



**HAL**  
open science

# Artificial Intelligence Application for Crude Distillation Unit: An Overview

Václav Miklas, Michal Touš, Vítězslav Máša, Sin Yong Teng

► **To cite this version:**

Václav Miklas, Michal Touš, Vítězslav Máša, Sin Yong Teng. Artificial Intelligence Application for Crude Distillation Unit: An Overview. IFIP International Workshop on Artificial Intelligence for Knowledge Management (AI4KMES), Aug 2021, Montreal, QC, Canada. pp.156-168, 10.1007/978-3-030-96592-1\_12 . hal-04120814

**HAL Id: hal-04120814**

**<https://inria.hal.science/hal-04120814v1>**

Submitted on 7 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Artificial Intelligence Application for Crude Distillation Unit: An overview

Václav Miklas<sup>1</sup>[0000-0003-4790-3680], Michal Touš<sup>1</sup>[0000-0002-0956-8511] Vítězslav Máša<sup>1</sup>[0000-0002-3646-9009] and Sin Yong Teng<sup>2</sup>[0000-0002-2988-8053]

<sup>1</sup> Brno University of Technology, Institute of Process Engineering & NETME Centre, Technická 2896/2, 616 69, Brno, Czech Republic

<sup>2</sup> Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, Netherlands  
masa@fme.vutbr.cz

**Abstract.** Artificial intelligence (AI) with its efficiency for complex systems is growing in popularity in many engineering fields. The ability of an AI method to be successfully applied is highly dependent on the previous research, which makes knowledge sharing within and across fields extremely valuable. This work focuses on crude distillations units (CDU), whose energy optimization has been a tremendous challenge because of its complexity. The presented overview shows that soft sensors are the most common application of artificial intelligence for a CDU, although a number of recent publications focus on optimization problems. The approaches for optimization are very diverse, which makes them hardly applicable in the current engineering practice. This work provides a guideline for selecting the right method, but also addresses the fact that different methods excel at different problems and with different data set sizes. For neural networks (NN), this further depends on their architecture and hyperparameter adjustment. This urges future research, whose goal could be a workflow that would automatically adapt methods and perform parameter tuning with minimum user input.

**Keywords:** Crude Distillation Unit (CDU), Artificial Intelligence (AI), Machine Learning (ML), Neural Network (NN), Sustainability.

## 1 Introduction

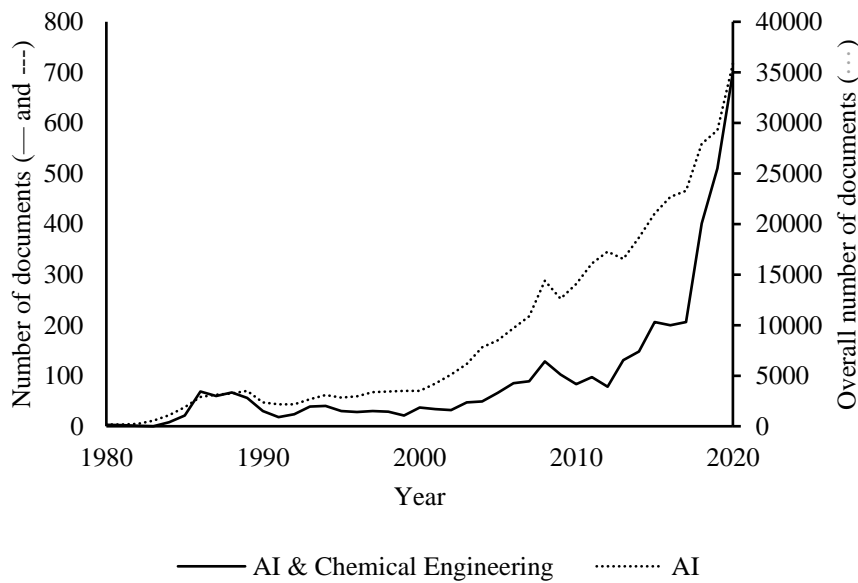
Artificial intelligence is a dynamically developed field, whose significance grew within various industries over the last few decades. While the origins of artificial intelligence (AI) and early neural networks (NN) can be traced 1940s [1], the breakthrough did not happen until the new millennium when rapid hardware development allowed for practical implementation [2].

