Eurasian Chem.-Technol. J. 27 (2025) 111–125
The journal of the Institute of Combustion Problems https://ect-journal.kz

https://doi.org/10.18321/ectj1659

Computational Modeling and Machine Learning for Predicting the Volumetric Flows in Crude Distillation Units: A Detailed Simulation and Validation Approach

Daniel Chuquín-Vasco[1]*, Julian Osorio-Getial[2], Nelson Chuquín-Vasco[3], Juan Chuquín-Vasco[3], Diana Aguirre-Ruiz[3] and Fernando Mejía-Peñafiel[3]

[1]Universitat Politécnica de Valencia, Valencia, España
[2]SOLMA, Advanced Mechanical Solutions, Mechanical Engineering, and Construction Services, Quito, Ecuador
[3]Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador

Article info

Received:
20 December 2024

Received in revised form:
2 February 2025

Accepted:
26 April 2025

Keywords:
Artificial Neuronal Networks
Crude
DWSIM
MATLAB
Naphta

Abstract

This research presents a predictive model based on Artificial Neural Networks (ANNs) for the prediction of molar flows in Crude Distillation Units (CDUs). Through rigorous simulation in DWSIM, a database of 350 points was generated, correlating the True Boiling Point (TBP) distillation temperatures of crude oil with the volumetric flows of light and heavy naphtha, distillates, and residue. An ANN with 10 inputs, 20 hidden neurons, and 5 outputs was trained using Levenberg-Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG) algorithms. The BR algorithm demonstrated superior performance, achieving a mean squared error (MSE) of 2.6904E-04 and a regression coefficient (R) of 0.9971 during the testing phase. Validation with experimental data confirmed the accuracy of the model, with average percentage errors of less than 0.68% for all products except residue (6.4%). ANOVA analysis (95% confidence) corroborated the statistical robustness of the ANN. This predictive tool will allow for the optimization of CDU design and operation, with a focus on energy efficiency and minimizing environmental impact. The study discusses the implications for real-time integration with control systems.

1. Introduction

Crude oil is a highly demanded raw material worldwide and has been exploited for over a century. Current research efforts focus on optimizing and maximizing the utilization of petroleum products. To address this issue, technology (modeling and/or optimization) is one of the most viable alternatives for the industry, as there are no costs or losses associated with these types of studies. Chemical process simulation is a powerful tool for designing, studying, optimizing, experimenting with, and redesigning processes in the industry. Additionally, artificial intelligence models are currently employed for the prediction and optimization of processes. In this study, an Artificial Neural Network (ANN) is developed with the objective of applying it to a Crude Oil Distillation Unit (CDU) to predict the final products of interest based on the feed conditions to the distillation column.

1.1 Rigorous simulations in crude oil distillation

Pannocchia et al. [1] presented the implementation of a multivariable predictive controller commonly used in the commercial field, applied to a simulated crude oil distillation process with high precision. After detailing the main characteristics of the process and the controller, they illustrate how it is possible to establish a connection between the simulation and the control environments. Various simulation results are shown, displaying characteristic changes in product quality and transitions between different types of crude oil.

*Corresponding author.
E-mail address: daniel.chuquin@unach.edu.ec

In an industrial process, Chang et al. [2] performed a dynamic simulation to identify optimal operating conditions in a 56-tray CDU with a capacity of 65,000 barrels per day. The developed model incorporates non-linear equations that analyze the fluid behavior in the column, drip loads, jet flooding, velocity limitations at the orifices, and potential blockages at the drains. This detailed analysis of the dynamic behavior of the unit during the start-up and shutdown phases is theoretical. It plays a crucial role in ensuring the efficient operation of such units.

Due to the complexity of interactions, such as feedstock variability and high process integration, modeling crude oil distillation processes is challenging. Gadalla et al. [3] incorporated a systematic algorithm to modernize an operating crude oil distillation column. The model was applied to a CDU processing a flow of 100,000 barrels per day. The results matched well with accurate data. The model was then used in optimization studies to reduce energy consumption and minimize carbon dioxide (CO_2) emissions.

1.2 Reconditioning with the use of simulators

Yan et al. [4] modernized and optimized a crude oil distillation unit (CDU) through exergy loss analysis. In this study, a qualitative and quantitative analysis was developed to improve the process's energy efficiency. The exergy loss of critical components in the CDU, including condensers, furnaces, and distillation columns, was analyzed. After optimization, energy efficiency increased by 12.5%, while total annual consumption decreased by 28.7%, keeping each product's flow rates and quality constant.

Enríquez-Gutiérrez et al. [5] used hydraulic correlations to improve the internal components of distillation columns, replacing existing plates with higher capacity ones and implementing structured packing. The objective of the study was to maximize the heat recovery of the process and the production of oil and derivatives.

Gadalla et al. [6] proposed optimizing the distillation columns and heat exchangers simultaneously to minimize the process's energy consumption and gas emissions and maximize production by modifying the structural conditions and interactions between the distillation system and the heat recovery system. Consequently, various modernization solutions were identified, ranging from no modifications to extensive upgrades involving additional heat exchanger areas and including supplementary units or equipment.

1.3 ANN as a prediction tool in chemical industries

Artificial neural networks (ANN) can predict any non-linear system through a learning process. This makes them useful artificial intelligence tools for modeling and representing the behaviour of highly non-linear systems without restrictions. This learning capacity gives them skills in classification, pattern recognition, error detection, and time series prediction [7–12].

Sharma and Singh [13] developed an ANN to manage the highly non-linear behavior of tert-amyl methyl ether (TAME). They used a conventional Proportional-Integral-Derivative (PID) controller and a neural network predictive control (NNPC) model. Data analysis concluded that the NNPC provided superior and smoother control performance than the PID controller.

Porrazzo et al. [14] presented a promising application of membrane distillation for desalination, particularly in renewable energy use. The ANN was trained to analyze process performance using varying process conditions. Its main objective was to propose improvements in process control and optimization based on analyzing distillate production, feed flow, solar radiation, and feed temperature.

Ochoa-Estopier et al. [15] used ANN to optimize production performance in a CDU with integrated heat exchange. They analyzed the energy conditions and efficiency of the equipment, heat recovery stages, and the technical feasibility of the process to find the optimal operating parameters of the system.

For a comprehensive application of neural networks in crude oil distillation, Liau et al. [16] developed an expert system using ANNs to optimize and maximize the oil production rate while maintaining required quality standards. This system utilized data from a CDU operating system, with crude oil properties as input variables and oil quality as output. It provided real-time operational information to operators regarding changes in crude oil properties.

Motlaghi et al. [17] developed an expert system using ANN models to predict the required product flow and temperature based on input feed characteristics. This system aimed to minimize model output errors and maximize oil production rates using data from an operating refinery.

In industrial processes, it is currently necessary and innovative to use artificial neural network (ANN) methodologies. Although there are no previous studies demonstrating the importance of using these methodologies, a CDU that utilizes neural networks can leverage their predictive capabilities,

thereby enhancing valuable aspects of the production of Light Naphtha (LN), Heavy Naphtha (HN), Light Distillate (LD), and Heavy Distillate (HD).

Our study proposes an artificial neural network (ANN) capable of simultaneously predicting the volumetric flow rates of five key products from a crude oil distillation unit (CDU): light naphtha (LN), heavy naphtha (HN), light distillate (LD), heavy distillate (HD), and residue. Unlike previous studies [16, 17] focusing on one or two specific products or using limited industrial databases, this work employs a robust database generated through detailed DWSIM simulations with 25 pseudocomponents and multiple operating conditions. In addition, three training algorithms (LM, SCG, and BR) are evaluated and compared. This methodology allows for accurate predictions and lays the groundwork for future integration into real-time control systems and energy optimization strategies.

2. Materials and Methods

2.1 Process Description

A Crude Distillation Unit (CDU), adapted from [18] is depicted in Fig. 1. The process consists of four columns: CDU, S-S (HN), S-S (LD), and S-S (HD). The last three are side strippers that operate with steam at high temperatures to remove light hydrocarbons.

