This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3615454

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version December 27, 2024.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Machine Learning for Fluid Thermodynamics: State of The Art and Open Challenges

FEDERICO SUCCETTI[1], MASSIMO PANELLA[1] (Senior Member, IEEE), PAOLO GIANNITRAPANI[2] JEAN-CHRISTOPHE RIGO[2] and STEFANIA COLONNESE[1] (Senior Member, IEEE)

[1]Department of Information Engineering, Electronics and Telecommunications (DIET), University of Rome "La Sapienza", Rome, Italy (e-mail: {federico.succetti, massimo.panella, stefania.colonnese}@uniroma1.it)
[2]Baker Hughes, Florence, Italy (e-mail: {paolo.giannitrapani, jeanchristophe.rigo}@bakerhughes.com)

Corresponding author: Prof. Stefania Colonnese (e-mail: stefania.colonnese@uniroma1.it).

ABSTRACT This paper presents a critical survey on adopting machine learning in solving complex real fluid thermodynamics problems. After reviewing the primary computational machine learning frameworks employed in thermodynamic modelling, we have analysed current research with a particular emphasis on properly estimating gas and liquid properties, vapour-liquid equilibrium, and supercritical fluids, focusing on pure gases. While ML offers a powerful paradigm for augmenting or even replacing traditional methods, its application faces significant open challenges. Key issues include the persistent trade-off between model accuracy and computational efficiency, the difficulty in capturing highly non-linear behaviour, especially near critical points or under extreme conditions, and the pervasive problem of data scarcity. We conclude the paper by introducing the main datasets available for thermodynamic property computation, such as the results of the GERG2008 project, and others relevant to turbomachinery applications. This survey provides a unified perspective on machine learning architectures used in thermodynamics and identifies open challenges and potential future advancements for enhancing predictive accuracy and efficiency while reducing execution time.

INDEX TERMS Thermodynamics applications, machine learning, deep learning, hybrid models, thermophysical properties, gas properties, liquid properties.

I. INTRODUCTION

REAL fluid thermodynamical problems require accurately measuring, modelling, and predicting physical variables to describe phenomena that are difficult to represent by traditional equations of state, such as phase transitions near critical points [1] or liquid–liquid transitions in tetrahedral fluids, where topological changes play a key role in phase stability [2]. Recent studies have shown that even in supercooled states, molecular disorder persists until a hidden transition occurs, revealing solid-like characteristics at a specific temperature threshold [3].

Similarly, non-ideal behaviour of multi-component mixtures under varying temperatures and pressures presents significant challenges, particularly in industrial applications such as turbomachinery design. For example, recent research has identified thermodynamic crossovers in supercritical fluids, refining the traditional understanding of the gas–liquid continuum beyond the critical point [4]. Prediction of high-pressure behaviour and supercritical regions, crucial for energy efficiency and industrial processes, often requires integrating molecular-level interactions into macroscopic models. A novel theoretical framework has also been proposed for understanding temperature regulation via phase transitions, providing a unified approach for applications in cooling and heating technologies [5].

Computational efficiency remains a significant issue for large-scale simulations, especially when estimating flow path characteristics, phase changes, heat transfer, and transport properties in a global context of energy loss minimization. These limitations call for innovative solutions in the real fluid thermodynamic domain. Therefore, in response to these persistent challenges, Machine Learning (ML) techniques have recently emerged as a powerful paradigm, offering a rich variety of solutions to different classes of problems, as visually summarized in Fig. 1.

While several surveys deal with thermodynamic problems [6]–[13], the literature lacks a systematical analysis of ML computational methods employed in this field. Here, we provide a comprehensive review of the current research trends on ML for thermodynamics, particularly focusing on prominent

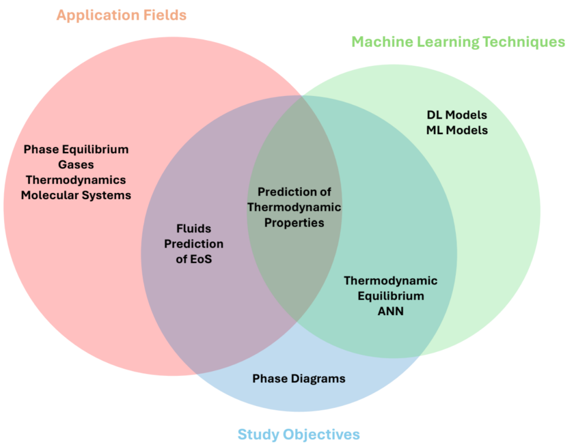


FIGURE 1. **Main thermodynamical problems and related Machine Learning approaches covered by the literature reported in this survey.**