Nowadays, AI has some extent of successful implementation in virtually all aspects of everyday life, including engineering. Some highly-cited published engineering applications range from optimization of constrained design problems [3], through multi-agent and holonic manufacturing systems [4], risk analysis and maintenance [5], to water quality prediction [6].

Some of the important engineering challenges to address using artificial intelligence are the reduction of energy consumption and related emissions, as well as sustainability of processes in general. Whereas a Scopus search “TITLE-ABS-KEY (artificial AND intelligence) OR TITLE-ABS-KEY (machine AND learning) OR TITLE-ABS-KEY (neural AND network)” returns almost 1.2 million results, mere 3,400 documents mention “sustainability” in their title, abstract or keywords.

One of the industries where the abovementioned sustainability-related keywords are relevant is petroleum refining, whose global primary energy consumption is estimated to be over 5 % [7]. Within oil refineries, the key equipment which is also related to the highest energy consumption is crude distillation unit (CDU), which might use around one-third of the refinery primary energy demands [8].

Chemical engineering, including oil refining, somewhat lagged behind the general increasing trend in artificial intelligence implementation at the beginning of the millennium. This is demonstrated in Fig. 1. However, the last couple of years showed an unprecedented growth, from around 200 publications indexed in Scopus per year between 2015 and 2017 to 700 records in 2020.

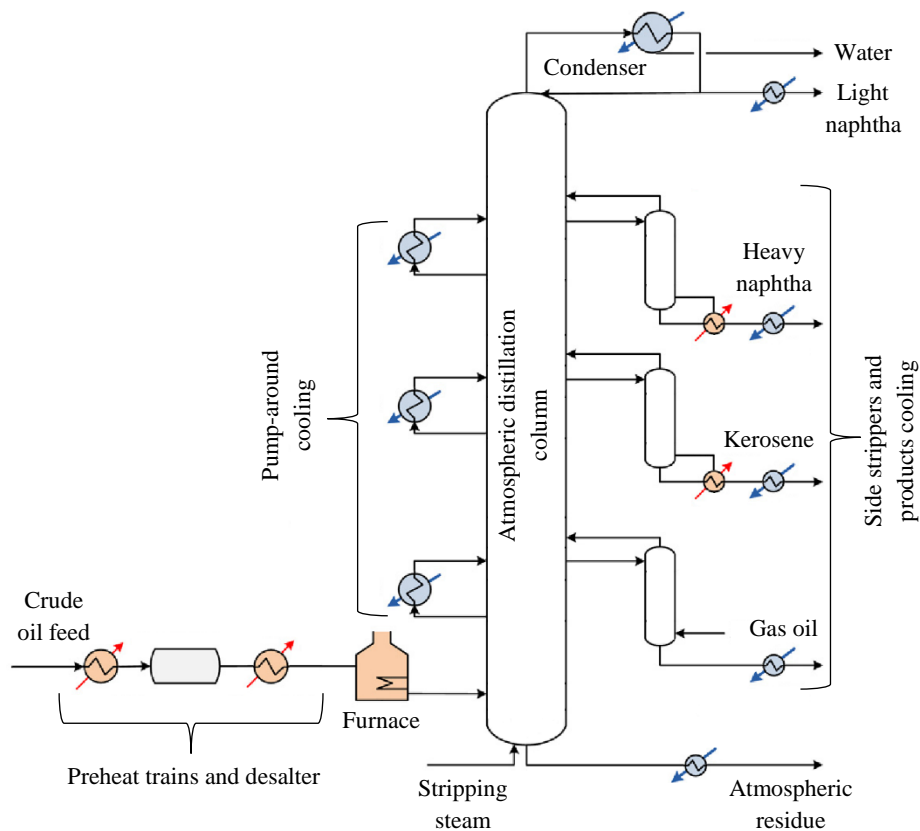


**Fig. 1.** All results for the “artificial intelligence” Scopus search compared to the same search narrowed down to the subject areas “Chemical Engineering”.