The distillation tower analyzed in this study is based on a crude oil that processes 100,000 barrels per day (BPD) of light crude oil of Venezuelan origin, named “Tia Juana” LN, HN, LD,HD and Residue (R).

Table 1. TBP distillation of Tia Juana crude oil

Name	Tia Juana light
Country of origin	Venezuela
Type	Light
Density (kg/m^3)	867.6
API gravity	31.6
Vol. %	Temperature (°C)
0	36.1
5	64.4
10	100.6
20	163.9
30	221.1
40	278.9
50	337.2
60	397.2
70	463.9
80	545.0

Source: [18]

Table 1 details the True Boiling Point (TBP) distillation data and other properties of crude oil which is fed into a heater at a 25 °C and 2.5 bar. It undergoes a preheating to reach a temperature of 365 °C before entering the distillation unit. This column, with its impressive total of 41 stages, sees the crude oil fed at stage 37 and steam a 1200 kmol/h (260 °C, 4.5 bar) at stage 41, enhancing the efficiency of the process.

The tower features three side product outlets (HN, LD, and HD) connected to side strippers and three return streams to the CDU to improve separation efficiency and obtain the desired products. The operating parameters for the process are detailed in Tables 2-5.

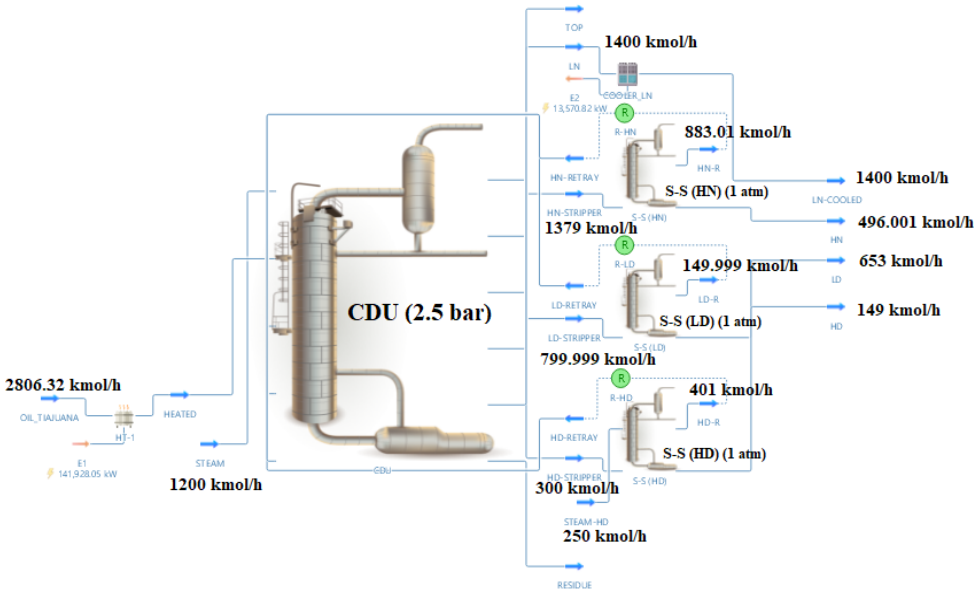

Fig. 1. Simulation of crude oil distillation with a Crude Distillation Unit (CDU).

Table 2. Crude Distillation Unit (CDU)

Parameter	Quantity	Unit
Pressure	2.5	bar
#Column stages*	41	-
#Feed stage OIL*	37	-
#Feed stage STEAM*	41	-
S-S (HN) draw and return trays*	9.8	-
S-S (LD) draw and return trays*	17.16	-
S-S (HD) draw and return trays*	27.26	-
PA1 draw and return trays*	9.7	-
PA1 Temperature	20	°C
PA1 Flowrate	1500	kmol/h
PA2 draw and return trays*	17.15	-
PA2 Temperature	50	°C
PA2 Flowrate	2500	kmol/h
PA3 draw and return trays*	27.25	-
PA3 Temperature	30	°C
PA3 Flowrate	500	kmol/h
Reflux ratio (RR)	4.17	-
Feed OIL Tia Juana molar flow	2806.32	kmol/h
Feed Temperature	365	°C
STEAM molar flow	1200	kmol/h

Notes:* Numbered from the bottom of the distillation tower. Source: [18]

Table 3. Heavy Naphtha

Parameter	Quantity	Unit
Pressure	1	atm
Phase	Liquid	-
#Column stages*	7	-
#Feed stage*	1	-
Draw tray molar flow	1379	kmol/h
Return tray molar flow	883	kmol/h
Product molar flow	496	kmol/h

Notes:* Numbered from the bottom of the distillatic tower. Source: [18]

Table 4. Light Distillate

Parameter	Quantity	Unit
Pressure	1	atm
Phase	Liquid	-
#Column stages*	8	-
#Feed stage*	1	-
Draw tray molar flow	800	kmol/h
Return tray molar flow	147	kmol/h
Product molar flow	653	kmol/h

Notes:* Numbered from the bottom of the distillation tower. Source: [18]

Table 5. Heavy Distillate

Parameter	Quantity	Unit
Pressure	1	atm
Phase	Liquid	-
#Column stages*	6	-
#Feed stage*	1	-
#Feed Steam stage*	5	kmol/h
Steam molar flow	250	kmol/h
Draw tray molar flow	300	kmol/h
Return tray molar flow	401	kmol/h
Product molar flow	149	kmol/h

Notes:* Numbered from the bottom of the distillation tower. Source: [18]

2.2 Methodology

The methodological approach used to develop the ANN is shown in Fig. 2. The operational parameters for the process simulation in DWSIM were chosen based on the conditions and parameters in Tables 1-5. Once the simulation was validated by Ibrahim et al. [18] a database was constructed from different crude oil characterizations, which served as the input variables for the ANN. Subsequently, training, validation, and testing of the ANN were conducted using the dataset generated by the simulation. Statistical analysis through graphical representations was used to evaluate the effectiveness of the ANN, ensuring a deep understanding.

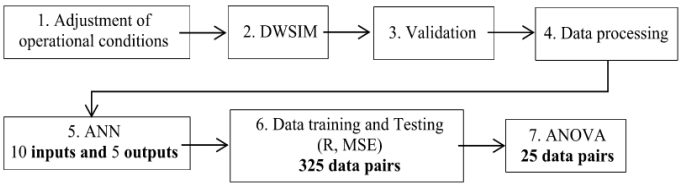

Fig. 2. Methodological scheme of the designed ANN.

2.3 DWSIM simulation

DWSIM is an open-source simulator that operates in a sequential modular steady state [19]. It is compatible with CAPE-OPEN and allows for the simulation of industrial plants involving chemical processes using thermodynamic models and highly precise unit operations [18].

For the crucial task of characterizing the crude oil, we considered 25 pseudocomponents. The methods and correlations we used for calculating properties such as temperature and critical pressure were

based on the widely recognized Riazi-Daubert model. The acentric factor, a key parameter, was determined using the Korsten model. Specific gravity, molar weight, and kinematic viscosity were calculated using the Riazi-Al-Sahhaf model, the Winn model, and the Abbott model respectively. The distillation towers, as shown in Fig. 1, were modeled using the 'Chem-Sep Column' model. Streams composed of hydrocarbons were simulated with the Peng-Robinson thermodynamic package. Raoult's Law/Ideal Gas Law/Ideal Solution/Antoine/Peng-Robinson 78 thermodynamic models were adjusted for the distillation towers [20–24]. Tables 1 to 5 specify the equipment parameters used in the process. The convergence of the simulation process was verified within a maximum of 100 iterations using the Newton's mathematical method.

2.4 Design and training of the ANN

ANNs provide numerous advantages: efficient data transmission between inputs and outputs, high learning capacity, and complex data and systems processing. These attributes collectively enhance the functionality of ANNs. It is mainly used in prediction and pattern recognition in non-linear systems [25, 26].