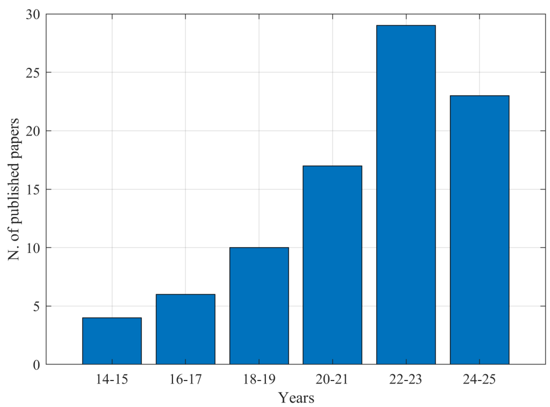


FIGURE 2. **Distribution of published papers as a function of age.**

issues such as vapour-liquid equilibrium, supercritical behaviour modelling, and accurate property evaluations across wide pressure and temperature ranges. Figure 2 gives an idea about the field's growth. It presents a bar diagram illustrating the frequency distribution of published papers related to the topics covered in this survey across different age groups, spanning from 2014-2015 to 2024-2025. The x-axis categorises individuals into distinct two-year age intervals, while the y-axis quantifies the number of papers published within each respective age group. The data reveals a clear trend of increasing publication activity with age, peaking in the 22-23 age group before a slight decline. To be specific, the latter is not properly a decline, since the year 2025 is not finished. The scientific survey is accompanied by an overview of the datasets available for computing thermodynamic properties. Finally, we draw a path for further relevant open research directions and architectural solutions, showing the largely unexplored potentiality of ML for thermodynamics.

The structure of the paper is as follows: Sect. II reviews the main ML models employed in the literature; Sect. III provides a critical overview of Artificial Intelligence (AI) in thermodynamic problems; Sect. IV describes the main components for the simulation environments, providing an overview of the available datasets; Sect. V describes open challenges and future developments; finally, Sect. VI provides some concluding remark.

II. MACHINE LEARNING MODELS IN THERMODYNAMICS: A REVIEW

ML solutions are characterized by a variety of architectures. Among others, some main classes of solutions are more frequently adopted in the herein surveyed papers. Specifically, the main ML methods adopted in the literature for solving thermodynamical problems belong to two classes: Deep Learning (DL) methods, and more in general, the class of Regression Techniques (RTs), whose computation architectures are shown in Fig. 3.

TABLE 1. **Common Metrics for Estimating the Accuracy of the Models.**

Metrics	Formula
Mean Absolute Error	$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$
Mean Squared Error	$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$
Root Mean Squared Error	$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$
Coefficient of determination	$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$
Mean Absolute Percentage Error	$\mathrm{MAPE}=\frac{100\%}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert$
Mean Percentage Error	$\mathrm{MPE}=\frac{100\%}{n}\sum_{i=1}^{n}\frac{y_i-\hat{y}_i}{y_i}$
Maximum Percentage Error	$\mathrm{PE}_{\max}=\max\left(\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}\times 100\%\right)$
Average Absolute Relative Deviation	$\mathrm{AARD}=\frac{100\%}{n}\sum_{i=1}^{n}\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}$
Average Absolute Percent Deviation	$\mathrm{AAPD}=\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert\times 100\%$
Mean Squared Logarithmic Error	$\mathrm{MSLE}=\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{1+y_i}{1+\hat{y}_i}\right)^2$
Mean Relative Error	$\mathrm{MRE}=\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}$
Mean Relative Percentage Error	$\mathrm{MRPE}=\frac{100\%}{n}\sum_{i=1}^{n}\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}$
F1-Score	$F1=2\cdot\frac{\mathrm{Precision}\cdot\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}$
Pearson Correlation Coefficient	$\rho=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}\cdot\sqrt{\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2}}$
Cross-Entropy	$\mathrm{CE}=-\sum_{i=1}^{n}\left(\log(\hat{y}_i^{y_i}\cdot 1-\hat{y}_i)^{1-y_i}\right)$