This work presents an overview of artificial intelligence applications for the CDU over the last decade, which, especially in its second half, recorded such a rapid increase in publishing activity within chemical engineering. However, only a limited number of publications deal with the energy intensity of the process, as will be shown below. The following chapters present the CDU process (2), literature search methodology (3) and the main findings in the literature (4).

## 2 Process description

A crude distillation unit (CDU) represents an essential separation step and the first of the fundamental unit operations in any petroleum refinery. Its main purpose is to separate the crude oil feed, a mixture of hydrocarbons with a wide range of boiling points, into semi-finished or final products based on their respective volatility. A common process layout of such a unit is shown in Fig. 2.



**Fig. 2.** A simplified process diagram of a typical crude distillation unit; adapted from [17].

Before entering the *atmospheric distillation column* (Fig. 2), the crude oil feed is pre-heated in a heat exchanger network or series of networks called *preheat trains* (Fig. 2). Heat is recovered from relatively hot streams elsewhere in the process. This heat integration allows for a higher energy efficiency of the unit.

The remainder of the required heat input is provided in a *furnace* (Fig. 2), at the outlet of which the crude oil is partially vaporized and enters the flash zone of the *column*. A relatively steep temperature gradient exists within the *column*, with the lowest temperature at the top. The gradient is caused and controlled by the overhead *condenser* (Fig. 2) as well as *pump-around cooling* (Fig. 2) along the column. The temperature at a specific point in the column is dependent on the vapor-liquid contact between the internal reflux stream and vapor flowing up from the flash zone. The temperature profile determines the position of side draws where different products can be extracted from the main *column*.

*Side strippers* (Fig. 2) are used to remove light ends (light hydrocarbons) from the products. This is necessary to adjust products' distillation curves and other properties according to the specification. Stripping is done by *stripping steam* (Fig. 2) injection, which is also used at the bottom of the main column, or via a reboiler.

The heaviest fraction, *atmospheric residue* (Fig. 2), leaves at the bottom of the column and is further processed in vacuum distillation and/or other processes. The number and character of lighter liquid products from the side strippers and condenser may vary, but these are commonly (from heaviest to lightest) *gas oil*, *kerosene*, *heavy naphtha*, and *light naphtha* (Fig. 2). The light ends and other non-condensable gases are separated in the *condenser* (this stream is omitted in Fig. 2 for simplicity).

Even the significantly simplified illustration and description of the unit demonstrates the high complexity of the system. This work's authors have recently dealt with a CDU system in a similar scope but in a high level of detail. The number of manipulated variables in this case study, already narrowed according to energy optimization purposes, was close to 50. The number of all variables, including dependent and nonrelevant variables, easily goes into hundreds for a typical CDU.

From the authors' experience, even conventional process simulation as the most popular method to design, optimize and troubleshoot crude distillation units can fail to obtain results, given an inadequate combination of manipulated variables. While this is not the method's drawback and often simply indicates an infeasible operating point, it calls for application of unconventional and progressive methods, effective for high-dimensional data and complex problems. For this reason, the following chapters explore the existing applications of artificial intelligence to CDU.

### 3 Literature search methodology

The goal of the methodology is to find relevant records, in which artificial intelligence (AI) is applied to a crude distillation unit (CDU) regardless of the specific purpose. To perform such search, four keyword phrases are used:

- crude distillation (the investigated process), AND
- artificial intelligence (as the most general term), OR
- machine learning (the largest subcategory of AI), OR
- neural network (the most pronounced method of AI),

The abstract and citation database Scopus is used for the search. The phrases above are looked for within titles, abstracts and keywords. Also, the year range is limited to 2011–2021. The initial search is narrowed by excluding the results which could not be retrieved (reason A), are not available in English (reason B), are not related to crude distillation or artificial intelligence (reason C), are not individual contributions (e.g., conference reviews; reason D), or use the same method or have the same purpose as in the authors' previous work (reason E).