The literature indicates that predictive regression algorithms using ANNs require at least 50 data points. Following this guideline, 350 data pairs were created by making random adjustments to the operational conditions.

The ANN's design and training were conducted using 70% of the database for the ANN training process (227 data), 15% for validation (49 data), and 15% for the testing phase (49 data).

2.5 ANN validation

The validation of the ANN was performed using performance metrics such as the mean square error (MSE) to assess the accuracy between predicted and actual values. In addition, the ANOVA methodology was applied to analyze the performance of the ANN. The equations used are defined in Eqs. (1) and (2) of the study [27, 28].

To enhance the ANN's accuracy, efficiency, and reliability, an iterative procedure was implemented to minimize the MSE and improve the correlation coefficients, continuously fine-tuning the ANN parameters.

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y - y')^2 \quad (1)$$

$$R = \frac{n\sum_{i=1}^{n}(y'y) - [\sum_{i=1}^{n}y'][\sum_{i=1}^{n}y]}{\sqrt{[n\sum_{i=1}^{n}y^2 - [\sum_{i=1}^{n}y^2]\,[n\sum_{i=1}^{n}y'^2 - [\sum_{i=1}^{n}y'^2]}} \quad (2)$$

where *n* refers to the number of observations, *y* indicates the experimental data (simulation), and *y'* signifies the predicted targets (ANN).

3. Results and discussion

3.1 Simulation validation

Table 6 summarizes the deviations in the volumetric flow rates of interest for the distillation products of the Crude Distillation Unit (CDU). As can be seen for LD and RESIDUE, the errors are greater than 10%, this can be attributed to the fact that DWSIM uses generalized correlations (Riazi-Daubert and Winn) to characterize the physicochemical properties of the 25 crude oil pseudocomponents, although these correlations are adequate, they can introduce errors in the extreme fractions of the cut (such as LD and residue), which are more sensitive to slight variations in the TBP curve. Furthermore, the reference data for validation come from Aspen Plus, with possible differences in thermodynamic models and internal column configurations, which may also contribute to discrepancies. It should be noted that the mean deviation is 8.5%.

Table 6. Simulation validation (Volumetric flow m^3/h)

Parameter	Aspen Plus Ibrahim et al. [18]	DWSIM	Error (%)
LN	103.5	103.498	0.002
HN	78.2	77.4107	1.009
LD	140.3	166.073	18.370
HD	48.1	51.4938	7.056
RESIDUE	292.5	339.481	16.062
			Average = 8.5%

3.2 Design and Training of the ANN

Based on the ASTM D86 Method [29], the volume percentages allow us to adequately represent the composition of the crude oil and its fractions; the ANN (Fig. 3) was designed with ten input parameters: Temperature (°C) at 0% Vol, 5% Vol, 10% Vol, 20% Vol, 30% Vol, 40% Vol, 50% Vol, 60% Vol, 70% Vol and 80% Vol and five output parameters: volumetric flow rate (m^3/h) of LN, HN, LD, HD and residue.

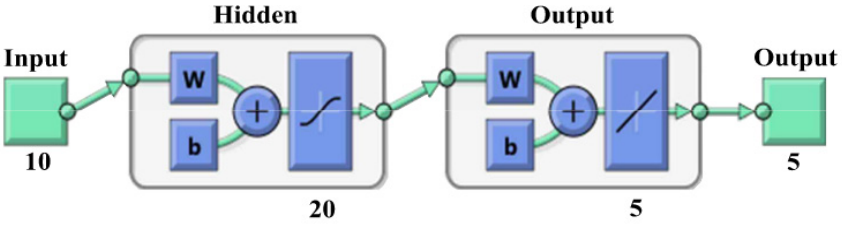

Fig. 3. Schematic of the designed ANN.

The ANN designed for this study consists of a single hidden layer with 20 neurons, selected after evaluating configurations between 10 and 100 neurons. The sigmoid activation function (logsig) was used in the hidden layer, and a linear function (purelin) was used in the output layer, which is suitable for regression problems. Three training algorithms were compared: Levenberg-Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG). The learning rate for LM and SCG was kept at its default value (0.01), while it was automatically adjusted in BR. Hyperparameter tuning was performed using an empirical approach, evaluating model performance based on mean square error (MSE) and R values in the training, validation, and testing phases.

3.3 ANN topology

3.3.1 ANN training algorithm

In this investigation, was employed with three distinct training algorithms: the Levenberg-Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient Backpropagation (SCG) algorithms. These methodologies have exhibited superior performance compared to other commonly used algorithms, showcasing a more effective convergence towards solutions characterized by reduced MSE values, as evidenced in the literature [30].

The neurons was adjusted to analyze how the hidden layer affects the results. This variation allowed the model accuracy to be assessed using R and MSE. The three training algorithms (Table 7) were evaluated, taking into account approaches used in predictive models [31–34].

This report presents the results of the ANN training using three algorithms provided by MATLAB. The purpose of this training was to identify the optimal configuration for each algorithm. At the end of the training, it was found that for LM, BR, and SCG, the optimal configuration was 20 hidden neurons (Testing), with the following performance metrics: for LM, MSE = 0.0008, and R = 0.9949; for BR, MSE = 0.0001 and R = 0.9986; and SCG, MSE = 0.0015 and R = 0.9824.

Bayesian regularization (BR) produced the most accurate and reliable model for predicting the experimental data, achieving an MSE of 0.0001 and reaching the highest R-value of 0.9986 (Table 7). Despite being computationally more intensive than other methods, Bayesian regularization is well suited for robust generalizations from limited data sets or intricate, thus outperforming techniques such as Levenberg-Marquardt [35]. Using Bayesian regression, it was identified that, unlike other models, this approach simplifies system optimization, thereby facilitating the energy optimization of the process [36].

3.3.2 Selection of the number of neurons in the hidden layer

Determining the optimal number of neurons facilitates locating the local minimum, which is a point where the error function reaches its lowest value.

Our analysis of the data in Table 7 has revealed a significant finding. Using 20 neurons in the hidden layer has resulted in the highest R-values during the testing phase, reaching a peak of 0.9986. This was accompanied by corresponding MSE values of 0.00014. These results strongly indicate that 20 neurons are the most suitable for this hidden layer, as they have elevated R-values and decreased error rates, thereby suggesting the most efficient model performance.

In this investigation, an ANN model was meticulously constructed utilizing the MATLAB Neural Net Fitting App R2020b. After a comprehensive analysis, it was determined that the ANN architecture consists of ten inputs, a single hidden layer (20 neurons), and five outputs.

Table 7. R and MSE values

# Hidden neurons		LM		BR		SCG	
		R	MSE	R	MSE	R	MSE
10	Training	0.00075	0.9919	0.00026	0.9971	0.0025	0.9730
	Validation	0.00174	0.9831	0	0	0.0028	0.9701
	Testing	0.00163	0.9833	0.00097	0.9916	0.0032	0.9377
20	Training	0.00010	0.9989	0.00009	0.9989	0.0014	0.9849
	Validation	0.00088	0.9906	0	0	0.0020	0.9796
	Testing	0.00052	0.9949	0.00014	0.9986	0.0015	0.9824
30	Training	0.00009	0.9989	4.2E-05	0.9995	0.0026	0.9733
	Validation	0.00105	0.9890	0	0	0.0046	0.9505
	Testing	0.00082	0.9910	0.00082	0.9912	0.0059	0.9271
40	Training	0.00058	0.9939	1.4E-05	0.9998	0.0019	0.9794
	Validation	0.00312	0.9667	0	0	0.0058	0.9386
	Testing	0.00278	0.9699	0.00183	0.9817	0.0045	0.9525
50	Training	0.00009	0.9990	7.4E-06	0.9999	0.0016	0.9824
	Validation	0.00176	0.9817	0	0	0.0050	0.9474
	Testing	0.00169	0.9838	0.00162	0.9832	0.0038	0.9598
60	Training	0.00033	0.9965	1.9E-06	0.9999	0.0011	0.9886
	Validation	0.00511	0.9429	0	0	0.0060	0.9395
	Testing	0.00620	0.9333	0.00107	0.9893	0.0054	0.9389
70	Training	4.3E-05	0.9995	8.1E-07	0.9999	0.0010	0.9888
	Validation	0.00480	0.9539	0	0	0.0059	0.9387
	Testing	0.00540	0.9436	0.00151	0.9844	0.0062	0.9383
80	Training	0.00052	0.9949	8.2E-08	0.9999	0.0003	0.9962
	Validation	0.00807	0.9062	0	0	0.0047	0.9410
	Testing	0.00597	0.9397	0.00186	0.9826	0.0031	0.9667
90	Training	5.6E-05	0.9993	1.1E-12	0.9999	0.0004	0.9951
	Validation	0.00620	0.9446	0	0	0.0069	0.9294
	Testing	0.00832	0.9277	0.00074	0.9927	0.0056	0.9415
100	Training	0.00111	0.9898	1.1E-12	0.9999	0.0001	0.9989
	Validation	0.01056	0.9036	0	0	0.0102	0.8854
	Testing	0.01046	0.8948	0.00145	0.9856	0.0089	0.9182