Neural Networks (NNs) encompass models like Feed-Forward Neural Networks (FFNNs), which consist of layers of interconnected neurons passing information in one direc-

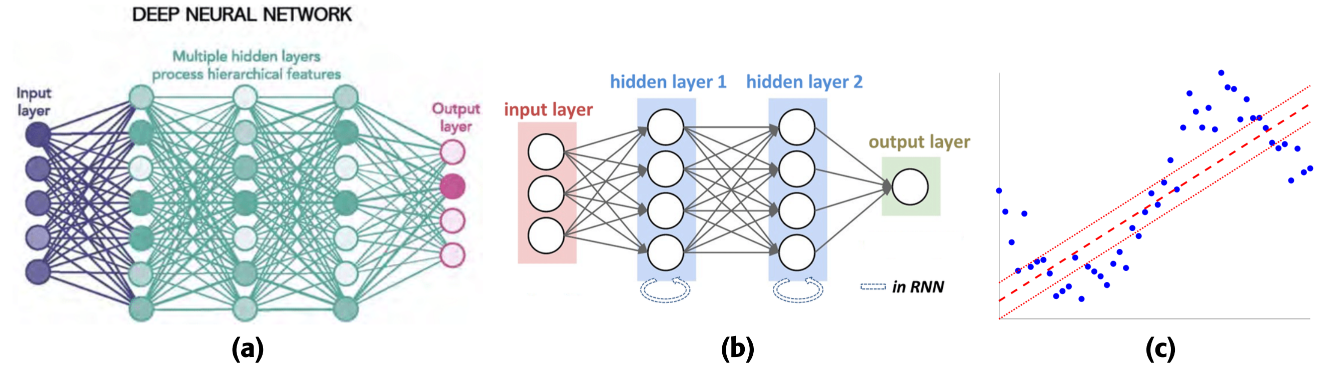


FIGURE 3. Machine Learning classes: Deep Neural Networks encompass models like Feed-Forward (a) and Recurrent (b) Neural Networks, which consist of layers of interconnected neurons passing information in one direction (a) [14] or both (b) [15]; (c) Regression Techniques involve statistical models for suitable parametric data fitting.

tion. Multilayer Perceptrons (MLPs) are a specific type of FFNNs that have multiple hidden layers between the input and output [16]. Physics-Informed Neural Networks (PINNs) incorporate physical laws into the training process to ensure solutions adhere to known scientific principles [17]. Since PINNs are a relatively recent and rapidly evolving field of application within ML, we provide a brief overview about their applications in fluid thermodynamics in the following of this Section. Graph Neural Networks (GNNs) operate on graph-structured data, capturing relationships between nodes [18]. Their variant, Graph Convolutional Neural Networks (GCNNs), extends GNNs by applying convolution operations to graph data. Regarding unsupervised models, Autoencoders (AEs) are designed for dimensionality reduction or feature learning by encoding data into a compressed representation, while Generative Autoencoders (GAEs) go a step further by generating new data samples from learned distributions [19].

PINNs represent a significant advancement in DL for addressing complex problems in fluid thermodynamics by integrating physical laws directly into the NN's training process. This approach ensures that solutions adhere to known scientific principles, making them particularly effective in scenarios with scarce or noisy data and when physical consistency is paramount [20]. PINNs achieve this by incorporating governing equations, typically partial differential equations (PDEs), into the loss function, leveraging automatic differentiation to evaluate differential operators without discretization errors [20], [21].