Within the final literature set included in the overview, four main data items are collected wherever available:

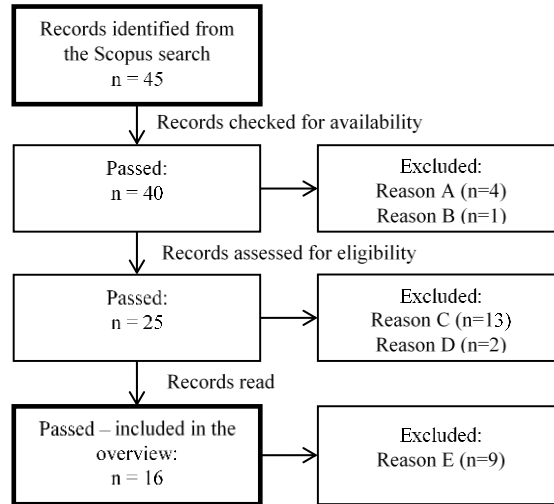
- used AI and non-AI methods,
- purpose of the work,
- size and/or division of a data set into training, test and validation sets if applicable,
- affinity to the topics of sustainability, energy and emissions.

The described multi-step procedure will lead to a relevant set of documents that can be analyzed with respect to qualitative and quantitative parameters that allow for discussion, recognition of well-covered aspects of the process and identification of potential future directions that are not as well-developed.

The main goal, however, is to provide an insight to engineers and researchers dealing with the same tasks that can already be found in the literature. Instead of “reinventing the wheel”, they can rely on the methods that have already been created and established.

## **4 Main findings in the literature**

The initial search resulted in a group of 45 records. Three following checks, which can be described as a) availability (exclusion reasons A and B), b) eligibility (exclusion reasons C and D), c) uniqueness (exclusion reason E), resulted in a final set of 16 documents that were thoroughly examined and included in the final overview. See Fig. 3 for details about how many documents were excluded in each step.



**Fig. 3.** Step-by-step selection of the final literature set from the initial search.

#### 4.1 Categorization

Based on their purpose, the 16 items can be divided into the following categories:

1. Control of the CDU (Table 1)
  - a. Soft sensors – 6 records;
  - b. Predictive control – 1 record;
2. Optimization of the CDU (Table 2)
  - a. Heat, energy and/or exergy budget – 4 records;
  - b. Cost – 2 records (1 design, 1 operation);
  - c. Non-specific opt. (1 record);
3. Other (Table 3)
  - a. Property estimation – 1 record;
  - b. Risk management – 1 record.

The results of the investigation in Tables 1–3 show what methods the authors used, the purpose of their efforts, a condensed description of their work, the size and distribution of their data set, and links to the topics of energy, emissions or sustainability (EES).

The application for soft sensors is prevalent, mostly for quantities that cannot be measured continuously, or where the related analysis is time- or cost-intensive (cut point temperature, dry point, vapour pressure).



**Table 1.** Overview of the records related to the process control.

Methods	Purpose	Description	Training / Test / Validation	EES	Ref.
1a. Soft sensors					
-K-means -K-nearest neighbors (K-NN) -Random forest (RFc) -Support Vector Classification (SVC) -multilayer perceptron (MLP)	soft sensor for the “needle penetration” analysis (UNE-EN 1426:2015)	-GridSearch used to optimize hyper parameters for each method (e.g. number of neurons in 1 hidden layer of the MLP) -94 inputs -1 output -K-means best for this application	90 % / 10 % (268 samples)	No	[9]
-multiple linear regression (MLR) -linear neural network (LNN) -multilayer perceptron (MLP) -radial basis function (RBF)	prediction of Reid vapor pressure (RVP) of the light naphtha product	-96 data samples -number of neurons and learning algorithms tested to find the best-performing combination -MLP (6-4-1) and RBF (6-7-1) best results	50 % / 25 % / 25 %	No	[10]
-deep belief network (DBN) -restricted Boltzmann machines (RBM)	soft sensor (ASTM 95% cut point temperature)	-DBN consists of a series of RBMs -16 inputs -1 output -architecture 16/20/16/1 -compared and superior to SVM, PLS, NNPLS	251/100	No	[11]
-multilayer perceptron (MLP) -least absolute shrinkage and selection operator (LASSO)	soft sensor for the estimation of the kerosene D95	-compared to sequential backward multiplayer perceptron (SBS-MLP), normalized mutual information feature selection (NMIFS) and ELM (Elman network) -25 variables -25-4-1 architecture	240 / 121	No	[12]