3.3.2 ANN training and testing

Table 8 illustrates the MSE values observed during the training and testing stages of the ANN process. This algorithm plays a crucial role in enhancing the robustness of the model, potentially rendering the validation process unnecessary. The obtained MSE values of 7.7695 E-05 for the training phase and 2.6904 E-04 for the testing phase demonstrate the ANN model's satisfactory performance and its ability to produce accurate predictions. Figure 4 represents the MSE in the training, ultimately converging to a value of 7.7695 E-05. The MSE performance on the training dataset (train) approaches zero, underscoring the exceptional predictive capacity of the network.

Table 8. ANN MSE values

PHASE	MSE
Train performance (training)	7.7695 E-05
Test performance (testing)	2.6904 E-04
Performance	1.0654 E-04

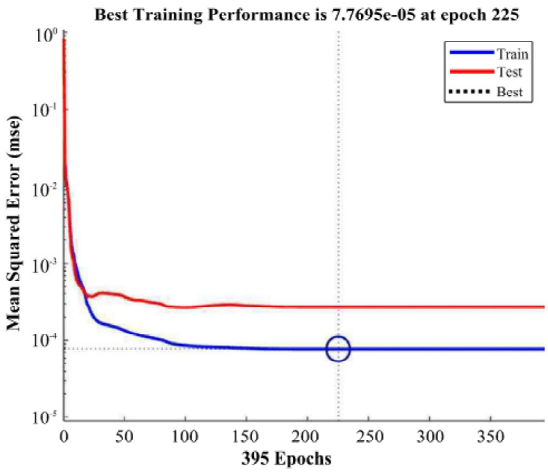

Fig. 4. ANN training performance (MSE).

The R values in training and testing phases were 0.995 and 0.934 for an overall R-value of 0.983 (Fig. 5). These significant R-values indicate a satisfactory correlation between the projected results and the intended objectives. R and MSE ranges were established to evaluate the ANN's efficiency. The R evaluation range was greater than 0.95, and MSE was less than 0.028. Compliance with these criteria validates that the ANN constitutes a reliable and accurate predictive model.

3.4 Prediction model for volumetric flows of LN, HN, LD, HD and RESIDUE

Figures 6-10 show the predictions based on ANN training data of the volumetric fluxes of CDU products: LN, HN, LD, HD, and RESIDUE. The developed model closely matches the experimental data, showing the robustness and suitability of ANNs for prediction.

The average percentage error (%E) of the predictions (Figs. 6-10) are: 0.1482% for LN, 0.43921% for HN, 0.14028% for LD, 0.16288% for HD, and 6.4% for residue.

Our ANN model shows significant accuracy and predictive capability improvements compared with previous studies [16, 17]. Previous results reported lower correlation coefficients (R) (0.947), and our model achieved an R of 0.9986 in the testing phase, thus providing a more accurate fit and reducing the dispersion between actual and predicted data. Furthermore, our model simultaneously predicts the flow rates of five main products (LN, HN, LD, HD, and residue). In future studies, we will consider energy and environmental factors for a hybrid optimization approach that combines genetic algorithms with artificial neural net-works.

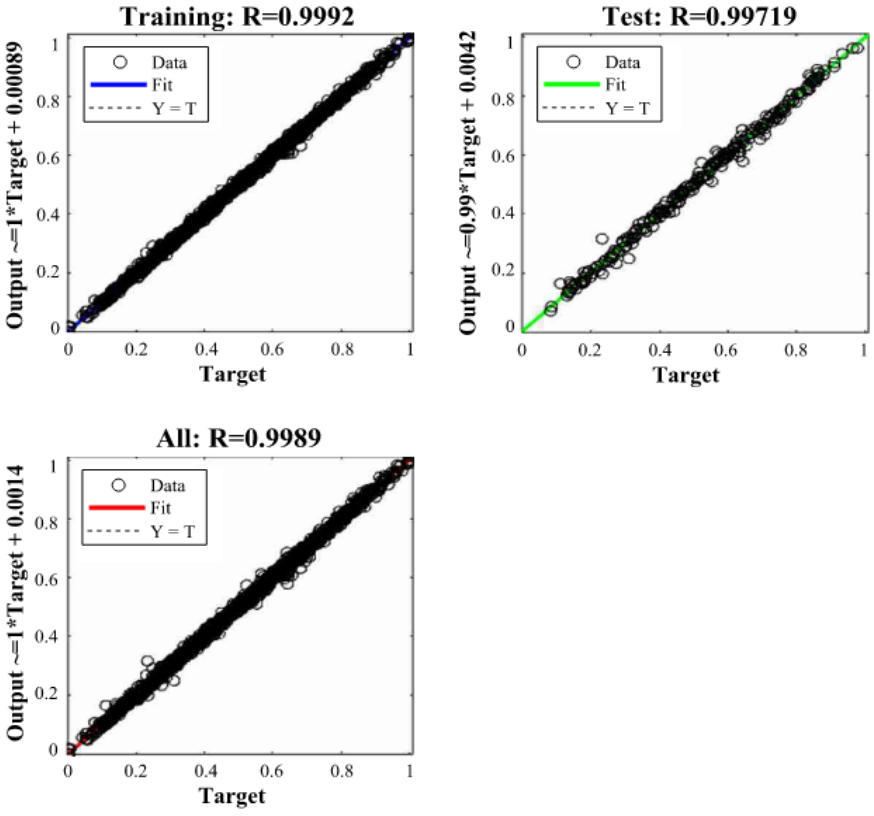

Fig. 5. ANN: R-values.

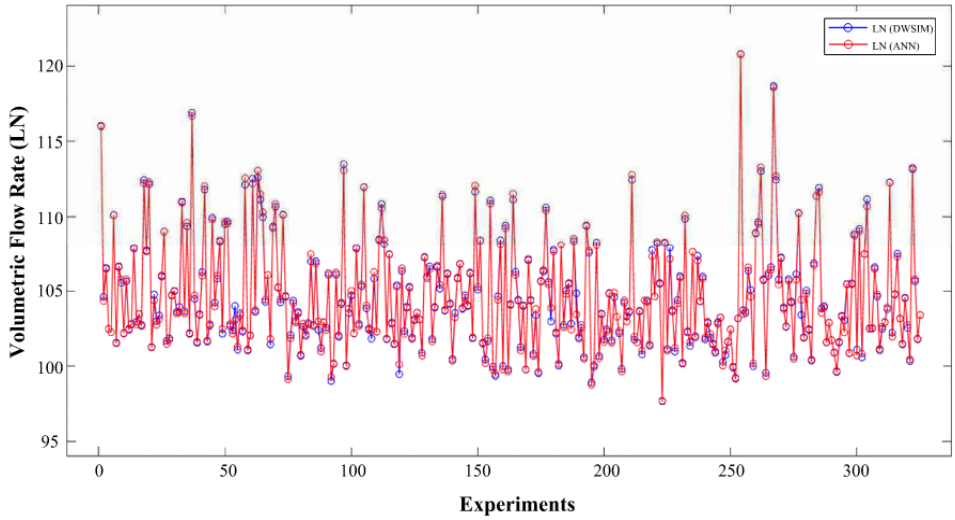

Fig. 6. Experimental (DWSIM) vs Predictions (ANN) - LN Volumetric Flow Rate.