Recent state-of-the-art developments in PINNs further enhance their capabilities in fluid dynamics and thermodynamics. These include the emergence of Physics-Informed Convolutional Long Short Term Memory Statistical Models, which significantly reduce computational costs for long-term fluid dynamics simulations while replicating key features of complex behaviors like Rayleigh-Bénard convection [22]. PINNs have also demonstrated success in capturing phase interfaces in two-phase flows and modeling hydrodynamics and heat transfer, showing robustness across various geometries and fluid properties [23]. Furthermore, Thermodynamics-Informed Graph Neural Networks are advancing the field by incorporating geometric and thermodynamic inductive biases to improve the accuracy and generalization of predictions in dissipative dynamic systems [18]. A notable evolution is the development of Physics-Informed Kolmogorov-Arnold Networks (PIKANs), offering promising alternatives to traditional PINNs through advancements in network design, feature expansion, optimization techniques, and uncertainty quantification [24]. While PINNs offer substantial benefits in enhancing model performance, improving data efficiency, and providing stable predictions, challenges persist in managing PDE discretizations, imposing boundary conditions, and addressing potential training instabilities, highlighting the ongoing need for more robust neural network architectures and advanced training algorithms, as well as a more rigorous mathematical framework [25].

From a ML point of view, Tree-Based Methods (TBMs) include Random Forest (RF), a model that builds an ensemble of decision trees to improve predictive accuracy by averaging their outputs [26]. Gradient Boosting (GB) regression, which sequentially builds models, where each one corrects errors made by the previous ones [27]. AdaBoost and XGBoost are boosting algorithms, where AdaBoost emphasizes correcting errors by weighting misclassified examples more heavily in subsequent models, while XGBoost adds regularization and optimization enhancements for faster and more accurate predictions [28], [29]. On a same level, RTs involve statistical models like Ridge Regression, which extends linear regression by adding a regularization term to prevent overfitting. Support Vector Regression (SVR) finds the hyperplane that best fits the data by minimizing errors and controlling model complexity [28]. Gaussian Process Regression (GPR) is a non-parametric method that provides probabilistic predictions by modelling data as a distribution of functions, offering uncertainty estimates in predictions [30].

In general, each method class is based on a distinct logic, with NNs relying on layered learning and feature extraction, TBMs focusing on ensemble learning and error correction, and RTs using mathematical formulations for predictive modelling. It is worth mentioning that transfer learning represents another important method in this context. However, we decided not to introduce it in this survey since it represents a highly specialised topic and is worthy of its dedicated review.

A. PERFORMANCE METRICS

The estimation accuracy performance metrics primarily evaluate the precision of predictive models by comparing predicted values to actual or experimental data. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE) are commonly used to assess the magnitude of errors and the correlation between predicted and actual values. Percentage-based errors, such as Mean Absolute Percentage Error (MAPE), Relative Percentage Error (RPE), and Average Absolute Relative Deviation (AARD), are employed to evaluate prediction accuracy in relative terms. Some metrics, like the Pearson correlation coefficient and the coefficient of determination (R^2), measure the degree of correlation between predicted and real data. Errors can also be computed for specific parameters like vapour-liquid equilibria, phase stability, and thermodynamic properties, reflecting the accuracy of predictions in specialized fields. Additionally, bias and deviation metrics gauge systematic errors and variability in predictions, often ensuring models avoid overfitting and maintain predictive reliability across various conditions. A summary of metrics is reported in Table 1.

III. ARTIFICIAL INTELLIGENCE IN THERMODYNAMICS PROBLEMS: A CRITICAL OVERVIEW

AI for addressing fluid thermodynamic challenges has found applications in numerous industrial problems. Before detailing the specific contributions, in Table 2 we report a comparison between the proposed review and similar ones in the literature. The table highlights the topics covered and the gaps filled by each review. For instance, Desgranges et al. [31]

TABLE 2. Overview of Machine Learning Surveys: Topics Covered and Gap Addressed.

