Accelerated modeling and design of a mixed refrigerant cryogenic process using a data-driven approach

Alimardani, Hosein; Asgari, Mehrdad; Shivaee-Gariz, Roohangiz; Tamnanloo, Javad

Published in:
Digital Chemical Engineering

Link to article, DOI:
10.1016/j.dche.2024.100143

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Digital Chemical Engineering 10 (2024) 100143

Contents lists available at ScienceDirect

Digital Chemical Engineering

journal homepage: www.elsevier.com/locate/dche

Original article

Accelerated modeling and design of a mixed refrigerant cryogenic process using a data-driven approach

Hosein Alimardani [a,1], Mehrdad Asgari [b,*,1], Roohangiz Shivaee-Gariz [c,d], Javad Tamnanloo [e]

[a] Department of Chemical Engineering, Caspian Faculty of Engineering, College of Engineering, University of Tehran, Tehran, Iran
[b] Department of Chemical Engineering & Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK
[c] Chemical engineering department, faculty of engineering, University of Tehran, Tehran, Iran
[d] Department of Chemical and Biochemical Engineering, Technical University of Denmark, DTU-Building 229, P. Søltofts Plads, DK-2800 Lyngby, Denmark
[e] School of Polymer Science and Polymer Engineering, University of Akron, 302 E Buchtel Ave, Akron, OH 44325, United States

ARTICLE INFO

Keywords:
Sub-ambient process
Machine learning
Data-driven techniques
Process modeling
Optimization

ABSTRACT

Cryogenic processes with mixed refrigerants are prevalent in energy-intensive chemical industries, enhancing energy efficiency while reducing costs and unit size. However, the curse of dimensionality and process design constraints pose significant hurdles for effective screening and optimization. To tackle this, we developed a neural network model for natural gas liquefaction prediction. Trained on an extensive Aspen HYSYS database, our ML model accurately simulates LNG processes, with an impressive R^2 test value of 99.63, operating almost ten million times faster than HYSYS. It effectively addresses vital process design constraints, including liquid slugging and temperature cross, crucial for optimization. By integrating the ML model with genetic and Nelder–Mead algorithms, we achieve an 8.9% reduction in total exergy, outperforming Aspen HYSYS within the same time frame. Our study underscores ML's significance in modeling energy-intensive chemical processes, providing insights into the exergy profile and enabling feature importance analysis.

1. Introduction

Cryogenic liquefaction systems and refrigerators have been integral to significant chemical engineering industries for over a century, with applications in ethylene recovery and gas liquefaction (Kumar et al., 2011). However, a major challenge faced by these processes is their high energy consumption. To enhance energy efficiency, the use of mixed refrigerant (MR) systems has emerged as a key solution, offering substantial advantages over conventional single-component refrigerant cascade or single-stage liquefaction systems (Aspelund et al., 2010; Lu and Wang, 2009; Cheng and Mah, 1980). Unlike single-component refrigerants, MR systems can substantially reduce energy consumption by matching their thermodynamic properties with the process cooling demands, enabling near-reversible operation and higher thermodynamic efficiency during the refrigeration process (Khan et al., 2013; Lee et al., 2002). Adjusting the refrigerant composition in MR systems allows fine-tuning of the evaporation temperature profile, resulting in reduced heat-transfer irreversibility and consequent energy savings. Additionally, MR liquefaction systems offer advantages in terms of reduced technical complexity and ease of maintenance (Lee et al., 2002).

Optimizing the thermodynamic performance of MR cryogenic systems heavily relies on selecting the appropriate refrigerant composition. However, this task poses two main challenges. Firstly, the vast range of available refrigerant options in MR systems results in a high-dimensional and extensive design space (Qyyum et al., 2019; Sangahn, 2019). Secondly, there are no explicit algebraic functions describing the behavior of MR refrigerants during cryogenic processes due to the evolving composition and physical properties as the MR stream evaporates (Khan et al., 2013; Ebrahimi et al., 2021; Shivaee-Gariz et al., 2020). A comprehensive modeling and analysis of the cryogenic process is crucial to predict the performance of designed MR refrigerants, often implemented using commercial chemical process simulation software or in-house developed computer programs. However, such models can be computationally demanding, especially with commercial software (Reddy et al., 2019; Xu et al., 2013).

In addition to the challenges associated with optimizing the composition of mixed refrigerant (MR) systems, there are further constraints that complicate the search for optimal designs in cryogenic chemical processes (Veloso et al., 2018; Leipold et al., 2008; Deb and Datta, 2013). Certain process design schemes involving temperature cross

* Corresponding author.
E-mail address: ma2000@cam.ac.uk (M. Asgari).
[1] The authors have contributed equally to this work.

https://doi.org/10.1016/j.dche.2024.100143
Received 17 October 2023; Received in revised form 7 January 2024; Accepted 26 January 2024
Available online 30 January 2024

in heat exchangers (Sanavandi and Ziabasharhagh, 2016) and liquid slugging in compressors (Xu et al., 2013) are deemed infeasible for many cryogenic chemical engineering processes due to technical limitations or operational considerations within chemical plants. This poses a significant challenge in distinguishing valid and invalid points within the complex and expansive high-dimensional design space of cryogenic processes. Though not straightforward, overcoming this challenge is crucial for successful process optimization and design (Mokhatab et al., 2013; Eldemerdash, 2011).

Furthermore, there are preferences expressed as restraints, both hard and soft, which further add complexity to the optimization of cryogenic processes. For instance, achieving a higher temperature difference across the heat exchanger is often considered a constraint in process design (Arie et al., 2015; Gundersen and Grossmann, 1990; Jimenez et al., 2004). The capital cost (CAPEX) and operating cost (OPEX) of cryogenic processes are influenced by the temperature profiles of the evaporation and condensation processes. Typically, the temperature approach within the heat exchangers of MR systems falls within the range of approximately 1–3 °C (Lee et al., 2002). Incorporating these restraints into the optimization and design strategies is challenging, as even slight variations in their definitions can lead to changes in the optimized solution (Deb and Datta, 2013; Curtis and Nocedal, 2008; Homaifar et al., 1994).

To address these challenges, various strategies, such as Genetic Algorithm (GA) and modified GA methods, have been employed to optimize cryogenic chemical processes using MR refrigerant systems (Nogal et al., 2008; Taleshbahrami and Saffari, 2010; Morin et al., 2011; He and Ju, 2014). GA, inspired by Darwin's theory of evolution and Mendelian inheritance, is an effective optimization method in multidimensional spaces where system behavior is not easily predictable using gradient-based approaches (Cammarata et al., 2001; Gen and Cheng, 1999). By balancing current knowledge and random searches, GA allows for the discovery of multiple solutions in the design space (Nogal et al., 2008). Researchers have successfully utilized GA for designing and optimizing MR cycles due to these advantageous characteristics (Nogal et al., 2008; He and Ju, 2014). For example, Alabdulkarem et al. applied GA to optimize the compression work in a propane precooled MR (C3MR) process, resulting in a 9% reduction in energy demand (Alabdulkarem et al., 2011).

In the pursuit of optimizing complex chemical engineering problems, Genetic Algorithm (GA) has shown promise (Angelova and Pencheva, 2011; Mosayebi and Sodhi, 2020). However, to prevent convergence to local optima (Rocha and Neves, 1999), especially in large search spaces with high dimensionality (McNelis, 2005), parameter tuning becomes crucial. To address this challenge, we previously employed a modified version of GA, called Discrete-Continuous Genetic Algorithm (DCGA), to optimize an MR cryogenic process (Ebrahimi et al., 2021). Our modification focused on enhancing randomization to avoid local minima and explore a larger design space, resulting in a 12.5% reduction in specific energy consumption compared to baseline GA optimization. Nonetheless, computational demands restrict the number of data points obtainable, particularly in MR cryogenic process modeling using commercial software.

Machine learning is rapidly outpacing traditional, rigorous methods in various chemical engineering fields e.g. fault detection, process control, and material discovery, resulting in substantial improvements in efficiency and productivity across various industrial settings (Schweidtmann et al., 2021; Pyzer-Knapp et al., 2022). While GA and conventional optimization algorithms have limitations when applied to chemical engineering process models (Keyvanloo et al., 2012), machine learning models offer a promising solution by addressing challenges related to model complexity, process understanding, and constraint handling (Karniadakis et al., 2021). Machine learning techniques provide insights into the significance of individual parameters (Rodríguez-Pérez and Bajorath, 2019; Wang et al., 2022) and their impact on thermodynamic performance across the entire design space, enabling intelligent design or retrofitting of industrial chemical plants (Khan and Lapkin, 2020). Additionally, machine learning facilitates recognizing and predicting valid operating conditions based on predefined design criteria, such as temperature cross and liquid slugging, supporting the development of modified design rules and targeted optimization, which conventional methods often struggle to achieve.

Recently, Kim et al. (2023) demonstrated the effectiveness of reinforcement learning (RL) in optimizing mixed refrigerant cryogenic processes, achieving a remarkable 13%–15% improvement in energy efficiency. This highlights the potential of artificial intelligence, specifically RL, in optimizing chemical engineering processes. However, certain aspects require further investigation regarding the utilization of machine learning (ML) for simulating and optimizing mixed-refrigerant liquefaction processes.

One crucial area that needs attention is the development of a fast and efficient simulation toolbox for mixed-refrigerant liquefaction processes using ML techniques. Current commercial software is often expensive, slow, and limited in accessibility. To overcome these limitations, supervised machine learning techniques can be explored to develop a versatile toolbox adaptable to various process design constraints and rules (Zhu et al., 2020; Ali et al., 2018).

Additionally, feature importance analysis has been overlooked in the RL-based approach studied recently (Kim et al., 2023). Combining supervised machine learning techniques with feature importance analysis can provide valuable insights into factors significantly influencing process performance, aiding in the optimization process.

Moreover, the challenges associated with identifying process design scenarios that satisfy constraints and their impact on finding valid solutions and optimizing energy consumption have not been adequately addressed. The role of data-driven techniques in identifying optimal process designs remains underexplored. This study aims to fill these research gaps and provide a comprehensive understanding of how data-driven techniques can effectively contribute to discovering optimal process designs in mixed-refrigerant cryogenic processes.

In our previous study (Shivaee-Gariz et al., 2020), we employed a basic Genetic Algorithm (GA) technique to model and optimize an LNG process utilizing a mixed-refrigerant cryogenic system. However, due to GA's inclination towards local minima instead of global optima, we now adopt a data-driven approach to model, understand, and optimize the targeted cryogenic process.

To enhance our process model in Aspen™, we expanded the range of considered refrigerants, enabling a more comprehensive understanding and improved thermodynamic efficiency. By systematically varying the process input variables, including mixed refrigerant composition, temperature, and pressure parameters, we constructed a substantial database linking these input variables to the resulting thermodynamic output parameters.

Using this extensive database, we developed and trained a neural network model to simulate the cryogenic process with mixed refrigerants. We evaluated the speed and accuracy of this model, assessing its potential as a viable alternative to commercial process simulation software. Additionally, we conducted a feature importance analysis using the model to identify the key factors contributing to a valid and efficient process design. The predicted exergy loss and adherence to process design criteria emerged as crucial indicators among the output variables.

Building upon the developed model, we fine-tuned the parameters associated with GA and optimized the natural gas liquefaction process utilizing the mixed-refrigerant system. To highlight the significance of data-driven techniques in designing and optimizing energy-intensive chemical processes, we compared the achieved thermodynamic efficiency in this study with conventional optimization methods. Our results underscore the importance of data-driven approaches in facilitating the understanding and design of efficient mixed-refrigerant cryogenic processes.