**Table 1 Continued.** Overview of the records related to the process control.

Methods	Purpose	Description	Training / Test / Validation	EES	Ref.
a number of statistical learning methods (partial least squares, neural networks, relevance vector machine, Bayesian linear regression, ridge regression)	soft sensor (95% cut point temperature)	-the process simulator PRO/II to generate data -relevance vector machine (RVM) was the best for this application (high accuracy, >10 times faster than NN, high generalization capacity)	225 (50) / 75	Yes	[13]
-bootstrap aggregated neural networks (BANN) -bootstrap aggregated partial least squares (BAPLS)	estimation of kerosene dry point with varying crudes	-16 inputs (selected out of 50) -Aspen HYSYS to generate simulated data, random noise added to the data -intentionally limited data set to represent limited dry point analyses -BANN for oil classification, BAPLS (partial least square) for the dry point prediction	200 / 70 / 87	No	[14]
<b>1b. Predictive control</b>					
- feedforward NN (FNN) -stochastic optimization algorithm Adam (training)	model predictive control for a CDU	-CVXOPT (quadratic programming solver used for data generation and as a benchmark) -architecture: 1072 / (1664–1920) × 3 hidden layers / 32 -NN milliseconds, QP solver tens of seconds, negligible performance loss	1500 epochs	No	[15]

The optimization purposes are almost as common as soft sensors and very often focus on or at least account for energy demands. However, the variety of used approaches make it uneasy to select the right methods for engineering applications. Authors often claim their approach is superior to others, and rightfully so – for the given problem, data set size, selected network architecture and hyperparameters.

**Table 2.** Overview of the records related to process optimization.

Methods	Purpose	Description	Training / Test / Validation	EES	Ref.
<b>2a. Heat, energy and/or exergy budget optimization</b>					
-neural network (NN) -Taguchi method -genetic algorithm (GA)	optimize cut point temperatures and specific energy, deal with feed uncertainty	- ANN predicts products' cut points based on the feed composition (reduces calculation time) - 3 hidden layers (13, 17, 16 neurons) - Aspen HYSYS used for energy demands calculation - Taguchi and GA optimize cut point temperatures	85 % / 15 % (192 data points)	Yes	[16]
-feedforward NN (FNN) -simulated annealing (SA)	retrofit of heat exchanger network (HEN) for optimizing CDU operation	-1 hidden layer -scenarios generated in Aspen HYSYS -profit is the objective function to maximize	70 % / 15 % / 15 % (800 data points)	Yes	[17]
-bootstrap aggregated neural networks (BANN)	maximization of exergy efficiency of a CDU	-BANN contains 30 NNs -Aspen HYSYS used for operating data generation	50 % / 30 % / 20 %	Yes	[18]
-evolutionary algorithm assisted by adaptive surrogate functions (EASF) -radial basis function (RBF) -approximation error fuzzy control strategy (AEFCS)	maximize distillation heat supply from reflux streams	-Aspen Plus – rigorous model -DEA (differential EA) – optimization -RBF – surrogate model, three layers -AEFCS – model management -ba	N/A	Yes	[19]

**Table 2 Continued.** Overview of the records related to process optimization.

Methods	Purpose	Description	Training / Test / Validation	EES	Ref.
<b>2b. Cost optimization</b>					
-neural network (NN) -support vector machine (SVM) -genetic algorithm (GA)	design of a crude distillation unit with optimized cost	-Aspen HYSYS used to build a surrogate model (ANN) while feasibility constraints are generated using a SVM to optimize the column configuration -annual cost as an objective function -18 inputs -46 output (divided into multiple NNs) -one hidden layer (10 neurons)	-70 % / 15 % / 15 % (NN) -75 % / 25 % (SVM)	Yes	[20]
-wavelet neural network (WNN) -line-up competition algorithm (LCA)	economic optimization of a CDU under prescribed constraints	-WNN predicts the CDU behavior, LCA applied to the constrained optimization problem -one hidden layer, 30 neurons -Aspen Plus used for the rigorous model -WNN superior to RBFNN and BPNN, LCA superior to GA and PSO	350/150	Yes	[21]
<b>2c. Non-specific optimization</b>					
efficient dropout neural network (EDN)	optimization problem of crude oil distillation units	-advantageous for high-dimensional multi-objective expensive optimization problem -up to 100 decision variables and 20 objectives -two hidden layers -sensitivity analysis on parameters	various	No	[22]

