 (8), 967–982. doi:10.1080/10916460500423304
- Parmar, K. K., Padmavathi, G., and Dash, S. K. (2019). Modelling and simulation of naphtha cracker. *Indian Chem. Eng.* 61 (2), 182–194. doi:10.1080/00194506.2018.1529633
- Plehiers, P. P., Symoens, S. H., Amghizar, I., Marin, G. B., Stevens, C. V., and Van Geem, K. M. (2019). Artificial intelligence in steam cracking modeling: a deep learning algorithm for detailed effluent prediction. *Engineering* 5 (6), 1027–1040. doi:10.1016/j.eng.2019.02.013
- Qian, F., Yu, J., and Jiang, W. (1994). "Modelling of industrial thermal cracking furnaces via functional-link artificial neural networks," in Proceedings of 1994 IEEE International Conference on Industrial Technology - ICIT '94, Guangzhou, China, 5–9 Dec. 1994. doi:10.1109/ICIT.1994.467033
- Ren, Y., Liao, Z., Sun, J., Jiang, B., Wang, J., Yang, Y., et al. (2019). Molecular reconstruction of naphtha via limited bulk properties: methods and comparisons. *Ind. Eng. Chem. Res.* 58 (40), 18742–18755. doi:10.1021/acs.iecr.9b03290
- Ren, Y., Guo, G., Liao, Z., Yang, Y., Sun, J., Jiang, B., et al. (2020). Kinetic modeling with automatic reaction network generator, an application to naphtha steam cracking. *Energy* 207, 118204. doi:10.1016/j.energy.2020.118204
- Sadrameli, S. M. (2015). Thermal/catalytic cracking of hydrocarbons for the production of olefins: a state-of-the-art review I: thermal cracking review. *Fuel* 140, 102–115. doi:10.1016/j.fuel.2014.09.034
- Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R., and Tkatchenko, A. (2017). Quantum-chemical insights from deep tensor neural networks. *Nat. Comm.* 8. doi:10.1038/ncomms13890
- Sedighi, M., Keyvanloo, K., and Towfighi, J. (2011). Modeling of thermal cracking of heavy liquid hydrocarbon: application of kinetic modeling, artificial neural network, and Neuro-Fuzzy models. *Ind. Eng. Chem. Res.* 50 (3), 1536–1547. doi:10.1021/ie1015552
- St John, P. C., Guan, Y., Kim, Y., Kim, S., and Paton, R. S. (2020). Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost. *Nat. Comm.* 11 (1). doi:10.1038/s41467-020-16706-7
- Tran, K., and Ulissi, Z. W. (2018). Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution. *Nat. Catal.* 1 (9), 696–703. doi:10.1038/s41929-018-0142-1
- Venkatasubramanian, V. (2019). The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* 65 (2), 466–478. doi:10.1002/aic.16489
- Willems, P. A., and Froment, G. F. (1988). Kinetic modeling of the thermal cracking of hydrocarbons. 1. Calculation of frequency factors. *Ind. Eng. Chem. Res.* 27 (11), 1959–1966. doi:10.1021/ie00083a001
- Yalamanchi, K. K., Monge-Palacios, M., van Oudenhoven, V. C. O., Gao, X., and Sarathy, S. M. (2020). Data science approach to estimate enthalpy of formation of cyclic hydrocarbons. *J. Phys. Chem. A* 124 (31), 6270–6276. doi:10.1021/acs.jpca.0c02785
- Zhang, S., Li, H., and Qiu, T. (2021). An innovative graph neural network model for detailed effluent prediction in steam cracking. *Ind. Eng. Chem. Res.* 60 (50), 18432–18442. doi:10.1021/acs.iecr.1c03728
- Zhang, Y., Hu, G., Du, W., and Qian, F. (2016). "Chapter three - steam cracking and EDC furnace simulation," in *Advances in chemical engineering*. Editor K. M. Van Geem (Academic Press), 199–272.
- Zhang, Y., Reyniers, P. A., Schietekat, C. M., Van Geem, K. M., Marin, G. B., Du, W. L., et al. (2017). Computational fluid dynamics-based steam cracking furnace optimization using feedstock flow distribution. *AIChE J.* 63 (7), 3199–3213. doi:10.1002/aic.15669



## Nomenclature

### Abbreviations

**CIP** Coil input pressure

**CIT** Coil input temperature

**COP** Coil output pressure

**COT** Coil output temperature

**E/E** Product ratios of ethylene to ethane

**M/P** Product ratios of methane to propylene

**P/E** Product ratios of propylene to ethylene

**PIONA** N-paraffins, iso-paraffins, olefins, naphthenes, aromatics

### Superscripts

**cal** Calculated

**exp** Experimental