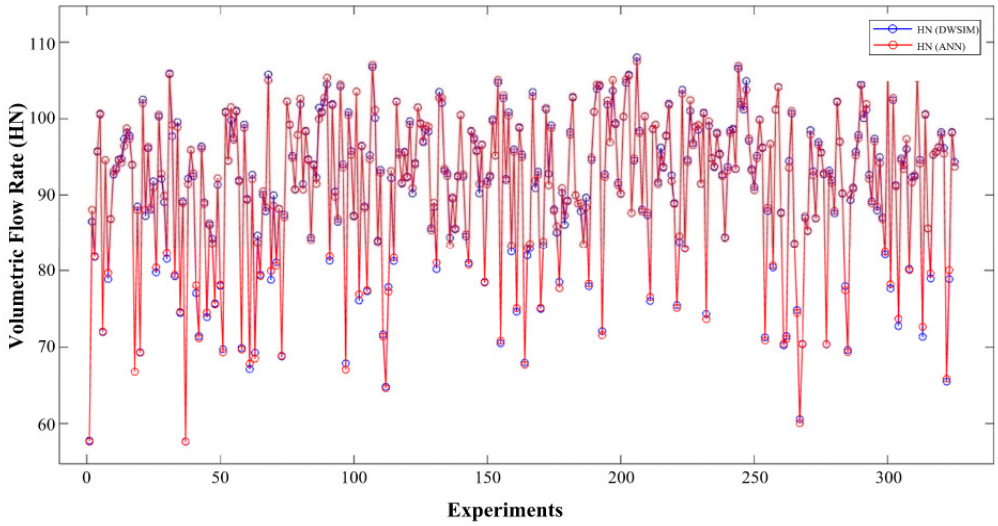

Fig. 7. Experimental (DWSIM) vs Predictions (ANN) - HN Volumetric Flow Rate.

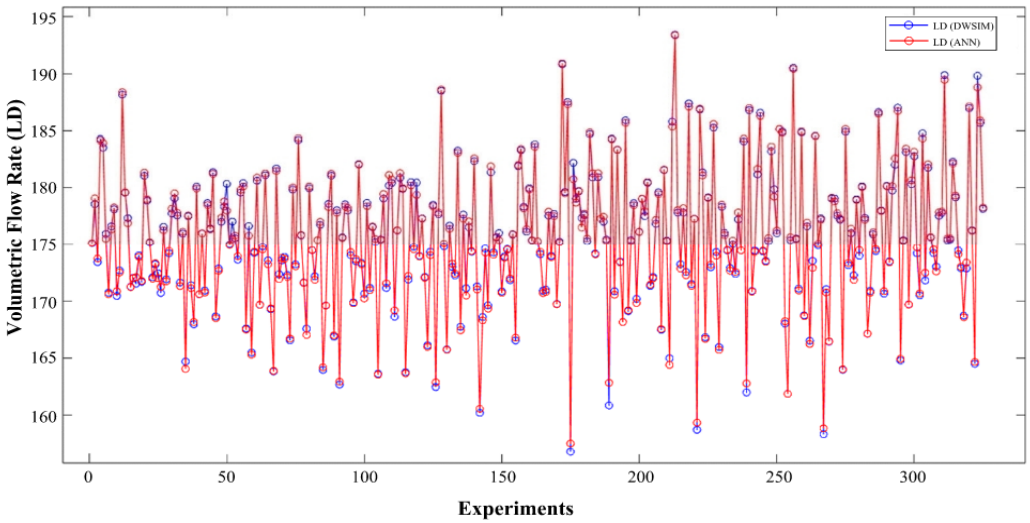

Fig. 8. Experimental (DWSIM) vs Predictions (ANN) - LD Volumetric Flow Rate.

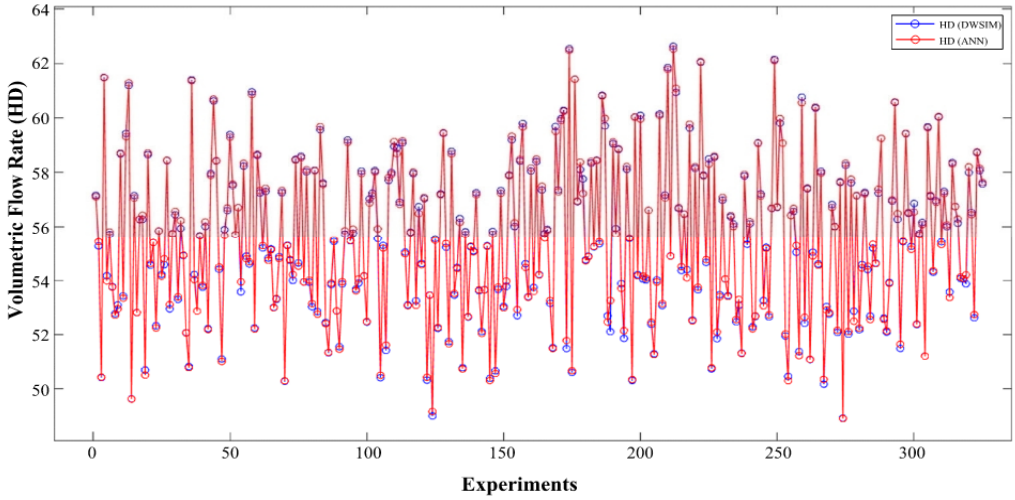

Fig. 9. Experimental (DWSIM) vs Predictions (ANN) - HD Volumetric Flow Rate.

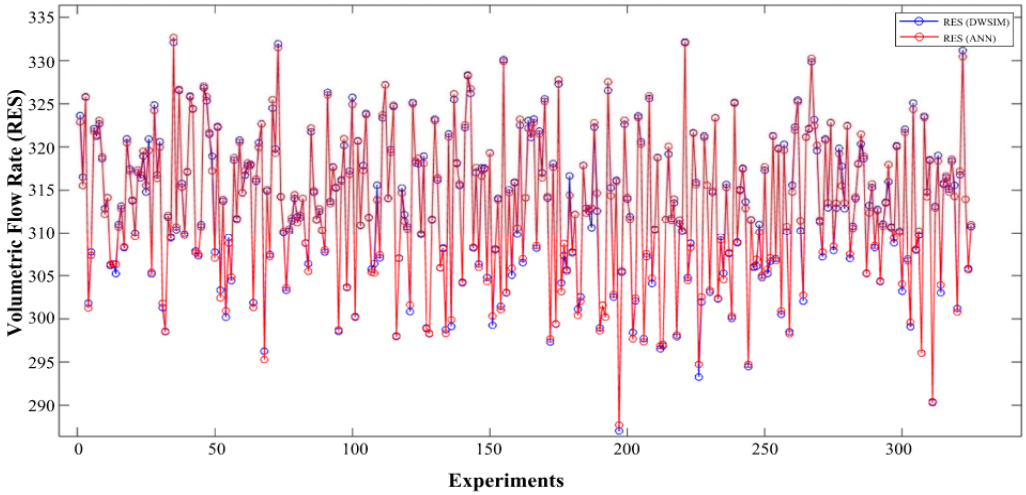

Fig. 10. Experimental (DWSIM) vs Predictions (ANN) - Residue Volumetric Flow Rate.

Hybrid methods combine various design approaches, such as utilizing detailed simulations to produce training datasets for AI models or employing heuristic optimization techniques to enhance the results [37, 38].