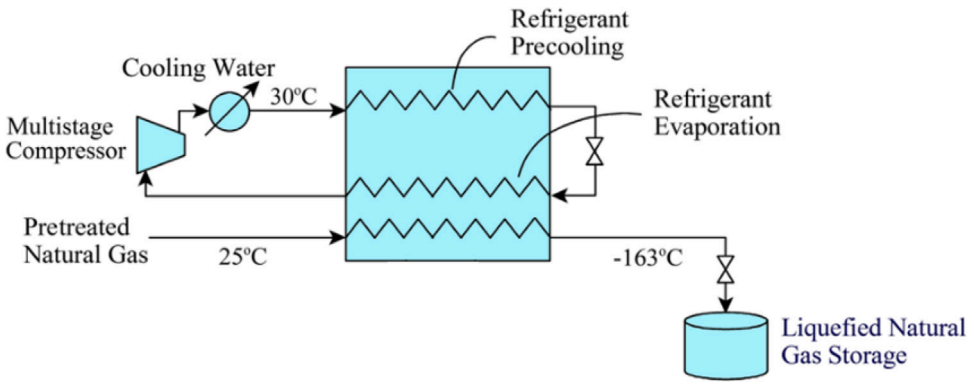

Fig. 1. Simple liquefied natural gas (LNG) process. *Source:* Reproduced with permission from Lee G-C, Smith R and Zhu XX (2003) Optimal Synthesis of Mixed-refrigerant Systems for Low-temperature Processes, Ind Eng. Chem Res, 41: 5016. © 2008 American Chemical Society.

2. Methods

2.1. Modeling of mixed refrigerant liquefaction process

Using mixed refrigerants instead of pure ones can significantly expand the operating range of refrigeration, especially when dealing with considerable temperature changes in the refrigeration cycle. The mixed refrigerant's evaporation is expected to occur within a similar temperature range as the process cooling range, resulting in a reduced temperature difference between the refrigerant and process streams and, consequently, decreased energy demands for refrigeration (Lee et al., 2002).

The design of a mixed refrigerant cycle is influenced by several parameters, including pressure level, refrigerant flow rate, and refrigerant composition. These parameters directly impact the temperature profile during refrigerant evaporation and, consequently, the temperature difference between the refrigerant and process streams. This temperature difference plays a critical role in both the refrigeration performance and the overall feasibility of the refrigeration process. If the temperature difference is negative, the refrigeration process becomes infeasible. Adjusting the compressor discharge pressure can lead to temperature cross, while increasing pressure results in a higher temperature difference, necessitating more power (Lee et al., 2002). Similarly, altering the refrigerant flow rate has a comparable impact on the refrigeration process.

Refrigerant composition plays a complex role in determining the refrigeration performance. Modifying the composition of the mixed refrigerant by adding or removing components can alter the entire evaporation profile, making the selection of an appropriate composition critical. The adjustment of refrigerant composition, flow rate, and pressure level is challenging due to the sensitivity of the temperature difference to refrigerant parameters and the intricate interactions among them.

Mixed refrigerants find significant use in the liquefaction of natural gas (LNG). The typical LNG process involves introducing natural gas into a heat exchanger, where it undergoes liquefaction through counter-current refrigerant flow. The refrigerant is then compressed and partially condensed using cooling water. The refrigerant's total condensation occurs through the heat exchanger by the evaporating refrigerant. Finally, the refrigerant is expanded through a valve and returned to the heat exchanger to continue the refrigeration cycle. Fig. 1 depicts the fundamental scheme of a simple LNG process.

2.2. Data harvesting of the chemical process

Sufficient data is crucial for the successful creation of machine learning models in chemical engineering applications. However, acquiring real-world data from an operational chemical plant can be time-consuming, expensive, and potentially hazardous. Additionally, the practical range of operating conditions that can be explored in a real plant is often limited by equipment constraints, leading to a lack of diversity in the data and poor generalization of the process.

To overcome these challenges, we utilized simulation software (specifically Aspen HYSYS in this study) to generate data for our machine learning model. Simulation software provides a fast, cost-effective, and safe way to collect data across a wide range of operating conditions, enabling us to explore a more comprehensive parameter space. Moreover, simulation software allows for in-depth analysis and examination of how the process behaves in previously unexplored regions.

In this study, we focused on simulating the PRICO (Poly Refrigerant Integral Cycle Operation) process, a proprietary mixed refrigerant process used for gas liquefaction. Our simulation involved a single mixed refrigerant loop with a mixture of ten components. Table 1 provides details of the feed conditions and additional assumptions used during data harvesting through simulation (Aspelund et al., 2010).

Through the use of Python's public library, pywin32, we were able to automatically access Aspen HYSYS Objects and efficiently harvest over 1 million data points. This abundant dataset provides a solid foundation for training our machine learning model, allowing us to develop accurate and reliable predictions for the PRICO process.

Once trained, the model becomes self-sufficient, eliminating the need for repeated data collection and streamlining the process. This approach contrasts with conventional methods that require repeated simulations for each prediction. Additionally, the model's adaptability to varying process constraints extends its versatility to diverse design scenarios, making it an invaluable tool for optimization tasks.

2.3. Features and labels

The input dataset for our study consisted of 13 features, which encompassed various aspects of the refrigerant's chemical composition. These features included the molar composition percentages of methane, ethane, propane, butane, nitrogen, ammonia, chlorine, ethylene, propene, and isobutane in the refrigerant. Additionally, the dataset contained information on the inlet and outlet pressure of the compressor, as well as the molar flow rate of the refrigerant. The operational range for molar flows spanned from 1 to 3.2 kg mole/s, the

Table 1
Natural gas (NG) parameters and operational assumptions.

Parameter	Value
NG molar flow (kg mole/s)	0.3
NG Composition (mol %)	100
Nitrogen	0.37
Methane	95.89
Ethane	2.96
Propane	0.72
n-butane	0.06
NG inlet pressure (bar)	60
NG inlet temperature (°C)	20
MR temperature after cooler (°C)	20

inlet pressure of MR to compressor ranged from 100 to 1000 kPa, and the outlet pressure of the compressor ranged from 2000 to 10,000 kPa.

In the study, generating compositions for the mixed refrigerant stream posed a significant challenge due to the interdependent nature of the coolant components. To address this, we employed the Dirichlet distribution, a statistical distribution that effectively captures the inherent dependencies among the components. Therefore we achieved an even distribution of values for each, ensuring a comprehensive exploration of the input space without the need for predetermined levels. This approach was particularly beneficial given the coarse nature of the exergy loss outputs, as it prevented a loss of resolution that could hinder the optimization phase. On the other hand, for the remaining three input variables – inlet and outlet pressure of the MR in the compressor and MR molar flow – a random sampling approach was employed within their respective operational ranges. Setting fixed levels for these variables would introduce artificial granularity and compromise resolution, whereas random sampling allowed for a continuous representation of values within the operational ranges. The resulting dataset achieved a uniform distribution across the entire input space, providing a robust foundation for the subsequent optimization phase without the limitations of limited resolution or fixed levels.

On the other hand, the output data contained crucial performance metrics. These included the exergy loss of the compressor, aftercooler, valve, and LNG heat exchanger. Furthermore, the dataset also incorporated the minimum approach temperature and the vapor fraction of the compressor inlet. Notably, the last two parameters could be considered as constraints during the optimization process. To be more precise about constraint handling, the numerical values of these parameters are predicted by the proposed regression model and the model does not directly determine the feasibility of the case. Therefore a post-processing step is needed to check the validity of a case depending on the specific criteria defined in the optimization scenario e.g the minimum approach temperature threshold.

Please refer to SI section 1 for further details on how we generated the features and normalized the dataset to ensure consistent and meaningful analysis.

2.4. Proposed model

In this study, we opted to employ a Deep Neural Network (DNN) as a surrogate model for our process. The decision to use a DNN was based on several advantageous characteristics it offers. Firstly, DNNs exhibit a relatively low computational load, making them computationally efficient. Moreover, they boast fast prediction speeds, allowing for rapid analysis and exploration of the data. Additionally, DNNs have the ability to effectively model complex and non-linear systems, making them suitable for our application. Furthermore, DNNs provide an adequate level of interpretability, allowing us to gain insights into the underlying processes.

The architecture of our DNN model comprises several layers. It is a dense fully connected feed-forward neural network, consisting of one input layer with 13 neurons, one output layer with 6 neurons, and six hidden layers. Each hidden layer contains 256 neurons. The DNN architecture is visually represented in Fig. 2.

The dataset was normalized using z-score normalization to eliminate the influence of different units or ranges of inputs. The activation function utilized in hidden layers was Mish due to its superior performance discussed in SI. The loss function was MAE to measure the model's performance during training. The backpropagation algorithm was employed to train the DNN using AMSGRAD as the optimizer that guides the model towards making more accurate predictions by minimizing the loss function iteratively. The compilation and training of the model were done using the Keras interface for the TensorFlow framework in Python (Tran and Phong, 2019; Misra, 2019). The training of the proposed DNN took approximately 85 min.

For a detailed understanding of the properties, architecture, and training of the DNN model, please refer to SI section 2.

2.5. MR cycle samples

The PRICO process was simulated using Aspen HYSYS, and exergy losses of unit operations were calculated using a novel graphical tool. The stream data obtained from Aspen HYSYS was incorporated into a Python script for exergy calculations. The script employed graphical tools, Ω-H diagrams, developed in (our paper reference) to determine the exergy loss within each unit operation. In these diagrams, Ω indicates the energy level and H states the enthalpy. The exergy loss in a unit operation can be calculated using the enclosed area in Ω-H diagram which is the difference between the exergy delivered by the source and the exergy accepted by the sink (Shivaee-Gariz et al., 2020). To build the dataset for our machine learning model, 60% of the exergy loss results were designated as the training set, 20% as the validation set, and the remaining 20% as the test set.

The independent variables, also referred to as features, represent the degrees of freedom for optimizing the mixed refrigerant (MR) system. These input variables were randomly varied within their specified lower and upper bounds. For each simulation that converged successfully, the corresponding values of the dependent variables, or outputs of the Deep Neural Network (DNN) model, were determined and recorded.

In this case, the system's inputs included the pressure of the MR input stream to the LNG unit, the MR output stream from the LNG unit, the molar flow of the MR stream, and the molar composition of the MR stream comprising ten components: methane, ethane, propane, n-butane, nitrogen, ammonia, Cl2, ethylene, propene, and i-butane.

The DNN model's outputs represented various variables describing the energy consumption of the MR cycle, such as exergy losses of the valve, compressor, LNG unit, and aftercooler, along with the minimum temperature approach and vapor fraction of the MR stream exiting the LNG unit. Exergy analysis provides valuable information about the location of process inefficiencies, enabling the identification of promising modifications in low-temperature refrigeration systems.

During the evaluation of actual physical validity for the cycles, two constraints were imposed. The first constraint required a minimum temperature approach of higher than 1.2 °C, and the second constraint necessitated complete condensation of the MR stream in the LNG unit. Consequently, the sum of exergy losses of all units for the valid cycles served as the objective function for the optimization task. Ultimately, the dataset consisted of evaluations for over 1 million cycles consisting of 18.5% valid cases.

2.6. Genetic algorithm

Genetic algorithms are a type of evolutionary algorithm that imitates natural selection to solve optimization problems. The process involves creating a population of potential solutions that evolve over time based on their fitness for the objective function (Mitchell, 1996). The main concept is to use rules to generate diverse solutions, evaluate

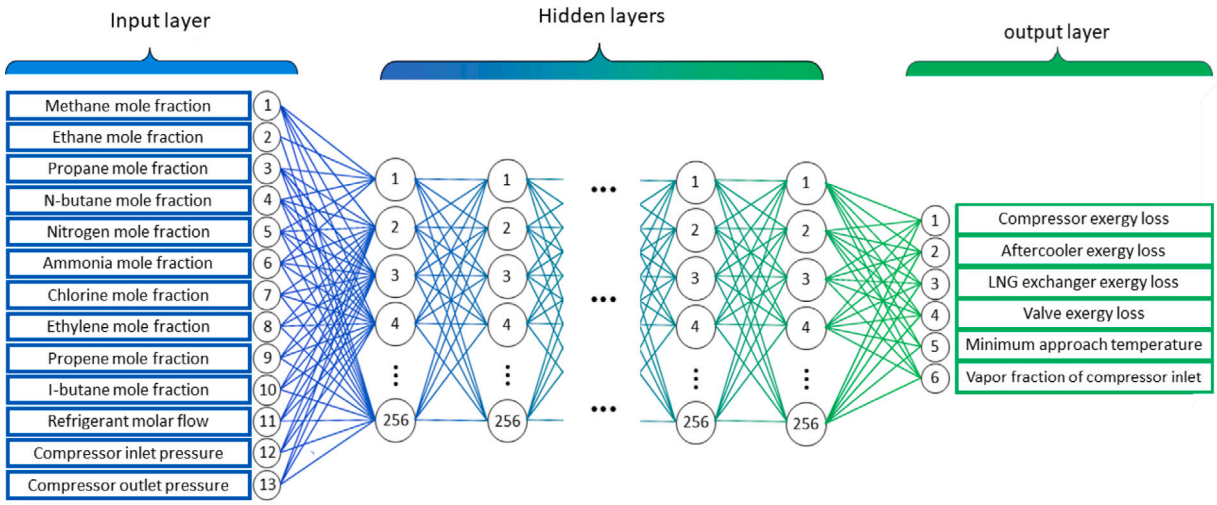

Fig. 2. The architecture of the DNN.

their quality, and combine the best ones to create even better solutions. This process continues iteratively until an optimal solution meeting the specified criteria is found (Kramer, 2017).