Survey	Topics covered	Gap addressed
[31]	Integration of ML with molecular simulation (MS) to improve the efficiency of thermodynamic predictions. Application of ML models to predict entropy, Helmholtz free energy, and Gibbs free energy. Challenges in computing these properties due to high computational costs of molecular simulations. ML-enhanced simulations for phase transitions and protein folding. Application in metal–organic frameworks (MOFs) and covalent organic frameworks (COFs) for energy and environmental applications.	Reducing computational costs associated with free energy calculations. Using ML-guided reaction coordinates for enhanced sampling. Enabling rapid screening of materials for gas storage, separation, and catalysis.
[32]	Data-driven and physics-informed ML models for Computational Fluid Dynamics (CFD). ML-enhanced numerical solvers and surrogate models. ML-driven inverse design and control strategies. Applications in aerodynamics, combustion, ocean/atmospheric modeling, and plasma physics. Challenges in multi-scale modeling, physics integration, and AI-driven scientific discovery.	Lack of a structured survey on recent ML advancements in CFD. Need for improved generalization and physics-aware models. Challenges in computational efficiency and numerical stability.
[33]	Importance of high-performance computing and data mining for fluid simulations. ML-based extraction of transport properties, phase behavior, and thermodynamic properties. Use of Physics-Informed Neural Networks (PINNs) and hybrid AI techniques to solve Navier-Stokes equations. Leveraging ML for parametrization of unresolved scales in high-Reynolds-number flows. Performance evaluation of ML models in bulk and confined Lennard-Jones (LJ) fluid simulations.	Lack of a structured review on AI/ML applications across different fluid research domains. Challenges in integrating ML with traditional physics-based models. Need for better interpretability and generalization in ML-based fluid models, especially for symbolic regression and hybrid AI methods.
[34]	Nanofluids, delving into the prediction of their thermo-fluid behaviors using ML techniques. Detailed analysis of ML models tailored for nanofluid behavior prediction. Specific algorithms and their performance. Data quality and the necessity for reliable datasets in training ML models for nanofluid predictions.	Complexity of nanofluid behavior. Difficulty in experimental and numerical simulations. Nonlinearity of thermo-fluid properties. Lack of robust prediction models. Optimization of model performance. Data quality and generalization.
[35]	The development of a thermodynamically consistent EoS using a machine-learning model trained on molecular simulation data. A focus on the Helmholtz free energy as a foundation for constructing physically meaningful fluid property models. Utilization of molecular dynamics (MD) simulations to generate training data for the Mie fluid, a generalization of the Lennard-Jones potential. Prediction of first- and second-order thermodynamic properties, including pressure, internal energy, heat capacities, compressibility, and phase equilibria. Use of derivative-based loss functions to incorporate higher-order thermodynamic properties into training.	Lack of a unified, thermodynamically consistent ML-based EoS. Challenges in integrating machine learning with fundamental thermodynamics. Computational inefficiency of traditional thermodynamic modeling.
[36]	Development of DNN-based models trained on datasets of mixture compositions and their computed critical points. Use of DNN predictions to initialize traditional root-finding and optimization methods. Integration with Damped Newton-Raphson (DNR) and Differential Evolution (DE) algorithms. Addresses challenges in modeling complex hydrocarbon mixtures.	Existing root-finding and global optimization techniques require many iterations, making them slow for multicomponent mixtures. DNNs significantly enhance convergence rate and robustness without compromising accuracy. Enforces thermodynamic constraints and validates predictions against traditional algorithms, ensuring scientific reliability.

review ML methods for predicting free energy landscapes in molecular and biological systems, with a particular focus on material science applications such as gas storage and separation, while also addressing computational efficiency in molecular simulations. On the other hand, Wang et al. [32] examine ML applications in CFD, covering fluid flow modelling, ML-enhanced solvers, and inverse design and control strategies for areas including aerodynamics, combustion, and climate modelling. On a same level, Sofos et al. [33] survey AI and ML approaches in CFD, with attention to turbulence modelling, multi-scale methods, and AI-enhanced solvers like physics-informed neural networks (PINNs). Guo et al. [34] focus on applying ML to predict the complex behaviour of nanofluids, emphasising the nonlinear relationships among their properties and the limitations of conventional experimental and numerical methods, and proposing ML models

TABLE 3. Overview of Machine Learning Approaches Applied in Recent Literature and Addressed Thermodynamic Problem.