In a sense, both implementations from the Other category (Table 3) can be understood as soft sensors, because both TBP data [10] and corrosion rate [12] are quantities that are costly to measure and not available in real time. They are excluded from the first category because of the limited relation to the process unit and the lack of a tangible location where the soft sensor would predict the quantities.

**Table 3.** Overview of the other records from the literature search.

Methods	Purpose	Description	Training / Test / Validation	EES	Ref.
<b>3a. Property Estimation</b>					
sure independence screening and sparsifying operator (SISSO)	ML used for the creation of a simple equation (TBP data to products' relative amounts)	laboratory distillation of crude oil	–	No	[23]
<b>3b. Risk management</b>					
-feedforward NN (FNN) -multiple linear regression analysis (MLRA)	corrosion prediction and the related risk management	-11 inputs -two hidden layers -3 outputs (pH, chloride and salt content in crude) -MLRA used to predict the corrosion rate	60–90 % / 40–10 % (24 points)	No	[24]

## 4.2 Discussion

Feedforward neural networks (FNN), and multilayer perceptron (MLP) which belongs to this category, were most dominant across all the applications. There was, however, a slight bias introduced to the search since NN were the only AI method that was explicitly searched for.

One common feature is the lack of data, which is often compensated using process simulation (Aspen HYSYS [14,16,17,18,20], Aspen Plus [19,21], PRO/II [13]) to generate a data set mimicking real operating data. While the motivation is sometimes the simplicity of data generation, this might indicate an underlying problem of the industry's conservatism toward artificial intelligence and limited willingness to share data with researchers.

While there were some attempts to optimize the architecture and parameters in the assessed literature [9,10], the trial-and-error approach is still most common for this challenge as well as method selection.

## 5 Conclusions

Artificial intelligence (AI) has been successfully applied to a number of engineering problems and has hugely grown in popularity in chemical engineering lately. In this work, some recent publications which applied artificial intelligence to the most energy-intensive oil refining process, crude distillation, were examined with respect to their purpose, methods, data set character and sustainability ideas.

Based on a multi-step search and selection, 16 records were included in the overview. It was found out that soft or virtual sensors are the most typical implementation of artificial intelligence for this process, while neural networks are the dominant AI method.

The variety of approaches, especially for the optimization of crude distillation units, make AI cumbersome to use for practical engineering problems. An interesting future research direction could focus on the automatic selection of an appropriate algorithm, together with the tuning of hyperparameters associated with machine learning.

## Acknowledgement

This research has been supported by the project No. CZ.02.1.01/0.0/0.0/16\_026/0008413 “Strategic partnership for environmental technologies and energy production”, which has been co-funded by the Czech Ministry of Education, Youth and Sports within the EU Operational Programme Research, Development and Education.

## References

1. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 1943 5:4. 5, 115–133 (1943).
2. Schmidhuber, J.: Deep Learning in neural networks: An overview. *Neural Networks*. 61, 85–117 (2015).
3. He, Q., Wang, L.: An effective co-evolutionary particle swarm optimization for constrained engineering design problems. *Engineering Applications of Artificial Intelligence*. 20, 89–99 (2007).
4. Leitão, P.: Agent-based distributed manufacturing control: A state-of-the-art survey. *Engineering Applications of Artificial Intelligence*. 22, 979–991 (2009).
5. Weber, P., Medina-Oliva, G., Simon, C., Iung, B.: Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*. 25, 671–682 (2012).
6. Ömer Faruk, D.: A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*. 23, 586–594 (2010).
7. Worrell, E., Bernstein, L., Roy, J., Price, L., Harnisch, J.: Industrial energy efficiency and climate change mitigation. *Energy Efficiency* 2008 2:2. 2, 109–123 (2008).
8. Abdulrahman, I., Máša, V., Teng, S.Y.: Process intensification in the oil and gas industry: A technological framework. *Chemical Engineering and Processing - Process Intensification*. 159, 108208 (2021).