After training and constructing the model, it serves as the fitness function for the genetic algorithm. Each solution is evaluated using the model to predict its fitness, and the genetic algorithm selects the best solutions based on these predictions. In our case, the goal is to find the least total energy loss subject to two operational constraints: (1) ensuring a 100% vapor fraction at the compressor inlet and (2) maintaining a minimum approach temperature higher than 1.2 °C.

To implement this method, the genetic algorithm is used with the model, and it significantly speeds up computations compared to connecting Python or MatLab to a mathematical engine using the Component Object Model (COM). The computations are 6–7 orders of magnitude faster with nearly negligible error.

2.7. Nelder–Mead

Genetic algorithms are commonly employed in the early stages of optimization due to their ability to explore a large search space and identify promising regions where the optimal solution may be located. However, as the search space narrows down and the problem becomes high-dimensional, the computational cost of using genetic algorithms can become prohibitive. Additionally, the stochastic nature of genetic algorithms and the rugged landscape of variables may cause them to get stuck in local minima during the final stages of optimization. Therefore, a deeper local search is necessary to fine-tune the solution.

To address this issue, we fixed the composition of the mixed refrigerant (MR) at values determined by the genetic algorithm initially. We then focused on optimizing only the operational variables, specifically the molar flow of refrigerant and the inlet and outlet pressure of the compressor, in the final local search. For this local search, we deployed the Nelder–Mead algorithm, which iteratively refines the current solution without requiring gradient information (Aspelund et al., 2010). Nelder–Mead uses a simplex, a geometric shape formed by a set of points, to effectively explore and exploit the promising regions found by the global search performed by the genetic algorithm (Nelder and Mead, 1965). The objective function remains the same as used in the genetic algorithm, ensuring consistency in the optimization process.

3. Results and discussion

3.1. Developing an ML model for simulating the mixed-refrigerant cryogenic liquefaction process

After developing an optimized machine learning (ML) model, we assess its performance in predicting exergy losses in a liquefied natural gas (LNG) plant. The ML model's predictions are compared to the simulated values obtained from Aspen HYSYS for different components of the plant. Fig. 3(a)–(d) present the predicted versus calculated values for the compressor exergy loss, LNG exergy loss, valve exergy loss, and aftercooler exergy loss. The main part of Fig. 3(e) illustrates the total exergy loss. To quantify the model's accuracy, we calculate the mean absolute error (MAE) and the coefficient of determination (R^2) for the total exergy loss, resulting in values of X and Y, respectively. These findings demonstrate the ML model's potential for accurately predicting exergy losses in an LNG plant. Fine-tuning and optimizing the model parameters further can enhance its performance, which is crucial for improving energy efficiency, minimizing exergy loss, and reducing the environmental impact of LNG plants.

Fig. 4 compares the convergence speed of the ML model, specifically a deep neural network (DNN) model, with Aspen HYSYS simulation for the mixed refrigerant cryogenic liquefaction process. The figure depicts the inverse logarithm of the simulation duration time. By utilizing a logarithmic scale, the figure effectively conveys the orders of magnitude difference in speed between the ML model and the conventional simulation approach. Notably, the DNN model exhibits a convergence speed approximately a million times faster than Aspen HYSYS. This advantage of having a rapid and reliable model plays a vital role in enabling efficient screening, optimizing design, and sensitivity analyses of the process. With this fast model, a vast number of design scenarios can be explored within a short timeframe, offering valuable insights into process design and optimization. Despite all the glaring aspects of utilizing a machine learning model, it is advisable to corroborate the final solution obtained through modeling and screening using Aspen HYSYS to ensure the process's safety and efficiency.

3.2. Studying the mixed refrigerant cryogenic liquefaction process: what affects the efficiency of the process?

In this section, we introduce a data-driven approach aimed at understanding the fundamental principles and criteria for achieving efficient mixed refrigerant cryogenic liquefaction processes. Our focus is on studying the relationship between process design and performance to identify crucial design parameters that have a significant impact on process exergy efficiency. Leveraging the available process data, we propose a methodology to optimize these parameters and enhance overall efficiency based on process exergy loss. Furthermore, we emphasize the detection of valid process design points, where temperature cross and vapor condensation are avoided, as it is of utmost importance.

To demonstrate the significance of our approach, Fig. 6 presents a box plot illustrating the relationship between the total exergy and

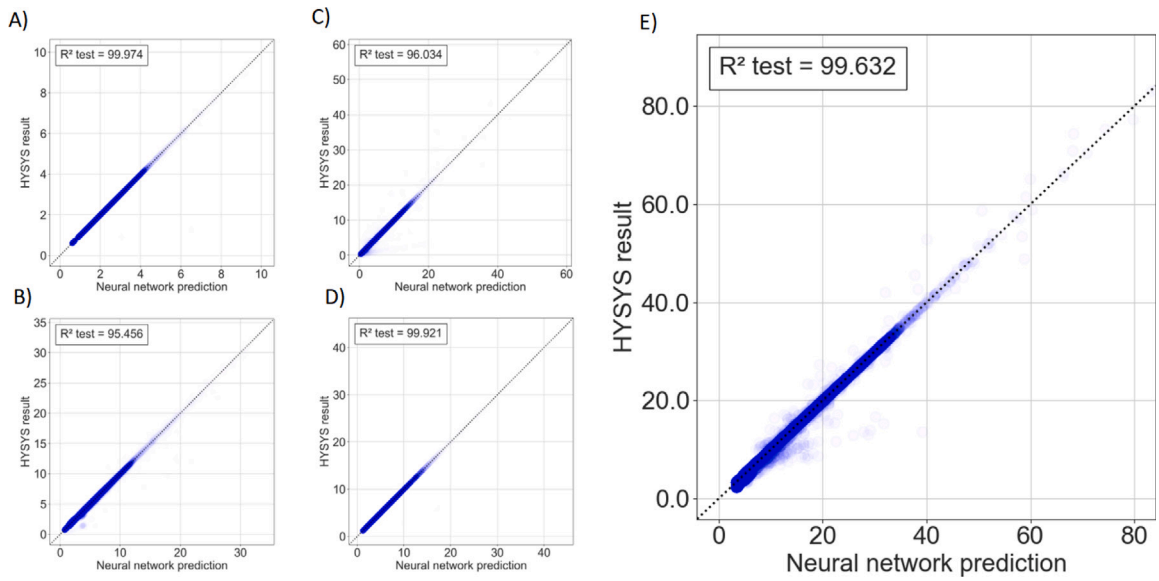

Fig. 3. Representative parity plots for the test sets of exergy loss components: (A) Compressor exergy loss, (B) LNG exergy loss, (C) Valve exergy loss, (D) Aftercooler exergy loss, and (E) Total exergy loss.

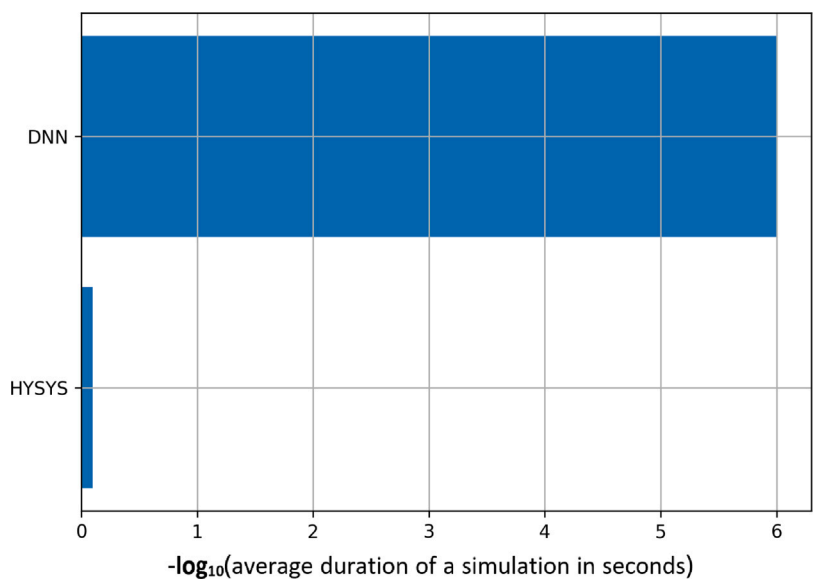

Fig. 4. Comparing the convergence speed for the developed DNN ML model and the Aspen HYSYS simulation engine.

the molar flow of the refrigerant in the mixed refrigerant cryogenic liquefaction process. As expected, the total exergy loss generally increases with the flow rate. However, the high standard deviation highlights the substantial influence of process design settings and variables on efficiency. By carefully selecting appropriate design settings and variables, it becomes possible to mitigate exergy loss at higher flow rates, surpassing the efficiency achieved at lower flow rates. Making informed decisions regarding process design and optimization is crucial to attaining optimal performance and efficiency in the context of mixed refrigerant cryogenic liquefaction processes.

Fig. 7 presents the impact of various features on the total exergy loss and its components, such as losses for the LNG heat exchanger, compressor, aftercooler, and valve. According to the figure, the nitrogen molar fraction, refrigerant molar flow, compressor inlet pressure, and methane molar fraction are the most influential factors for total exergy loss. The refrigerant molar flow and compressor inlet pressure are the most significant factors in determining the loss of the LNG heat exchanger. The nitrogen molar fraction, refrigerant molar flow, and compressor outlet pressure are the most critical factors contributing to the compressor exergy loss. The nitrogen molar fraction, methane molar fraction, and compressor molar fraction are the most critical factors for the aftercooler exergy loss. The nitrogen molar fraction is the most influential factor in determining the valve exergy loss, with a significant difference from the other contributing factors. In this study, the effect of increasing the compressor inlet pressure on the total exergy loss in a mixed refrigerant cryogenic liquefaction process is investigated. The kernel density plot presented in Fig. 8 provides valuable insights into this relationship. The observed shift in the probability distribution towards lower exergy losses indicates an improved average performance as the compressor inlet pressure is raised. This phenomenon can be attributed to the inherent characteristics of the process and the thermodynamic principles governing it.

When the compressor inlet pressure is increased, higher pressure ratios across the compressor stages are achieved, leading to more efficient compression of the refrigerant. This efficient compression process reduces exergy losses, resulting in a shift towards lower exergy

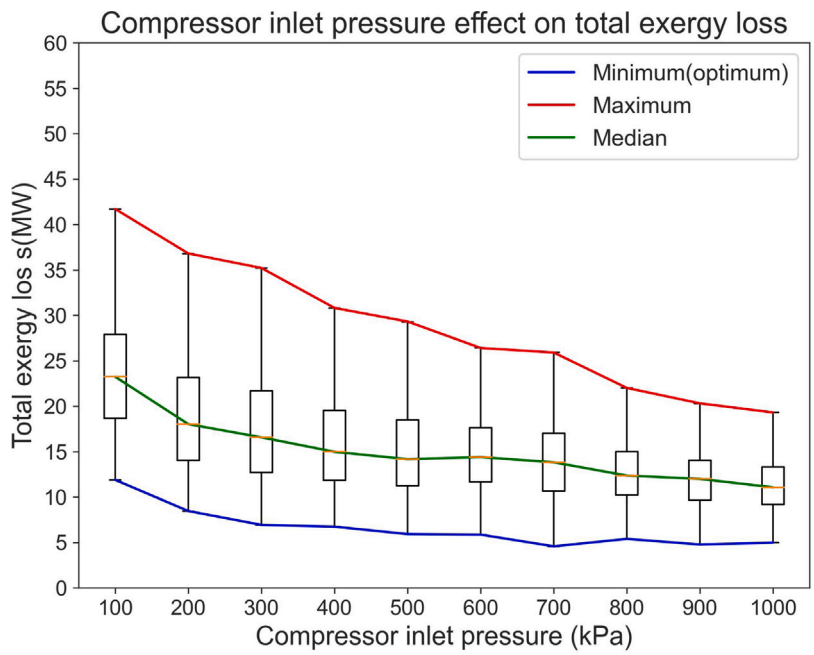

Fig. 5. Box plot of total exergy loss based on the compressor inlet pressure.

losses in the probability density plot. The narrower range of exergy losses observed with higher inlet pressure indicates a higher level of constraint in the system, suggesting that the compressor operates under more regulated conditions. This relationship highlights the importance of optimizing the compressor operation and considering other process variables to achieve improved efficiency and minimize exergy losses in the mixed refrigerant cryogenic liquefaction process.