3.5 ANN model verification

The evaluative capacity of the Artificial Neural Network (ANN) for determining the Volumetric Flow Rate of LN, HN, LD, HD, and RESIDUE was evaluated using a dataset comprising 25 random input values that were unknown to the ANN. This dataset was carefully selected to represent a wide range of operating conditions and was not used during the training phase of the ANN. The findings reveal a strong correspondence among the observations and the predictions. This implies that the ANN exhibits remarkable forecasting proficiency for the volumetric flow rates involved in the distillation process (Figs. 11-15).

To ensure the validity of the ANN's predictions, this study conducted comprehensive ANOVA analyses using Statgraphics 19 [39]. The outcomes of these tests, summarized in Table 9, reveal that all the P values (indicating the significance level in statistical tests) exceed 0.05. This rigorous analysis suggests the absence of statistically notable distinctions among the averages of the actual observations and the model predictions. Therefore, these statistical assessments provide strong confirmation that the developed ANN is statistically reliable in forecasting the volumetric flow rates of LN, HN, LD, HD, and RESIDUE with a 95% confidence level.

Although 350 data points were used for training, validating, and testing the ANN model, this size was considered sufficient due to the quality of the data generated through rigorous simulations in DWSIM and the selection of input variables. Furthermore, previous studies [16, 17] have demonstrated prom-

ising results in regression models with similar datasets. The model's validity was supported by ANOVA analysis with a 95% confidence interval and the high performance obtained (R = 0.9986, MSE = 0.0001). Although data augmentation techniques were not applied at this stage, their incorporation and hybrid models are considered in future research.

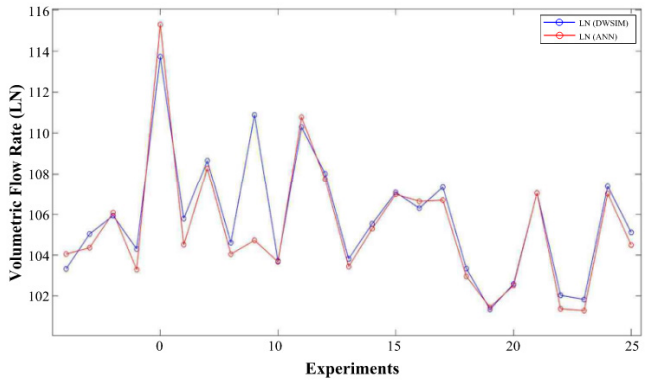

Fig. 11. Experimental and prediction data of LN volumetric flow rate.

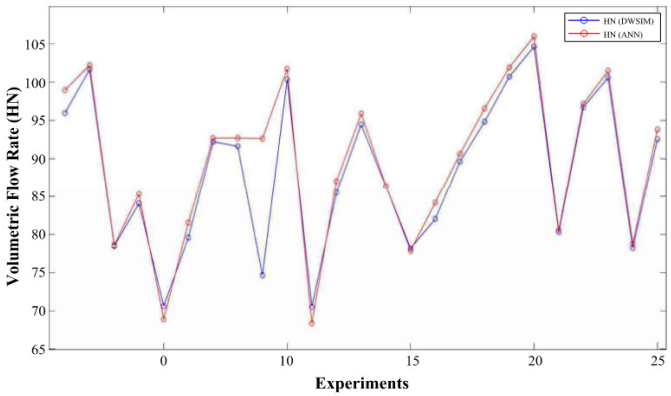

Fig. 12. Experimental and predictions data of HN volumetric flow rate.

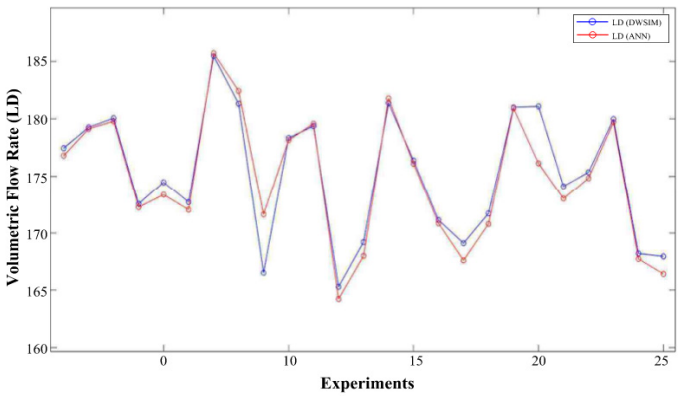

Fig. 13. Experimental and predictions data of LD volumetric flow rate.

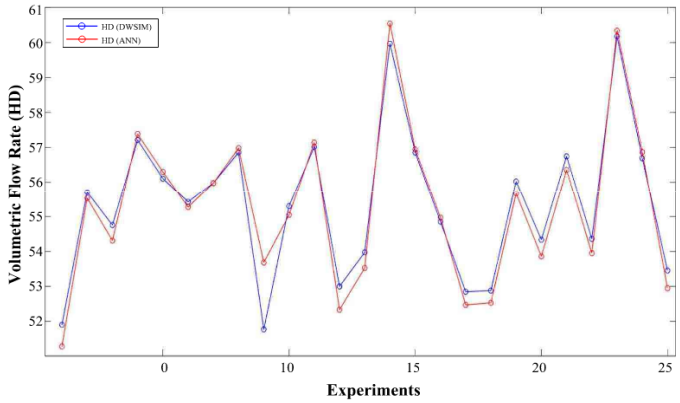

Fig. 14. Experimental and predictions data of HD volumetric flow rate.

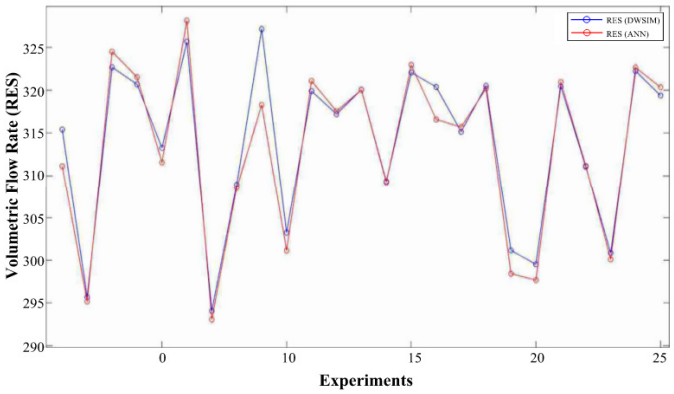

Fig. 15. Experimental and predictions data of RESIDUE volumetric flow rate.

3.6 Benefits of using ANNs for process optimization

ANNs have the potential to optimize various parameters as pressure, concentrations, temperature, reactants and flow rates. This not only enhances the effectiveness and efficiency of industrial operations but also leads to significant cost savings. By reducing resource usage while enhancing system productivity, ANNs can contribute to a more sustainable and profitable operation.

One of ANNs' characteristics is their adaptability, which can be used in energy and environmental optimization processes. ANNs can assess a process's optimal conditions to improve efficiency. In addition, the ability of ANNs to forecast process behavior and predict outcomes based on operating conditions allows operators to make preventive decisions, contributing to energy and environmental sustainability. A schematic representation in Fig. 16 illustrates the fundamental steps and tasks essential for integrating the ANN into industrial processes.