Method	Thermodynamic prediction problem	AI Model	AI Area
[37]–[42]	Properties of gases and supercritical fluids. Vapour-liquid equilibrium behaviour of binary mixtures. Stable phases and crystal structures of inorganic compounds. Phase diagrams. Self-diffusion coefficients in dense fluids.	Random forest	Machine learning
[38], [40]	Properties of gases. Self-diffusion coefficients in dense fluids.	Gradient Boosting Regression	
[38]	Properties of gases.	AdaBoost, XGBoost, Ridge Regression	
[38], [41], [43]–[45]	Properties of fluids and supercritical fluids.	Support Vector Regression, Gaussian Process Regression	
[39]	Phase diagrams.	Machine-learned potentials	
[46], [47]	Isobaric heat capacity of natural gases. Performance of organic Rankine cycle (ORC) based on the thermodynamic properties of working fluids.	Genetic algorithms	
[40], [41]	Properties of supercritical fluids. Self-diffusion coefficients in dense fluids.	Decision Tree, K-nearest neighbors	
[43]	Properties of fluids.	Multiple linear regression	
[48]	Equation of state models for fluid systems.	Symbolic regression	
[38], [41]–[43], [45]–[47], [49]–[64]	Properties of fluids and supecritical fluids. Gas compressibility factor. Isothermal flash calculations. Activity coefficient of binary mixtures. Phase equilibria. Vapour pressure of pure compounds. Phase diagrams. Vaporization properties of pure substances. Isobaric heat capacity of natural gases. Performance of organic Rankine cycle based on the thermodynamic properties of working fluids. vapour-liquid equilibrium of binary mixtures. Equations of state of noble gases and molecular liquids.	Feed-forward neural networks	Deep learning
[41], [50], [65]–[70]	Gas compressibility factor. Properties of supercritical fluids. Speed up the calculations of thermodynamic properties of real gases. Properties of biogas. Boyle temperature of pure gases. Properties of fluids in vapour-liquid equilibrium. Vapor-liquid equilibrium in binary systems.	Multilayer perceptrons	
[38], [71]	Properties of gases. Flow parameters in complex thermodynamic machinery systems.	Physics-Informed Neural Networks	
[38]	Properties of gases.	Graph convolutional neural network	
[72]	Speed up flash calculation.	Mixture of Experts	
[37]	Stable phases and crystal structures of inorganic compounds.	Generative Autoencoders	
[54]	Activity coefficient of binary mixtures.	Message passing neural networks	
[65]	Speed up the calculations of properties of real gases.	Residual neural network (ResNet)	
[73]	Phase diagrams.	Transformers	

TABLE 4. Thermodynamics Papers and Related Focus.

Method	Gas Properties	Liquid Properties	Supercritical Fluids	Refrigerants	Vapour-liquid Equilibrium	Phase Diagrams	High pressure Systems
[40], [63]	✓	✓	✓	-	-	-	-
[42], [72]	✓	✓	-	-	✓	✓	
[60], [61], [69]	✓	✓	-	✓	✓	-	-
[44]	-	✓	-	-	✓	✓	✓
[62]	-	✓	-	✓	✓	-	
[54], [70]	✓	✓	-	-	✓	-	-
[38], [68]	✓	-	-	-	-	✓	-
[41]	-	-	✓	-	-	✓	✓
[46], [49], [50], [64], [66], [67], [71]	✓	-	-	-	-	-	✓
[37], [39], [43], [55]	✓	✓	-	-	-	✓	-
[47]	-	✓	-	-	✓	✓	-
[59], [74]	✓	-	-	-	✓	-	-
[73]	-	✓	-	-	-	✓	-
[58]	-	✓	-	-	-	✓	✓
[57], [75]–[77]	✓	-	-	-	✓	✓	✓
[65]	✓	-	✓	-	-	-	✓
[52], [53], [56]	-	✓	-	-	✓	-	-
[78]	✓	✓	✓	-	✓	-	✓
[79]	✓	-	✓	-	✓	-	✓
[48]	✓	-	✓	-	-	✓	-
[80]	✓	✓	✓	-	✓	✓	✓
[51]	✓	-	-	-	✓	✓	-
[81]	✓	-	-	✓	✓	✓	-
[82]	-	-	✓	✓	-	✓	-
[45]	-	-	✓	-	✓	✓	-

combined with optimisation techniques to improve predictive accuracy and computational efficiency. Chaparro et al. [35] develop an ML-based equation of state (EoS) for Mie fluids, ensuring thermodynamic consistency by incorporating physical constraints into artificial neural networks trained on molecular simulation data. Lastly, Jayaprakash et al. [36] investigate accelerating critical point calculations in fluid mixtures using deep neural networks (DNNs), integrating ML into numerical solvers such as Damped Newton-Raphson (DNR) and Differential Evolution (DE) algorithms to reduce computational time.