The findings from Fig. 5, illustrating the box plot of total exergy loss based on the compressor inlet pressure, align with the results presented in Fig. 8, which depicted a kernel probability density plot of total exergy with respect to the compressor inlet pressure. These findings reaffirm the relationship between the compressor inlet pressure and the efficiency of the liquefaction process. Notably, as the compressor inlet pressure increases, both the average value and the range of variations in total exergy loss decrease, indicating a more energy-efficient liquefaction process. These results emphasize the significance of carefully tuning and optimizing process variables to achieve enhanced efficiency and reduced exergy losses in mixed refrigerant cryogenic liquefaction processes. Fig. 9 presents the random forest feature importance plot, which measures the mean decrease in impurity and indicates how much each feature contributes to the overall performance of the model. This analysis provides valuable insights into the importance of different parameters in determining the total exergy loss of a mixed-refrigerant cryogenic process.

The results reveal that the nitrogen molar fraction, compressor inlet pressure, methane molar fraction, and butane molar fraction are the most influential factors affecting the exergy loss, while the molar fractions of propane, chlorine, propene, and ammonia have less impact.

The nitrogen molar fraction plays a significant role as it affects the overall composition of the refrigerant mixture. Nitrogen is commonly used to optimize the thermodynamic behavior of the mixture, and deviations from the desired nitrogen molar fraction can disrupt the phase behavior, leading to inefficiencies in cooling and condensation processes, ultimately impacting the total exergy loss.

Another critical parameter is the compressor inlet pressure, which plays a crucial role in the compression process. Increasing the inlet pressure enhances compression work, resulting in higher refrigerant pressures and temperatures. This improved compression work leads to better cooling and condensation downstream, ultimately reducing the total exergy loss. Thus, optimizing the compressor inlet pressure is essential for maximizing process efficiency.

Additionally, the methane and butane molar fractions significantly influence the properties and behavior of the refrigerant mixture. Methane, being the primary component in cryogenic processes, affects the temperature profile of refrigerant evaporation and heat transfer characteristics, thereby influencing the exergy loss. Similarly, the butane molar fraction impacts the composition and performance of the mixture, affecting cooling and condensation processes and, consequently, the overall exergy loss. These findings highlight the importance of carefully considering and optimizing these parameters to enhance the efficiency and performance of mixed-refrigerant cryogenic processes.

In contrast, parameters such as propane, chlorine, propene, and ammonia molar fractions have a lower importance, indicating their limited influence on the exergy loss. Within the specific context of the cryogenic process under study, these components have minimal effects on phase behavior, heat transfer, or thermodynamic properties.

Fig. 10 displays a reduced-dimension t-SNE plot of variables, aiming to visualize the total exergy loss of mixed-refrigerant cryogenic processes. The plot categorizes the design scenarios into different points represented by colors. Points in black and red indicate scenarios with minimum exergy loss, signifying high efficiency. On the other hand, the blue and yellow points represent design scenarios with less efficiency, characterized by higher energy loss. The observed characteristics of the points in the plot, such as their scarcity, randomness, and dispersion, shed light on the coarse nature of the problem. The scarcity of points suggests that finding design scenarios with optimal efficiency is challenging. Only a small subset of all possible design scenarios leads to highly efficient processes.

Additionally, the randomness and dispersion of the points illustrate the difficulty of exploring the design space and identifying efficient configurations. The scattered distribution of the points signifies that the relationship between the variables and the resulting exergy loss is complex and non-linear. There is no simple, direct mapping between the input variables and the desired outcome of minimizing energy loss. The randomness in the distribution of points suggests that small changes in the input variables can lead to significant variations in the exergy loss, making it difficult to predict the optimal design configurations.

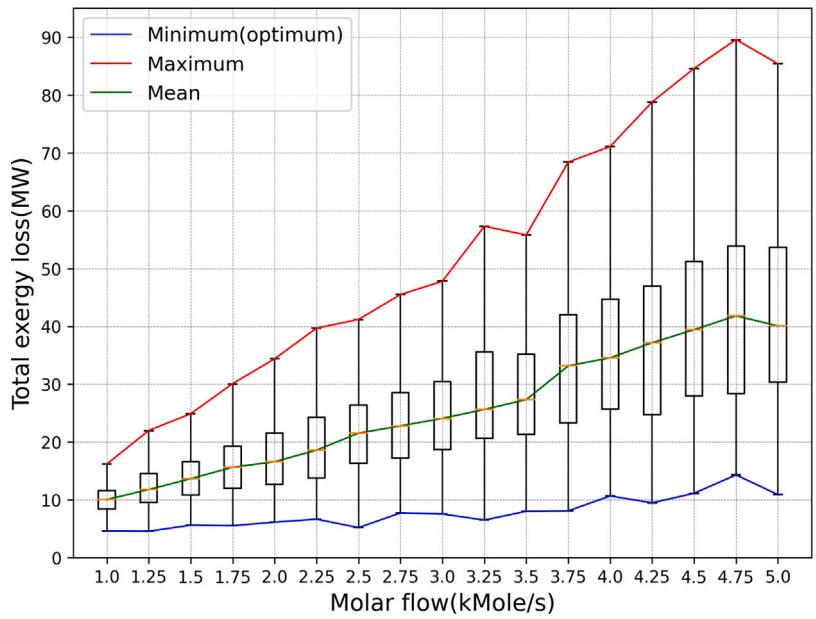

Fig. 6. Total exergy loss as a function of the refrigerant molar flow.

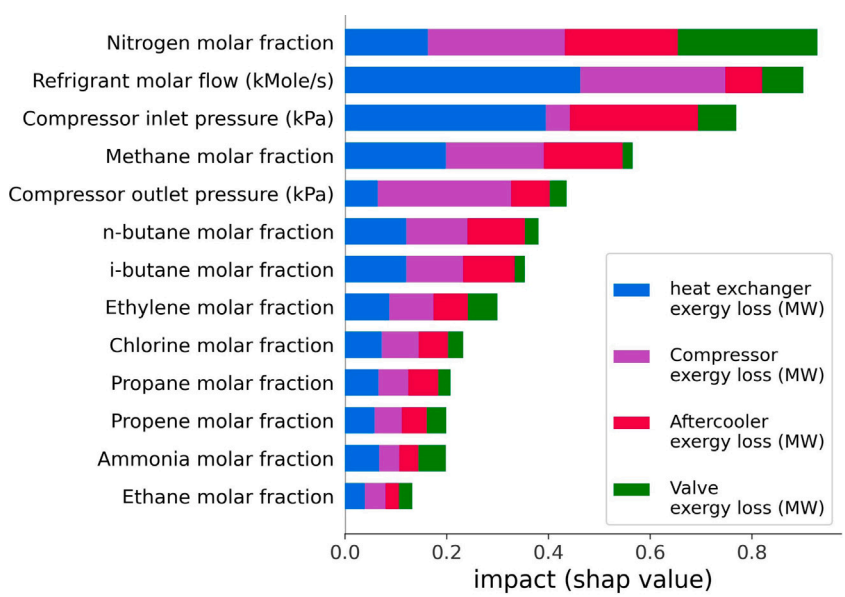

Fig. 7. Calculated SHAP values showing the feature importance for different features.

Furthermore, the coarseness of the problem is highlighted by the high-dimensional space in which the search for efficient design scenarios occurs. With multiple variables at play, the design space becomes vast and intricate. Exploring all possible combinations and their respective exergy losses becomes computationally challenging and time-consuming. The high dimensionality adds complexity and makes it challenging to navigate the design space effectively, further exacerbating the difficulty of finding efficient solutions. These challenges emphasize the need for advanced optimization techniques and data-driven approaches to tackle the intricacies of mixed-refrigerant cryogenic processes and achieve optimal efficiency.

3.3. Developing an ML model for predicting the validity of design cases

The identification of valid process design points plays a critical role in cryogenic liquefaction processes. Valid points are defined as operating conditions where temperature cross and vapor condensation are avoided, representing feasible optimal conditions for the process. The detection of these points is not only very important for detecting feasible design choices but is also essential for achieving maximum efficiency, as discussed in later sections. This detection is particularly crucial during the design phase, where the optimization of various parameters is required to attain the desired performance.

Fast and reliable tools for detecting valid points provide a means to quickly screen the design space, facilitating the identification of optimal design parameters and innovative processes. These tools enable the generation of large datasets, which can be utilized to train machine learning models for process optimization. This approach allows for the accurate identification and optimization of design parameters, leading to a search for cost-effective and efficient cryogenic liquefaction processes among the feasible design choices. The detection of valid points also supports sensitivity analyses and exploration of the design space, offering valuable insights into process performance and behavior. Ultimately, the detection of valid points ensures successful design

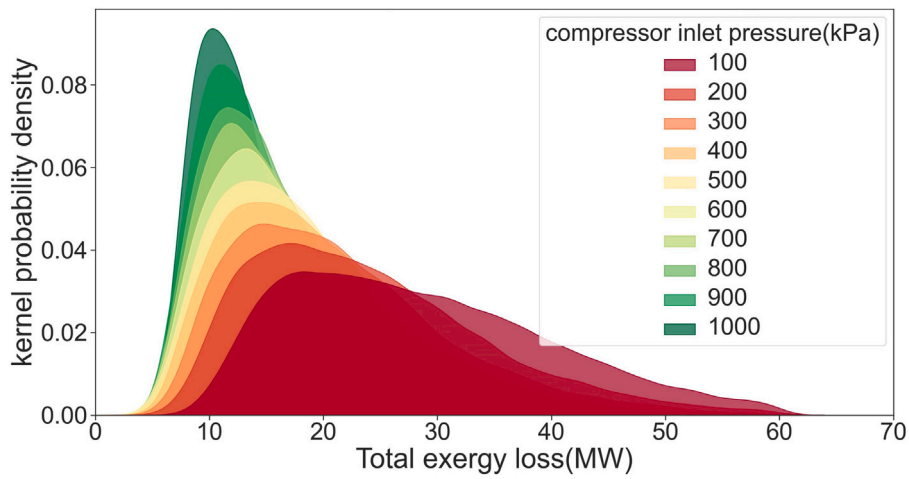

Fig. 8. The kernel probability density showing the effect of compressor inlet pressure on total exergy loss.

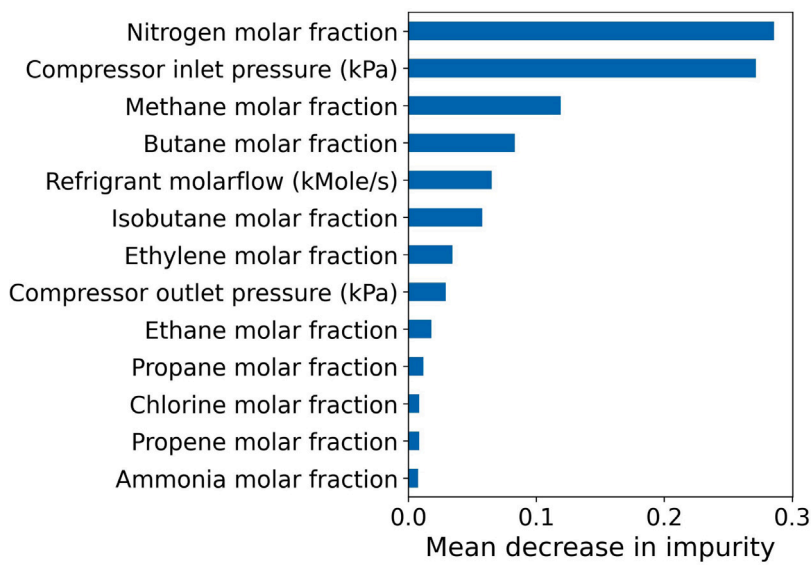

Fig. 9. The feature importance factors calculated by random forest technique depict the importance of different process variables on total exergy loss.

and operation, impacting the economic viability and competitiveness of cryogenic liquefaction processes.