9. Niño-Adan, I., Landa-Torres, I., Manjarres, D., Portillo, E.: Soft-sensor design for vacuum distillation bottom product penetration classification. *Applied Soft Computing*. 102, (2021).
10. Rogina, A., Šiško, I., Mohler, I., Ujević, Ž., Bolf, N.: Soft sensor for continuous product quality estimation (in crude distillation unit). *Chemical Engineering Research and Design*. 89, 2070–2077 (2011). <https://doi.org/10.1016/j.cherd.2011.01.003>
11. Shang, C., Yang, F., Huang, D., Lyu, W.: Data-driven soft sensor development based on deep learning technique. *Journal of Process Control*. 24, 223–233 (2014).
12. Sun, K., Huang, S.-H., Wong, D.S.-H., Jang, S.-S.: Design and Application of a Variable Selection Method for Multilayer Perceptron Neural Network With LASSO. *IEEE Transactions on Neural Networks and Learning Systems*. 28, 1386–1396 (2017).
13. Urhan, A., Ince, N.G., Bondy, R., Alakent, B.: Soft-Sensor Design for a Crude Distillation Unit Using Statistical Learning Methods. (2018)
14. Zhou, C., Liu, Q., Huang, D., Zhang, J.: Inferential estimation of kerosene dry point in refineries with varying crudes. *Journal of Process Control*. 22, 1122–1126 (2012).
15. Kumar, P., Rawlings, J.B., Wright, S.J.: Industrial, large-scale model predictive control with structured neural networks. *Computers and Chemical Engineering*. 150, (2021).
16. Durrani, M.A., Ahmad, I., Kano, M., Hasebe, S.: An artificial intelligence method for energy efficient operation of crude distillation units under uncertain feed composition. *Energies*. 11, (2018).
17. Ochoa-Estopier, L.M., Jobson, M., Smith, R.: Operational optimization of crude oil distillation systems using artificial neural networks. *Computers and Chemical Engineering*. 59, 178–185 (2013). <https://doi.org/10.1016/j.compchemeng.2013.05.030>
18. Osuolale, F.N., Zhang, J.: Multi-objective Optimisation of Atmospheric Crude Distillation System Operations Based on Bootstrap Aggregated Neural Network Models. *Computer Aided Chemical Engineering*. 37, 671–676 (2015).
19. Shi, X., Tong, C., Wang, L.: Evolutionary optimization with adaptive surrogates and its application in crude oil distillation. In: 2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016 (2017)
20. Ibrahim, D., Jobson, M., Li, J., Guillén-Gosálbez, G.: Optimization-based design of crude oil distillation units using surrogate column models and a support vector machine. *Chemical Engineering Research and Design*. 134, 212–225 (2018).
21. Shi, B., Yang, X., Yan, L.: Optimization of a crude distillation unit using a combination of wavelet neural network and line-up competition algorithm. *Chinese Journal of Chemical Engineering*. 25, 1013–1021 (2017).
22. Guo, D., Wang, X., Gao, K., Jin, Y., Ding, J., Chai, T.: Evolutionary Optimization of High-Dimensional Multiobjective and Many-Objective Expensive Problems Assisted by a Dropout Neural Network. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. (2021).
23. Giordano, G.F., Vieira, L.C.S., Gomes, A.O., de Carvalho, R.M., Kubota, L.T., Fazzio, A., Schleder, G.R., Gobbi, A.L., Lima, R.S.: Distilling small volumes of crude oil. *Fuel*. 285, (2021).
24. Hassanudin, S.N., Aziz, I.A., Jaafar, J., Qaiyum, S., Zubir, W.M.A.M.: Predictive analytic dashboard for desalter and crude distillation unit. In: 2017 IEEE Conference on Big Data and Analytics, ICBDA 2017. pp. 55–60 (2018)