Table 9. ANOVA results comparing predicted and experimental volumetric flow rates

Source	Sum of squares	DO	Mean square	F-Value	P-value
			LN		
Inter groups	2.42205	1	2.42205	0.26	0.6131
Intra groups	**448.742**	**48**	**9.3488**		
Total (Corr.)	451.164	49			
			HN		
Inter groups	27.8574	1	27.8574	0.26	0.6109
Intra groups	**5098.23**	**48**	**106.213**		
Total (Corr.)	5126.09	49			
			LD		
Inter groups	2.18015	1	2.18015	0.07	0.7930
Intra groups	**1502.45**	**48**	**31.3011**		
Total (Corr.)	1504.63	49			
			HD		
Inter groups	0.0782628	1	0.0782628	0.02	0.9009
Intra groups	**239.782**	**48**	**4.99546**		
Total (Corr.)	239.86	49			
			RESIDUE		
Inter groups	6.71428	1	6.71428	0.07	0.7953
Intra groups	4733.2	48	98.6084		
Total (Corr.)	4739.92	49			

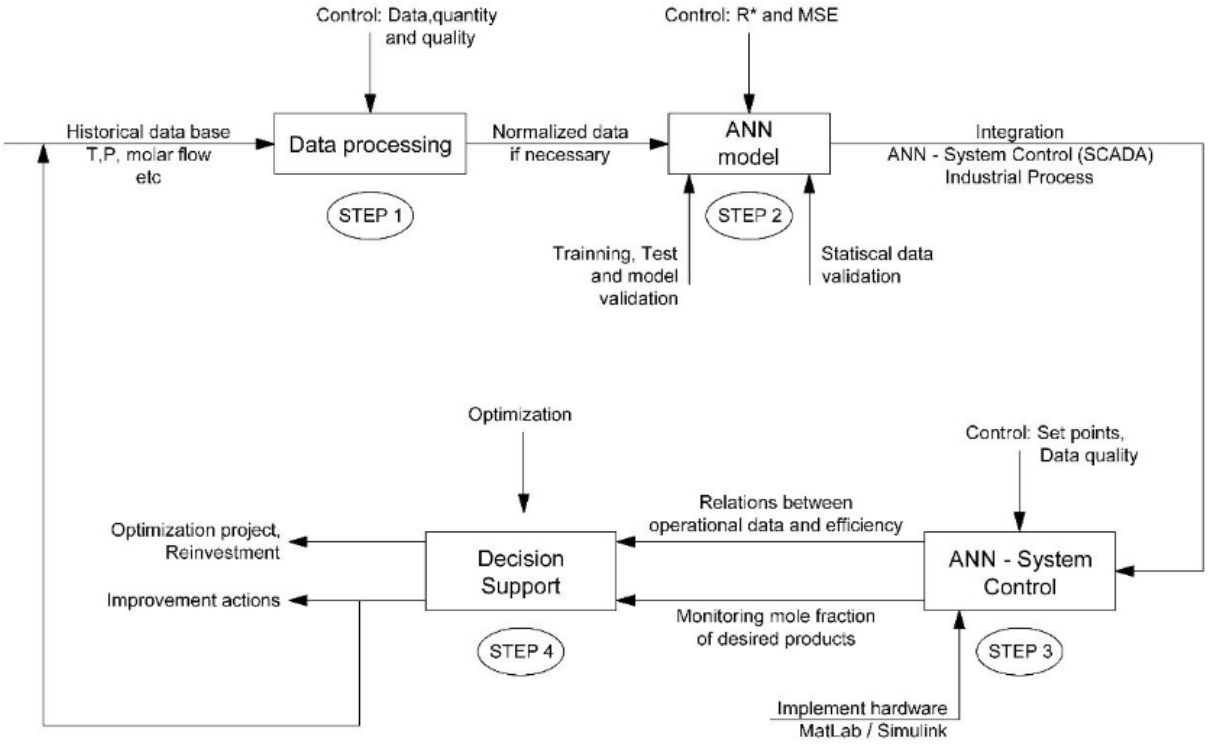

Fig. 16. Generalized diagram to implement the ANN.

To ensure successful real-time incorporation, the subsequent suggestions are proposed: Define the goals of the ANN, establish the historical data set, which is a collection of past process data, for retraining the ANN, this data is used to teach the ANN how to respond to different inputs and conditions, preprocess the data set, train the ANN and link it to the hardware of the automated control system. Establish real-time monitoring strategies to assess ANN functionality and generate an automatic system for continuous ANN training to improve its performance continuously.

4. Conclusions

This investigation harnessed the power of an Artificial Neural Network (ANN) to precisely predict the desired products, including LN, HN, LD, HD, and RESIDUE. The ANN was meticulously trained using a dataset of 325 entries and rigorously validated with 25 additional data points. This thorough process ensured the reliability of the simulation, leading to the establishment of a robust database that identified the variables significantly impacting the outcomes of interest in the process. Specifically, the temperatures corresponding to % VAP in the TBP distillation curves were considered inputs (10 data), while the products from the fractionation tower were considered as outputs (5 data).

During the neural network design, we evaluated three algorithms available in MATLAB (LM, BR, and SCG). These algorithms evaluate the values of R (tendency to 1) and MSE (tendency to 0). After analysis, the Bayesian Regularization (BR) algorithm was selected due to its MSE of 2.6904 E-04 and R of 0.9971 during the network testing phase. The network architecture was defined with ten inputs, 20 hidden neurons, and five outputs. Statistical analyses confirm the accuracy of the ANN in predicting mass fractions in crude distillation plant processes. This study indicates that the developed ANN can be used as a dependable tool for forecasting the volumetric flows of a Crude Distillation Unit (CDU) based on TBP distillation curves. Future investigations may explore the residue derived from the CDU for subsequent vacuum distillation processes. In addition, subsequent research endeavours are anticipated to concentrate on hybrid energy optimizing operations within a current plant setting.

References

[1]. G. Pannocchia, L. Gallinelli, A. Brambilla, G. Marchetti, Rigorous simulation and model predictive control of a crude distillation unit, *IFAC Proc. Vol.* 39 (2006) 635–340. DOI: 10.3182/20060402-4-BR-2902.00635

[2]. B. Chang, S. Lee, H. Kwon, I. Moon, Rigorous industrial dynamic simulation of a crude distillation unit considered valve tray rating parameters, *Comput. Chem. Eng.* 22 (1998) S863–S866. DOI: 10.1016/S0098-1354(98)00167-7

[3]. M.A. Gadalla, O.Y. Abdelaziz, D.A. Kamel, F.H. Ashour, A rigorous simulation-based procedure for retrofitting an existing Egyptian refinery distillation unit, *Energy* 83 (2015) 756–765. DOI: 10.1016/j.energy.2015.02.085

[4]. C. Yan, L. Lv, S. Wei, et al., Application of retrofitted design and optimization framework based on the exergy analysis to a crude oil distillation plant, *Appl. Therm. Eng.* 154 (2019) 637–649. DOI: 10.1016/j.applthermaleng.2019.03.128

[5]. V.M. Enríquez-Gutiérrez, M. Jobson, L.M. Ochoa-Estopier, R. Smith, Retrofit of heat-integrated crude oil distillation columns, *Chem. Eng. Res. Des.* 99 (2015) 185–198. DOI: 10.1016/j.cherd.2015.02.008

[6]. M. Gadalla, D. Kamel, F. Ashour, H.N. El Din, A New Optimisation based Retrofit Approach for Revamping an Egyptian Crude Oil Distillation Unit, *Energy Procedia* 36 (2013) 454–464. DOI: 10.1016/j.egypro.2013.07.051

[7]. R. Chen, H. Qian, M.K. Khashan, et al., Triple-objective optimization using ANN+NSGA-II for an innovative biomass gasification-heat recovery process, producing electricity, coolant, and liquefied hydrogen, *Case Stud. Therm. Eng.* 60 (2024) 104647. DOI: 10.1016/j.csite.2024.104647

[8]. Y. Zhou, W. Li, Y. Song, ANN model of co-gasification process in bubbling fluidized bed: Contribution ratio analysis, *Int. J. Hydrog. Energy* 73 (2024) 10–19. DOI: 10.1016/j.ijhydene.2024.06.007

[9]. D. Antanasijević, V. Pocajt, I. Popović, et al., The forecasting of municipal waste generation using artificial neural networks and sustainability indicators, *Sustain. Sci.* 8 (2013) 37–46. DOI: 10.1007/s11625-012-0161-9

[10]. N. Elshaboury, E.M. Abdelkader, G. Alfalah, A. Al-Sakkaf, Predictive Analysis of Municipal Solid Waste Generation Using an Optimized Neural Network Model, *Processes* 9 (2021) 2045. DOI: 10.3390/pr9112045