Fig. 11 shows the impact of various features on the overall validity and invalidity of the designed process, as well as the two contributing factors to validity and invalidity, i.e., liquid slugging at the compressor inlet and temperature cross in the LNG heat exchanger. Fig. 11 demonstrates that the nitrogen molar fraction and methane molar fraction are the most important factors in determining the validity and invalidity of the process design. A higher concentration of nitrogen in the feed stream may lead to liquid slugging, which can cause significant damage to the compressor. On the other hand, a lower concentration of methane in the feed stream can result in a temperature cross, which can adversely affect the feasibility of the process. Overall, understanding the impact of various features on the process design can help identify the critical design parameters that significantly impact the feasibility of the process, and this information can be utilized to design the process in a valid manner.

In addition to our focus on understanding the principles and criteria for efficient mixed refrigerant cryogenic liquefaction processes, we have also developed a robust machine learning (ML) model capable of predicting valid process design scenarios from invalid data points. The effectiveness of this ML model has been validated through a confusion matrix, as shown in Fig. 12.

The test set comprises a total of 209715 data points, and the ML model successfully predicts over 99.5% of the design scenarios (37,047 True/True and 171,610 False/False) as valid or invalid. This finding is highly significant as it enables the efficient exploration of the valid portion within the design space. Without such a model, devising a strategy to identify optimum valid design cases becomes a challenging task.

The successful prediction of valid process design scenarios by our ML model opens up new possibilities for optimizing cryogenic liquefaction processes. By leveraging this model, researchers and engineers can focus their efforts on the valid design space, significantly reducing the time and resources required for identifying optimal process configurations. This finding will be discussed in more detail later, emphasizing its implications for the development of effective strategies for achieving optimum performance and efficiency in mixed refrigerant cryogenic liquefaction processes.

With the successful development of our machine learning (ML) model, we have gained the ability to explore the patterns and distributions of allowed compositions within the mixed refrigerant stream. This knowledge is crucial for optimizing the design of mixed refrigerant systems and ensuring their optimal performance. To provide a visual representation of these compositions, Fig. 13 illustrates the distribution

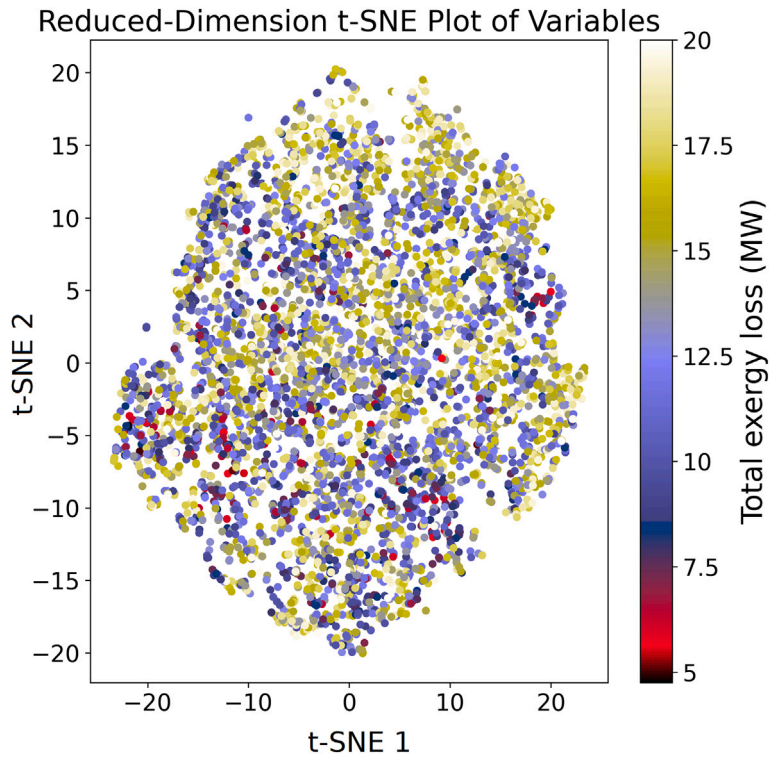

Fig. 10. Reduced-dimension t-SNE plot of variables. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

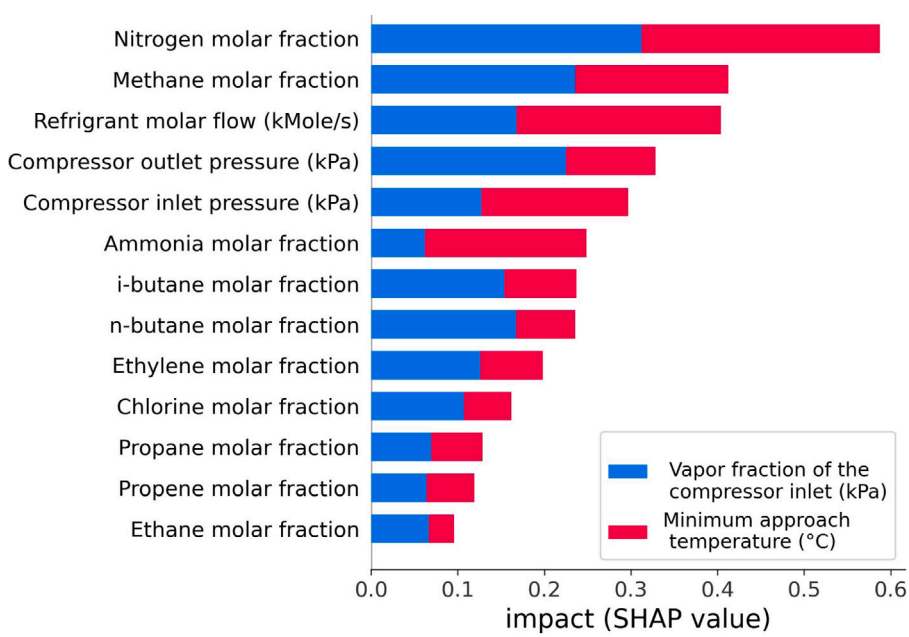

Fig. 11. Calculated SHAP values showing impact of features on finding valid design space.

of the mixed refrigerant composition as a function of the molar fraction of different components, with a specific focus on the valid design cases.

Upon analyzing the composition distribution, we uncover distinct patterns and ranges for various components present in the mixed refrigerant stream. Notably, gases such as butane, nitrogen, and methane exhibit broader permissible ranges in terms of their mole fraction. These components can occupy higher proportions within the mixed stream without compromising the validity of the design scenario. Ethylene behaves more like nitrogen and methane in comparison with butane according to Fig. 13. This is due to their similarity in boiling temperatures. In contrast, other gases exhibit narrower distributions, typically at lower concentrations or mole fractions. The existence of wider allowed ranges for the mole fraction of nitrogen, methane (and possibly butane) compared to other gases can be attributed to several factors. One significant factor is the thermodynamic properties and behavior of these specific gases. Nitrogen, methane, and butane possess characteristics that enable them to maintain stable and efficient operation within a broader range of mole fractions. These gases may have favorable thermodynamic interactions, allowing them to maintain their integrity and perform optimally even at higher concentrations. In

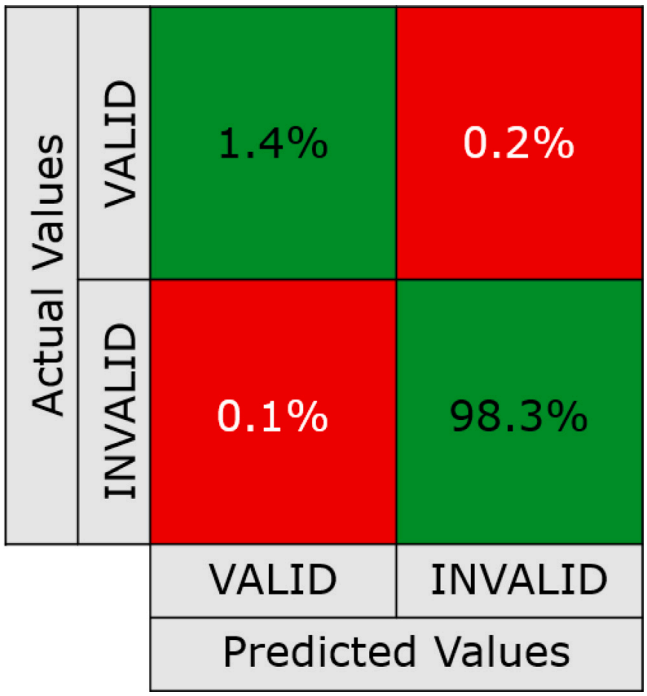

Fig. 12. Confusion matrix for the ML model predicting the validity of the design cases.

contrast, other gases may exhibit more sensitive behavior, experiencing limitations in stability or efficiency outside of narrower ranges.

Furthermore, the broader permissible range of nitrogen, methane, and possibly butane offers potential advantages in the design and operation of mixed refrigerant systems. The ability to utilize higher quantities of these gases opens up opportunities for enhanced refrigeration capacity and improved system efficiency. By allowing greater proportions of nitrogen, methane, and butane in the mixed stream, researchers and engineers can harness their desirable properties to achieve better performance and optimize the system design.

On the other hand, gases with limited distribution at lower concentrations or mole fractions require careful consideration and optimization during the design process. While their presence may be essential for specific applications or desired performance characteristics, it is crucial to ensure their effective integration within the mixed refrigerant stream. This may involve precise control, tailored system configurations, or additional optimization techniques to maintain their stability and prevent any detrimental effects on the overall system performance.

By incorporating the predictions of our ML model and the distribution of valid compositions, researchers and engineers gain valuable insights for establishing optimum ranges in the design of mixed refrigerant systems. This comprehensive approach ensures that the resulting system operates within the valid design space, minimizing the risk of temperature cross and liquid slugging, while maximizing overall performance and efficiency.

3.4. Optimizing LNG process using a data-driven approach

In this section, our research focus revolves around enhancing the energy efficiency and sustainability of cryogenic liquefaction processes through the optimization of exergy loss from a thermodynamic perspective. Cryogenic liquefaction processes are of great importance in improving energy efficiency and promoting sustainable practices, with particular emphasis on minimizing exergy loss. Although conventional process simulation tools like Aspen HYSYS have been extensively employed for process calculations and optimization, their computational speed limitations pose challenges when exploring the vast design space within reasonable timeframes.

To address these limitations, we employ a genetic algorithm (GA) in two distinct scenarios. Firstly, we utilize the GA in conjunction with the Aspen HYSYS model, where the results obtained from the model serve as inputs for the GA-based calculations. Secondly, we leverage a machine learning (ML) model as the engine for process calculations, which possesses remarkable computational speed, surpassing Aspen HYSYS by a factor of a million. Consequently, when both models are subjected to similar time constraints, the ML model enables a much wider coverage of the design space compared to Aspen HYSYS. Thus, implementing the GA on the ML model holds the promise of achieving improved results, specifically in terms of significantly lower optimized exergy loss.

To quantitatively evaluate the anticipated improvement, we set a total computational time limit of 24 h for each optimization case using the Aspen HYSYS model. While conducting the optimization process, we recognize that the efficiency of the GA is influenced by variables such as population size and mutation probability. Therefore, we systematically vary the GA parameters while keeping the total computational time fixed at 24 h, as depicted in Fig. 14. The optimized exergy loss values obtained range from 3.33 MW to 7.15 MW, depending on the specific GA variables employed.

In contrast, when implementing the GA and local search on the ML model, we achieve a remarkable global minimum exergy loss of 3.03 MW within a computational time of less than 10 min. This achievement represents a substantial 9% improvement in exergy loss compared to utilizing the ML model as the engine for process calculations, irrespective of the specific GA settings employed. The exceptional speed, robustness, and efficiency demonstrated by the developed ML model provide compelling reasons to consider similar strategies for optimizing industrial-oriented processes in a rapid and reliable manner.

This study's findings have significant implications for the field. Integrating advanced ML models in cryogenic liquefaction process optimization overcomes computational limitations and achieves unprecedented efficiency and sustainability. The substantial reduction in exergy loss improves energy efficiency, enabling greener practices. ML-driven approaches transform energy-intensive sectors, paving the way for eco-friendly and economically viable cryogenic processes. Industries like natural gas processing and petrochemicals can benefit, revolutionizing process design and operations for a more sustainable future.

3.5. Significance of the data-driven approach in finding valid process design scenarios with minimized exergy loss

The Fig. 15 provides a detailed visualization of the optimization results obtained for the mixed refrigerant liquefaction process. To achieve an optimal design, the genetic algorithm was employed with different weight factors assigned to the exergy loss, vapor fraction, and approach temperature.

The primary parameter of interest in this study is the exergy loss, as it directly impacts the thermodynamic efficiency of the process. However, it is essential to ensure that the resulting process design remains valid. To address this concern, the fitness function incorporates the vapor fraction and approach temperature as well. This allows any deviation from the desired process design conditions to be penalized.

The figure illustrates the distribution of the optimization results using color-coded points. The gray points represent invalid design cases, where the process parameters fail to meet the predefined constraints. Conversely, the dark purple points represent valid design cases with exergy losses above the threshold of 7.2 MW. The points falling within the range of exergy losses between 3.3 and 7.2 MW are color-coded between red and blue, indicating different levels of exergy loss.

The key takeaway from this figure is the significant role played by the weight factors in determining the success of achieving valid optimized design cases. It is evident that different combinations of

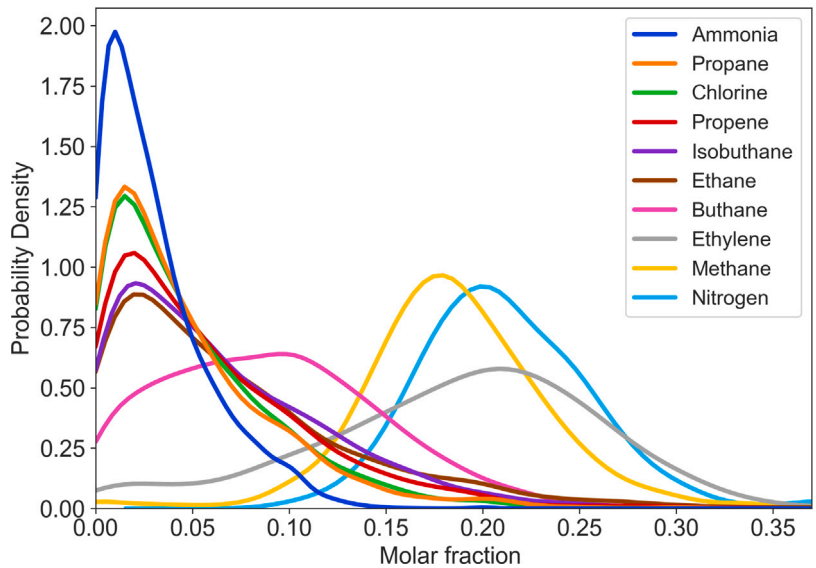

Fig. 13. The allowed molar compositions for the designed mixed refrigerant system.

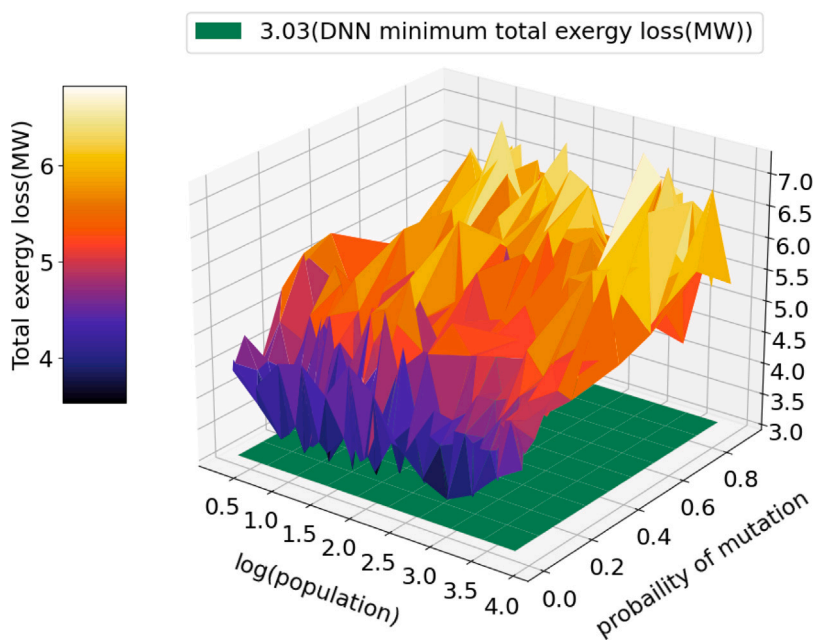

Fig. 14. Optimization of the exergy losses by genetic algorithm using the Aspen HYSYS and machine models.

weight factors may only lead to the identification of invalid design points, indicating the genetic algorithm's inability to converge on valid solutions despite the optimization process.

For the subset of weight factors that result in a valid design scenario, the figure showcases variations in both the rate of reaching the minimum exergy loss and the exact value of the final optimized exergy loss. This variability is evident from the changing colors and the final exergy values displayed in Fig. 15. The inset in the figure provides a closer examination of the final region of optimization.

To provide a clearer understanding of this variability, Fig. 16 presents the final points of the optimization process for each set of weight factors, using a triangle plot where each axis represents the value of a specific weight factor. Upon closer inspection, several optimization cases are observed to result in invalid design points, as indicated by the gray crosses.

Furthermore, among the sets of weight factors that led to valid design points, only a few managed to achieve design scenarios with an optimum exergy loss within a 5% error range, represented by the purple points. The other data points exhibit errors ranging between 5% and 15%, denoted by different colors such as green, yellow, and red.

This analysis highlights the challenges in finding a suitable combination of weight factors that lead to both valid and highly optimized design scenarios. Achieving a balance between the conflicting objectives of minimizing exergy loss, ensuring process validity, and reaching an optimal design configuration is complex and depends on the specific weight factor values employed. The variability observed in the triangle plot emphasizes the need for careful consideration and fine-tuning of weight factors to attain optimal results in mixed refrigerant liquefaction processes. Further investigations and sensitivity analyses are required to identify the most favorable weight factor settings for practical implementation and consistent performance improvements.

These findings provide valuable insights into the complexity of the optimization problem in the context of the mixed refrigerant liquefaction process. The presence of numerous invalid design scenarios

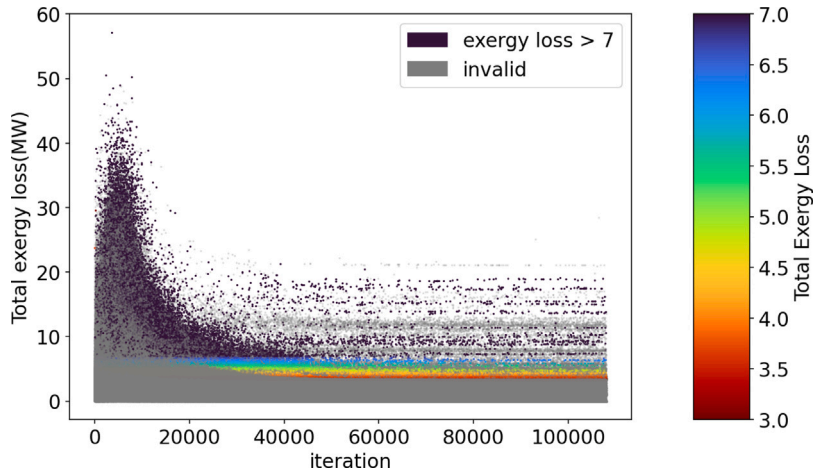

Fig. 15. Sensitivity of the optimization problem to the chosen weight factors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

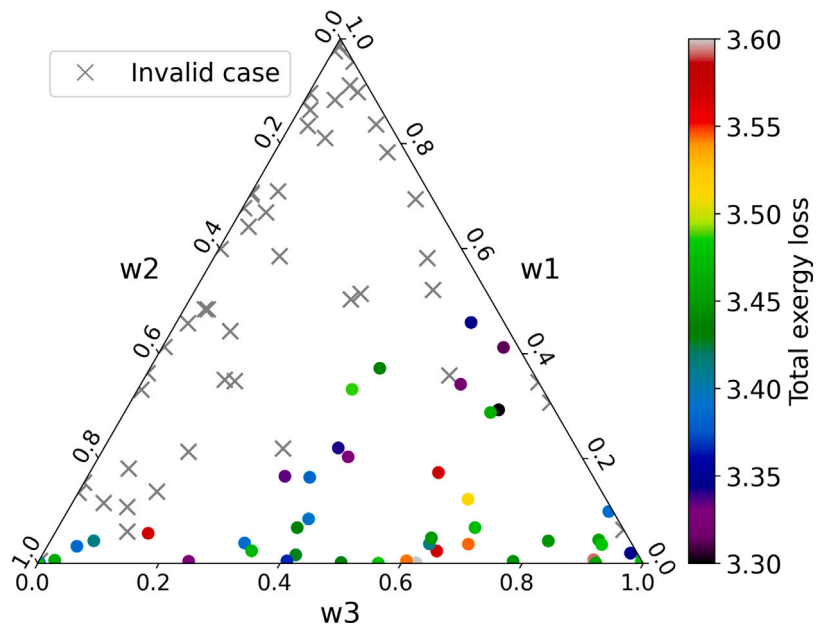

Fig. 16. The impact of fitness function weights on the optimal solution showing the complexity of the optimization problem. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

within the parameter space underscores the challenges associated with obtaining optimal solutions. Moreover, the significance of selecting appropriate weight factors becomes apparent when attempting to achieve valid design points with minimal exergy loss. The results depicted in both Figs. 15 and 16 emphasize the intricacies of optimizing the mixed refrigerant liquefaction process and shed light on the importance of weight factor selection for achieving desirable outcomes.

Fig. 17 provides a comprehensive visualization of the optimization process carried out using a genetic algorithm. The figure presents the evolution of the optimization in two distinct cases: one where appropriate weight factors for the vapor fraction and approach temperature were applied, and another where only the exergy loss was considered without incorporating the vapor fraction and approach temperature.

In the first case, the inclusion of suitable weight factors for the vapor fraction and approach temperature enabled the identification of valid design points with minimal exergy loss (green points) around the optimal value of 3.03 MW. This outcome signifies the significance of considering these factors in the optimization process to ensure the attainment of valid design scenarios with optimal performance. The weight factors provide a means to assign relative importance to different process outputs, allowing for a balanced optimization that takes into account various aspects of the system.

Conversely, when the optimization focused solely on minimizing the exergy loss without considering the vapor fraction and approach temperature, near-zero exergy loss values were achieved (red points). However, it is crucial to note that these low exergy losses were associated with invalid design cases. This observation emphasizes the necessity of incorporating all relevant parameters into the fitness function to avoid reaching invalid design areas. Neglecting important factors can lead to misleading results and undermine the practicality and feasibility of the optimized designs.