[11]. B. Fang, J. Yu, Z. Chen, et al., Artificial intelligence for waste management in smart cities: A review, *Environ. Chem. Lett.* 21 (2023) 1959–1989. DOI: 10.1007/s10311-023-01604-3

[12]. H.K. Balsora, K.S.J.B. Joshi, A. Sharma, et al., Artificial Neural Network-Based Models for the Prediction of Biomass Pyrolysis Products from Preliminary Analysis, *Ind. Eng. Chem. Res.* 62 (2023) 14311–14319. DOI: 10.1021/acs.iecr.3c01662

[13]. N. Sharma and K. Singh, Model predictive control and neural network predictive control of TAME reactive distillation column, *Chem. Eng. Process.: Process Intensif.* 59 (2012) 9–21. DOI: 10.1016/j.cep.2012.05.003

[14]. R. Porrazzo, A. Cipollina, M. Galluzzo, G. Micale, A neural network-based optimizing control system for a seawater-desalination solar-powered membrane distillation unit, *Comput. Chem. Eng*. 54 (2013) 79–96. DOI: 10.1016/j.compchemeng.2013.03.015

[15]. L.M. Ochoa-Estopier, M. Jobson, R. Smith, Operational optimization of crude oil distillation systems using artificial neural networks, *Comput. Chem. Eng*. 59 (2013) 178–185. DOI: 10.1016/j.compchemeng.2013.05.030

[16]. L.C.K. Liau, T.C.K. Yang, M. Te Tsai, Expert system of a crude oil distillation unit for process optimization using neural networks, *Expert Syst. Appl.* 26 (2004) 247–255. DOI: 10.1016/S0957-4174(03)00139-8

[17]. S. Motlaghi, F. Jalali, M.N. Ahmadabadi, An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework, *Expert Syst. Appl.* 35 (2008) 1540–1545. DOI: 10.1016/j.eswa.2007.08.105

[18]. D. Ibrahim, M. Jobson, J. Li, G. Guillén-Gosálbez, Optimal design of flexible heat-integrated crude oil distillation units using surrogate models, *Chem. Eng. Res. Des*. 165 (2021) 280–297. DOI: 10.1016/j.cherd.2020.09.014

[19]. DWSIM, DWSIM – The Open Source Chemical Process Simulator, 2023. https://dwsim.org/ (accessed January 10, 2024).

[20]. K. Ankita Singh, D. Kundu, Comprehensive thermodynamic modeling framework for the estimation of physico-chemical properties of deep eutectic solvent-heavy crude oil system, *J. Mol. Liq.* 392 (2023) 123489. DOI: 10.1016/j.molliq.2023.123489

[21]. M.A. Jamali, A. Bissenbay, N. Nuraje, Thermodynamic Modeling and Process Simulation of Kumkol Crude Oil Refining, *Eurasian Chem.-Technol. J.* 25 (2023) 183–192. DOI: 10.18321/ectj1521

[22]. Souck Jenny, Modelling of Crude Oil Distillation – Modellering av råoljedestillation, Stockholm, 2012., p. 57. URN: urn:nbn:se:kth:diva-146195

[23]. H. Al-Muslim, I. Dincer, Thermodynamic analysis of crude oil distillation systems, *Int. J. Energy Res.* 29 (2005) 637–655. DOI: 10.1002/er.1097

[24]. J.P. Gutierrez, L.A. Benítez, J. Martínez, et al., Thermodynamic properties for the simulation of crude oil primary refining, *Int. Journal of Engineering Research and Applications* 4 (2014) 190–194. https://www.ijera.com/papers/Vol4_issue4/Version%201/AA044190194.pdf

[25]. P. Zafari, A. Ghaemi, Modeling and optimization of CO_2 capture into mixed MEA-PZ amine solutions using machine learning based on ANN and RSM models, *Results Eng.* 19 (2023) 101279. DOI: 10.1016/j.rineng.2023.101279

[26]. J. Rasulzade, Y. Maksum, M. Nogaibayeva, et al., Reduction of Material Usage in 3D Printable Structures Using Topology Optimization Accelerated with U-Net Convolutional Neural Network, *Eurasian Chem.-Technol. J.* 24 (2022) 277–286. DOI: 10.18321/ectj1471

[27]. R. Al Dwood, Q. Meng, A.W. Ibrahim, et al., A novel hybrid ANN-GB-LR model for predicting oil and gas production rate, *Flow Meas. Instrum.* 100 (2024) 102690. DOI: 10.1016/j.flowmeasinst.2024.102690

[28]. F. Abbasian, G.I. Hadaller, R.A. Fortman, et al., An Artificial Neural Network (ANN) Model to Predict Critical Heat Flux (CHF) in a CANDU Fuel Element Simulation (FES) with Various Nonuniform Axial Heat Flux Shapes and Flow Liner Creep Profiles, *Nucl. Eng. Des.* 432 (2025) 113736. DOI: 10.1016/j.nucengdes.2024.113736

[29]. R.N.G. Santos, E.R.A. Lima, M.L.L. Paredes, ASTM D86 distillation curve: Experimental analysis and premises for literature modeling, *Fuel* 284 (2021) 118958. DOI: 10.1016/j.fuel.2020.118958

[30]. M. Troiano, E. Nobile, F. Mangini, et al., A Comparative Analysis of the Bayesian Regularization and Levenberg–Marquardt Training Algorithms in Neural Networks for Small Datasets: A Metrics Prediction of Neolithic Laminar Artefacts *Information* 15 (2024) 270. DOI: 10.3390/info15050270

[31]. A. Taheri-Garavand, M. Beiranvandi, A. Ahmadi, N. Nikoloudakis, Predictive modeling of Satureja rechingeri essential oil yield and composition under water deficit and soil amendment

conditions using artificial neural networks (ANNs), *Comput. Electr. Agric.* 222 (2024) 109072. DOI: 10.1016/j.compag.2024.109072

[32]. S. Abubakar, D. Okoh, B.I. Tijjani, R.S. Said, Speed and accuracy investigations of neural network algorithms for ionospheric modelling at an equatorial region, *J. Atmos. Sol.-Terr. Phys*. 265 (2024) 106365. DOI: 10.1016/j.jastp.2024.106365

[33]. D. Chuquin-Vasco, G. Torres-Yanacallo, C. Calderón-Tapia, et al., ANN for the prediction of isobutylene dimerization through catalytic distillation for a preliminary energy and environmental evaluation, *AIMS Environ. Sci.* 11 (2024) 157–183. DOI: 10.3934/environsci.2024009

[34]. D. Chuquin-Vasco, D. Chicaiza-Sagal, C. Calderón-Tapia, et al., Forecasting mixture composition in the extractive distillation of n-hexane and ethyl acetate with n-methyl-2-pyrrolidone through ANN for a preliminary energy assessment, *AIMS Energy* 12 (2024) 439–463. DOI: 10.3934/energy.2024020

[35]. F.L. Marcos, B. Plangklang, Design and simulation of hybrid renewable systems for rural electrification, *Energy Rep*. 11 (2024) 2220–2235. DOI: 10.1016/j.egyr.2024.01.062

[36]. B. Choudhuri, R. Sen, Optimization of WEDM parameters for Machining Inconel 800 by ANN based Bayesian hybrid algorithm. *Mater. Today Proc*. 62 (2022) 1098–1101. DOI: 10.1016/j.matpr.2022.04.318

[37]. A. Tgarguifa, T. Bounahmidi, S. Fellaou, Optimal Design of the Distillation Process Using the Artificial Neural Networks Method," 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 2020, pp. 1-6. DOI: 10.1109/IRASET48871.2020.9092266

[38]. S.P. Pushkala, R.C. Panda, Design and analysis of reactive distillation for the production of isopropyl myristate, *Cleaner Chem. Eng*. 5 (2023) 100090. DOI: 10.1016/j.clce.2022.100090

[39]. Statgraphics Technologies, STATGRAPHICS® Centurion 19 User Manual, 2020. https://www.statgraphics.com