Furthermore, an intriguing finding in Fig. 17 is that when no weight factors were applied for the vapor fraction and approach temperature, very few valid points (blue points) were obtained throughout the entire optimization process. This suggests that these parameters play a vital role in defining the validity and performance of the design. Their omission severely limits the algorithm's ability to find valid design points, indicating the complexity and challenges involved in optimizing

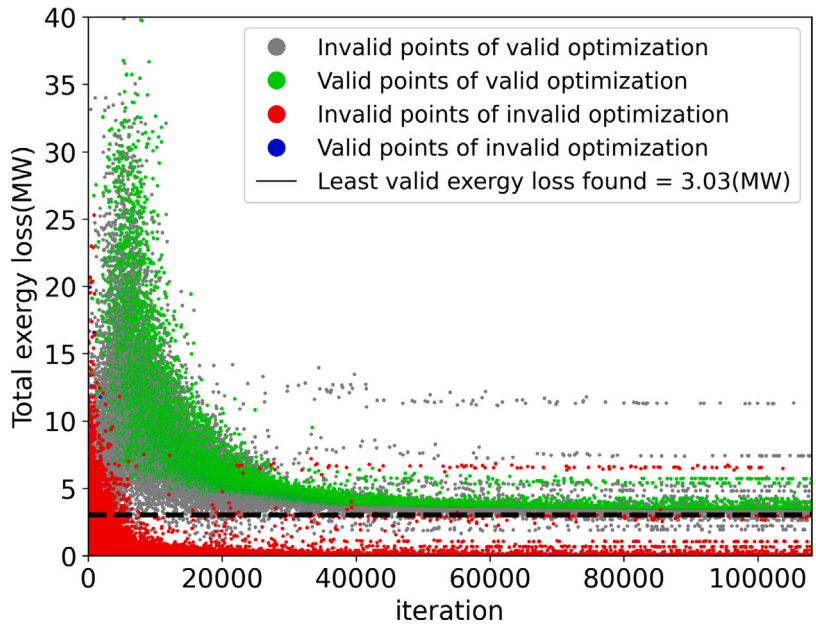

Fig. 17. The impact of detecting valid design space for reaching technically feasible process design solutions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

within a design space that contains a substantial number of invalid scenarios.

In conclusion, Fig. 17 provides valuable insights into the optimization process using a genetic algorithm, particularly when considering the weight factors for the vapor fraction and approach temperature. The results highlight the importance of incorporating these factors to achieve valid design scenarios with minimal exergy loss. Additionally, the figure demonstrates the pitfalls of focusing solely on exergy loss without accounting for other critical parameters, resulting in low exergy losses but with invalid designs. It underscores the significance of a comprehensive approach that considers multiple process outputs to ensure the feasibility and efficiency of the optimized designs.

Design optimization in a high-dimensional design space poses significant challenges, primarily due to the presence of numerous invalid points compared to valid design points. To address this issue, it is crucial to develop effective strategies for efficiently recognizing valid data points. In this context, we compare two different approaches to highlight the importance of fast recognition using machine learning (ML) in achieving optimized designs with minimal exergy loss.

In the first case, we employ a genetic algorithm with appropriate weight factors, utilizing Aspen HYSYS as the calculation engine. This approach relies solely on Aspen HYSYS for evaluating the exergy losses of different design points. On the other hand, in the second case, we integrate ML models into the genetic algorithm, allowing for the fast classification of valid and invalid design points. For the valid points, the exergy losses are then calculated using Aspen HYSYS, ensuring accurate assessments (see Figure S.2 for the full process diagram).

Fig. 18 illustrates the evolution of exergy losses in these two cases over multiple iterations. Despite both approaches ultimately utilizing Aspen HYSYS for exergy loss calculations, it becomes evident that the second case, where ML models are used for initial validation, achieves optimized values at a significantly faster rate. The primary reason for this expedited optimization is the ability of ML models to swiftly filter out invalid design points, thus avoiding the need for time-consuming exergy loss calculations using Aspen HYSYS.

The significance of this approach lies in the time-saving aspect it offers. By leveraging ML models to identify valid design points early on, the optimization process becomes more efficient. The computational burden on Aspen HYSYS is reduced as only valid design points are subjected to exergy loss calculations. Consequently, the overall optimization procedure is streamlined, allowing for faster convergence towards optimal design solutions with minimal exergy loss.

Table 2
GA results for composition in Case 1.

Solvent	Composition (mol %)
Methane	19
Ethane	7.82
Propane	0.4
n-butane	18.17
Nitrogen	19.49
Ammonia	1.98
Chlorine	0.51
Ethylene	24.16
Propene	3.22
i-butane	5.25

Furthermore, integrating ML models for preliminary validation provides an additional layer of confidence in the design selection process. By confirming the validity of design points using ML, we can mitigate the risk of investing computational resources in evaluating invalid points. This ensures that the subsequent exergy loss calculations carried out by Aspen HYSYS are performed on a reliable subset of design points, leading to more accurate and trustworthy results.

The findings from Fig. 18 highlight the potential and significance of employing fast ML models for recognizing valid design points in a coarse, high-dimensional design space. By incorporating these models into the optimization process, we expedite the search for optimized design cases with minimal exergy loss. This not only improves computational efficiency but also enhances the reliability of the design optimization process.

3.6. A discussion on comparison with previous studies

In Case 1, our cryogenic process model's exergy loss calculations are directly compared with previous works. The results are based on natural gas feed with a molar flow of 0.3 kg mole/s and Minimum Temperature Approach (MTA) set to 1.2 °C. Tables 2 and 3 present the outcomes of the genetic algorithm (GA) optimization and local search for Case 1, respectively.

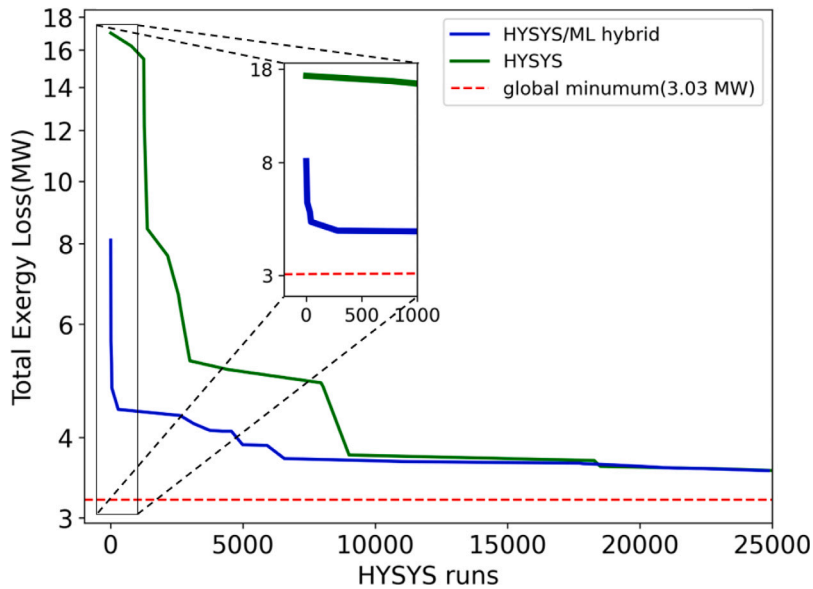

Fig. 18. The comparison of the search speed for the HYSYS/ML hybrid model compared to HYSYS GA optimization showing the importance of detection of the valid points by ML.

Table 3
Local search results for operational variables in Case 1.

Property	Value(Before local search)	Value(After local search)
Specific work (kJ/kg LNG)	5.92	5.79
Refrigerant molar (kg mole/s) flow	1.38	1.43
Compressor inlet pressure (kPa)	761.65	806.14
Compressor outlet pressure (kPa)	2862.15	2860.33

Comparing our results with Shivaee Gariz et al.'s current process (Shivaee-Gariz et al., 2020), we observe a significant 20.4% reduction in total exergy loss and a 9.5% decrease in shaft work required for the PRICO process (Table 4).

The GA's global minimum for total exergy loss is 3.23 MW, and after applying the local search, this value is further reduced to 3.03 MW, representing a 5.92% reduction. This highlights the coarseness of the feature landscape, necessitating deep scanning in the local zone to find optimal points (see Figure S.3 for the full Nelder–Mead progression). The need for the local search step indicates the complexity of the optimization problem and the presence of multiple local minima in the design space.

In Case 2, we compared the specific power requirement per amount of LNG produced in previous studies with our work, setting the minimum approach temperature to 0.1 °C. Tables 5 and 6 show the global and local optimization results.

Our result of 1005 kJ/kg LNG using 13 variables falls between the reported values of 967 kJ/kg LNG with 21 variables and 1105 kJ/kg LNG with 8 variables (Table 7). These results reveal the correlation between the number of variables to optimize and the invisible limit for the least specific work attainable. As the number of solvents increases, the potential to decrease shaft work using the right optimization techniques also increases.

In conclusion, our study showcases the effectiveness of the genetic algorithm and local search in optimizing cryogenic processes and achieving significant reductions in exergy loss and shaft work. By comparing with existing works, we demonstrate the superiority of our approach, providing valuable insights into the optimization challenges and underlying correlations between the number of variables and process efficiency. Our findings contribute to the advancement of energy-efficient and environmentally conscious industrial practices, paving the way for more sustainable and economically viable cryogenic liquefaction processes with broader applications in various industries. Embracing such advanced optimization techniques promises to revolutionize process design and operations, ushering in a new era of energy efficiency and sustainability.

4. Conclusions

In this study, we optimized mixed refrigerant cryogenic liquefaction processes through a data-driven approach. Our novel methodology integrates machine learning models with genetic algorithms to identify design points with minimized exergy loss, which significantly improves energy efficiency and sustainability. On the computational side, integration of machine learning accelerates valid design scenario recognition and enhances computational efficiency, while utilizing genetic algorithms with local search techniques substantially reduces exergy loss and specific work.

The study's implications for industries relying on cryogenic liquefaction processes, such as natural gas processing and petrochemicals, are noteworthy. Optimizing these energy-intensive sectors not only promotes environmentally sustainable practices but also promises economic benefits through reduced operational costs.

Our approach demonstrates superior efficacy when compared to conventional optimization techniques. The substantial reductions in total exergy loss achieved highlight the significant potential of our methodology in advancing cryogenic process design and operations. The success of our data-driven approach suggests promising outcomes for future research in integration of advanced optimization algorithms and heuristics with machine learning models. Additionally, exploring the impact of various process parameters and constraints provides valuable insights into the design space, contributing to achieving more stringent performance targets.

Furthermore, conducting sensitivity analyses and uncertainty quantification studies can deepen our understanding of the robustness and reliability of the optimized designs. Addressing the coarseness and high dimensionality of the design space remains an important challenge, and devising strategies to overcome these issues will be crucial for achieving even more efficiency.

Moreover, as data-driven approaches evolve, investigating the integration of real-time process data and advanced control strategies could enable the development of adaptive and self-optimizing cryogenic processes. The combination of online data-driven optimization

Table 4
Results for Case 1 and comparison to previous studies.

Property	Lee et al. (2002)	Shivaee-Gariz et al. (2020)	This work(before local search)	This work(after local search)
Specific work (MW)	7.98	7.27	5.92	5.79
Total exergy loss (MW)	3.74	3.35	3.23	3.03

Table 5
GA results for composition in Case 2.

Solvent	Composition (mol %)
Methane	17.20
Ethane	11.48
Propane	0.00
n-butane	18.85
Nitrogen	16.04
Ammonia	0.00
Chlorine	0.00
Ethylene	23.86
Propene	0.00
i-butane	12.56

Table 6
Local search results for Case 2.

Property	Value(Before local search)	Value(After local search)
Specific work (kJ/kg LNG)	1165	1005
Refrigerant molar (kg mole/s) flow	1.24	1.26
Compressor inlet pressure pressure (kPa)	575.61	681.62
Compressor outlet pressure (kPa)	2651.4	2556.41

Table 7
Comparison of specific power requirements in this work with previous studies for Case 2.

Property	Aspelund et al. (2010)	Ebrahimi et al. (2021)	This work
Specific power (kJ/kg LNG)	1105	967	1005
Number of variables	8	21	13
Number of solvents	5	17	10
Minimum approach temperature	0.1	0.1	0.1

and machine learning-based control could lead to dynamic and responsive operations that adapt to changing conditions, thereby ensuring continued efficiency and sustainability.

This study serves as a stepping stone for further advancements in energy-efficient and environmentally conscious industrial practices, opening doors for future works to explore a multitude of opportunities in cryogenic liquefaction and beyond. With the ever-increasing computational power and advancements in data-driven methodologies, we are confident that our approach will continue to pave the way for a more sustainable and economically viable future in energy-intensive processes.

Declaration of competing interest

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

Acknowledgments

M.A. acknowledges an Early Postdoc.Mobility (EPM) fellowship from the Swiss National Science Foundation (SNSF) (Project ID: P2ELP2_195134). M.A. expresses gratitude to the European Research Council (ERC) for evaluating the project with the reference number 101106377 titled "CLARIFIER" and accepting it for funding under the HORIZON TMA MSCA Postdoctoral Fellowships - European Fellowships. Furthermore, M.A. acknowledges the funding provided by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (Grant Reference: EP/Y023447/1; Organization Reference: 101106377).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.dche.2024.100143.

References

Alabdulkarem, A., Mortazavi, A., Hwang, Y., Radermacher, R., Rogers, P., 2011. Optimization of propane pre-cooled mixed refrigerant LNG plant. Appl. Therm. Eng. 31 (6–7), 1091–1098.
Ali, W., Khan, M.S., Qyyum, M.A., Lee, M., 2018. Surrogate-assisted modeling and optimization of a natural-gas liquefaction plant. Comput. Chem. Eng. 118, 132–142.
Angelova, M., Pencheva, T., 2011. Tuning genetic algorithm parameters to improve convergence time. Int. J. Chem. Eng. 2011.
Arie, M., Shooshtari, A., Dessiatoun, S., Al-Hajri, E., Ohadi, M., 2015. Numerical modeling and thermal optimization of a single-phase flow manifold-microchannel plate heat exchanger. Int. J. Heat Mass Transfer 81, 478–489.
Aspelund, A., Gundersen, T., Myklebust, J., Nowak, M., Tomasgard, A., 2010. An optimization-simulation model for a simple LNG process. Comput. Chem. Eng. 34 (10), 1606–1617.
Cammarata, G., Fichera, A., Guglielmino, D., 2001. Optimization of a liquefaction plant using genetic algorithms. Appl. Energy 68 (1), 19–29.
Cheng, W.B., Mah, R.S., 1980. Interactive synthesis of cascade refrigeration systems. Ind. Eng. Chem. Process Des. Dev. 19 (3), 410–420.
Curtis, F.E., Nocedal, J., 2008. Flexible penalty functions for nonlinear constrained optimization. IMA J. Numer. Anal. 28 (4), 749–769.
Deb, K., Datta, R., 2013. A bi-objective constrained optimization algorithm using a hybrid evolutionary and penalty function approach. Eng. Optim. 45 (5), 503–527.
Ebrahimi, A., Tamnanloo, J., Mousavi, S.H., Soroodan Miandoab, E., Hosseini, E., Ghasemi, H., Mozaffari, S., 2021. Discrete-continuous genetic algorithm for designing a mixed refrigerant cryogenic process. Ind. Eng. Chem. Res. 60 (20), 7700–7713.
Eldemerdash, U., 2011. Technology review of natural gas liquefaction processes. J. Appl. Sci..
Gen, M., Cheng, R., 1999. Genetic Algorithms and Engineering Optimization, vol. 7, John Wiley & Sons.
Gundersen, T., Grossmann, I.E., 1990. Improved optimization strategies for automated heat exchanger network synthesis through physical insights. Comput. Chem. Eng. 14 (9), 925–944.
He, T., Ju, Y., 2014. Design and optimization of a novel mixed refrigerant cycle integrated with NGL recovery process for small-scale LNG plant. Ind. Eng. Chem. Res. 53 (13), 5545–5553.
Homaifar, A., Qi, C.X., Lai, S.H., 1994. Constrained optimization via genetic algorithms. Simulation 62 (4), 242–253.
Jimenez, E.S., Salamon, P., Rivero, R., Rendon, C., Hoffmann, K.H., Schaller, M., Andresen, B., 2004. Optimization of a diabatic distillation column with sequential heat exchangers. Ind. Eng. Chem. Res. 43 (23), 7566–7571.
Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nat. Rev. Phys. 3 (6), 422–440.
Keyvanloo, K., Sedighi, M., Towfighi, J., 2012. Genetic algorithm model development for prediction of main products in thermal cracking of naphtha: Comparison with kinetic modeling. Chem. Eng. J. 209, 255–262.
Khan, A., Lapkin, A., 2020. Searching for optimal process routes: A reinforcement learning approach. Comput. Chem. Eng. 141, 107027.
Khan, M.S., Lee, S., Rangaiah, G., Lee, M., 2013. Knowledge based decision making method for the selection of mixed refrigerant systems for energy efficient LNG processes. Appl. Energy 111, 1018–1031.
Kim, S., Jang, M.-G., Kim, J.-K., 2023. Process design and optimization of single mixed-refrigerant processes with the application of deep reinforcement learning. Appl. Therm. Eng. 223, 120038.
Kramer, O., 2017. Genetic Algorithm Essentials, first ed. In: Studies in Computational Intelligence, Springer International Publishing, Basel, Switzerland.
Kumar, S., Kwon, H.-T., Choi, K.-H., Lim, W., Cho, J.H., Tak, K., Moon, I., 2011. LNG: An eco-friendly cryogenic fuel for sustainable development. Appl. Energy 88 (12), 4264–4273.
Lee, G., Smith, R., Zhu, X., 2002. Optimal synthesis of mixed-refrigerant systems for low-temperature processes. Ind. Eng. Chem. Res. 41 (20), 5016–5028.

Leipold, M., Gruetzmann, S., Fieg, G., Maschmeyer, D., Sauer, J., Wiederhold, H., 2008. Development of a sophisticated framework for complex single-and multi-objective optimization tasks. In: Computer Aided Chemical Engineering, vol. 25, Elsevier, pp. 695–700.

Lu, T., Wang, K., 2009. Analysis and optimization of a cascading power cycle with liquefied natural gas (LNG) cold energy recovery. Appl. Therm. Eng. 29 (8–9), 1478–1484.

McNelis, P.D., 2005. Neural Networks in Finance: Gaining Predictive Edge in the Market. Academic Press.

Misra, D., 2019. Mish: A self regularized non-monotonic neural activation function. CoRR abs/1908.08681.

Mitchell, M., 1996. An Introduction to Genetic Algorithms. MIT Press, London, England.

Mokhatab, S., Mak, J., Valappil, J., Wood, D.A., 2013. Handbook of Liquefied Natural Gas. Gulf Professional Publishing.

Morin, A., Wahl, P.E., Mølnvik, M., 2011. Using evolutionary search to optimise the energy consumption for natural gas liquefaction. Chem. Eng. Res. Des. 89 (11), 2428–2441.

Mosayebi, M., Sodhi, M., 2020. Tuning genetic algorithm parameters using design of experiments. In: Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion. pp. 1937–1944.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7 (4), 308–313.

Nogal, F.D., Kim, J.-K., Perry, S., Smith, R., 2008. Optimal design of mixed refrigerant cycles. Ind. Eng. Chem. Res. 47 (22), 8724–8740.

Pyzer-Knapp, E.O., Pitera, J.W., Staar, P.W.J., Takeda, S., Laino, T., Sanders, D.P., Sexton, J., Smith, J.R., Curioni, A., 2022. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. NPJ Comput. Mater. 8 (1).

Qyyum, M.A., Duong, P.L.T., Lee, S., Lee, M., et al., 2019. Dual mixed refrigerant LNG process: Uncertainty quantification and dimensional reduction sensitivity analysis. Appl. Energy 250, 1446–1456.

Reddy, H.V., Bisen, V.S., Rao, H.N., Dutta, A., Garud, S.S., Karimi, I.A., Farooq, S., 2019. Towards energy-efficient LNG terminals: Modeling and simulation of reciprocating compressors. Comput. Chem. Eng. 128, 312–321.

Rocha, M., Neves, J., 1999. Preventing premature convergence to local optima in genetic algorithms via random offspring generation. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer, pp. 127–136.

Rodríguez-Pérez, R., Bajorath, J., 2019. Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. J. Med. Chem. 63 (16), 8761–8777.

Sanavandi, H., Ziabasharhagh, M., 2016. Design and comprehensive optimization of C3MR liquefaction natural gas cycle by considering operational constraints. J. Nat. Gas Sci. Eng. 29, 176–187.

Sangahn, K., 2019. Variable selection-based SPC procedures for high-dimensional multistage processes. J. Syst. Eng. Electron. 30 (1), 144–153.

Schweidtmann, A.M., Esche, E., Fischer, A., Kloft, M., Repke, J.-U., Sager, S., Mitsos, A., 2021. Machine learning in chemical engineering: A perspective. Chem. Ing. Tech. 93 (12), 2029–2039.

Shivaee-Gariz, R., Tahouni, N., Panjeshahi, M.H., Abbasi, M., 2020. Development of a new graphical tool for calculation of exergy losses to design and optimisation of sub-ambient processes. J. Clean. Prod. 275, 123161.

Taleshbahrami, H., Saffari, H., 2010. Optimization of the C3MR cycle with genetic algorithm. Trans. Can. Soc. Mech. Eng. 34 (3–4), 433–448.

Tran, P.T., Phong, L.T., 2019. On the convergence proof of AMSGrad and a new version. IEEE Access 7, 61706–61716.

Veloso, T.G.C., Sotomonte, C.A.R., Coronado, C.J., Nascimento, M.A., 2018. Multi-objective optimization and exergetic analysis of a low-grade waste heat recovery ORC application on a Brazilian FPSO. Energy Convers. Manage. 174, 537–551.

Wang, Z., Zhou, T., Sundmacher, K., 2022. Interpretable machine learning for accelerating the discovery of metal-organic frameworks for ethane/ethylene separation. Chem. Eng. J. 444, 136651.

Xu, X., Liu, J., Jiang, C., Cao, L., 2013. The correlation between mixed refrigerant composition and ambient conditions in the PRICO LNG process. Appl. Energy 102, 1127–1136.

Zhu, W., Chebeir, J., Romagnoli, J.A., 2020. Operation optimization of a cryogenic NGL recovery unit using deep learning based surrogate modeling. Comput. Chem. Eng. 137, 106815.

Hosein Alimardani received his B.Sc. degree in Chemical Engineering from the University of Tehran (UT) in 2020. Currently, he is pursuing his M.Sc. degree at the University of Tehran (UT), Tehran, Iran, with a focus on the applications of machine learning for the design and optimization of chemical processes. His research interests lie at the intersection of chemical engineering and data-driven methodologies, seeking innovative approaches to enhance process efficiency and sustainability.

Dr. Mehrdad Asgari completed his Ph.D. at EPFL in January 2020, conducting research under the guidance of Prof. Wendy Lee Queen. His doctoral work focused on characterizing metal–organic frameworks for carbon capture applications. Following his Ph.D., Dr. Asgari continued as a postdoctoral researcher at EPFL, collaborating with Prof. Berend Smit and Prof. Francois Maréchal. During this time, he worked at the interface of experiments and computations, utilizing computational toolboxes to study and characterize metal–organic frameworks for gas adsorption and storage. In October 2021, Dr. Asgari received the SNSF Early Postdoctoral Mobility Fellowship and became a Marie-Curie postdoctoral fellow at the University of Cambridge. His current research builds upon his expertise in metal–organic frameworks and explores advanced applications in gas adsorption, storage, and catalysis. Throughout his academic journey, Dr. Asgari has demonstrated a strong commitment to cutting-edge research and the development of sustainable technologies. His contributions to the fields of chemical engineering and materials science hold great promise for addressing critical challenges in energy and environmental sustainability.

Roohangiz Shivaee Gariz is a chemical engineer with a passion for process optimization and energy-intensive industries. In 2017, he earned his B.Sc. degree in Chemical Engineering from the prestigious University of Tehran (UT), followed by an M.Sc. degree in Chemical Engineering from UT in 2019, achieving honors in both programs. Currently, Roohangiz is pursuing a doctoral degree at DTU (Technical University of Denmark) with a focus on Process Studies of Electrically Heated Clay Calciners. As a Ph.D. student, she is dedicated to exploring innovative approaches for enhancing energy efficiency and sustainability in industrial processes. Her research contributes to the advancement of environmentally conscious and economically viable solutions in the field of chemical engineering. With a strong background in chemical engineering and a keen interest in data-driven optimization, Roohangiz aims to address complex challenges in cryogenic liquefaction processes and other energy-intensive systems. Her work exemplifies a commitment to shaping a greener and more sustainable future for industries that rely on cryogenic processes, natural gas processing, and petrochemicals.

Javad Tamnanloo is a dedicated researcher in chemical engineering, having earned his B.Sc. degree from the University of Tehran (UT) in 2012 and his M.Sc. degree from UT in 2014, both with honors. Currently, he is pursuing his Ph.D. at the School of Polymer Science and Polymer Engineering at the University of Akron, OH. Javad's research focuses on utilizing machine learning for creating innovative representations and quantitative structure–property relationships in polymer science, aiming to advance materials design and engineering.