Our work provides a comprehensive overview of ML applications specifically in fluid thermodynamics, going beyond the narrow focus on CFD or specific fluid types. It covers the prediction of gas and liquid properties, vapour-liquid equilibria, phase diagrams, supercritical fluids, and industrial applications such as turbomachinery and energy systems, leveraging datasets like GERG2008. Furthermore, our survey systematically compares a broad range of ML models, including support vector regression, NNs, DL architectures, symbolic regression, and PINNs, with an emphasis on their performance in predicting thermodynamic properties for in-

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3615454

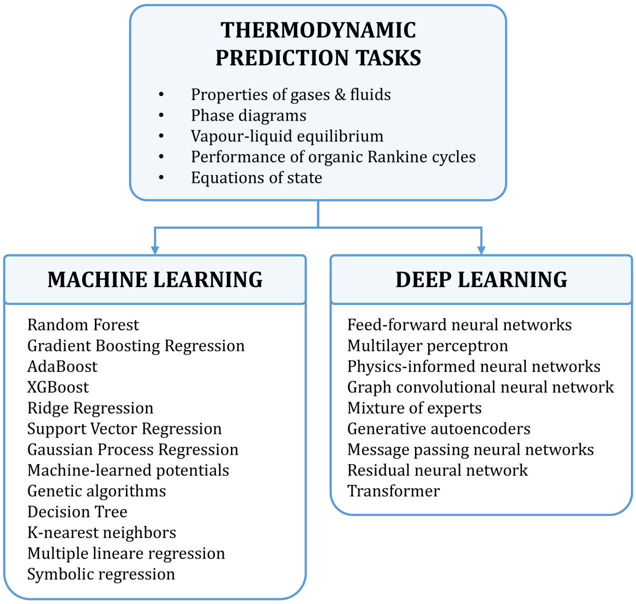


FIGURE 4. **A visual summary of the Main Thermodynamic Prediction Tasks and related ML and DL Approaches.**

dustrially relevant systems. By encompassing these diverse aspects, this work fills the gaps left by previous surveys and provides a unified perspective on the use of ML for solving complex fluid thermodynamic problems.

In Table 3, we present an overview of the methods adopted in the different papers analysed in this survey, together with the addressed thermodynamic problem. The table categorises these studies by the AI Area (ML or DL), specific AI Model used, and the Thermodynamic prediction problem investigated. Within the ML domain, a diverse range of models has been employed. RF, GB, AdaBoost, XGBoost, and Ridge Regression are frequently used for predicting properties of gases and supercritical fluids, as well as for vapour-liquid equilibrium and diffusion coefficients. SVR and GPR have also shown promise in predicting fluid and supercritical fluid properties. For specialised tasks like phase diagram prediction, Machine-learned potentials have been utilised. On the other hand, Genetic algorithms have been applied to determine isobaric heat capacity and evaluate the performance of organic Rankine cycles. Simpler models like Decision Tree, K-nearest neighbours (KNN), and Multiple linear regression still find applications for predicting properties of supercritical fluids, self-diffusion coefficients, and general fluid properties. Last but not least, Symbolic regression has been explored for developing equation of state models for fluid systems. The DL paradigm has seen a significant surge in applications. FFNNs are particularly prevalent, addressing a broad spectrum of problems including properties of fluids and supercritical fluids, gas compressibility factors, phase equilibria, and vaporisation properties of pure substances. MLPs are also widely used, especially for gas compressibility factor prediction and speeding up thermodynamic property calculations for real gases. More advanced DL architectures, such as PINNs and GCNNs are emerging for predicting gas properties and flow parameters in complex systems. Other worth mentioning DL models include Mixture of Experts for accelerating flash calculations, GAEs for predicting stable phases and crystal structures, Message passing neural networks (MPNNs) for activity coefficients of binary mixtures, Residual neural networks (ResNet) for speeding up calculations of real gas properties, and Transformers for phase diagrams. For further clarity, we also provide a visual summary of the main thermodynamic prediction tasks and related AI approaches in Figure 4.

Overall, Table 3 highlights a clear trend toward the increasing adoption of sophisticated AI models, particularly within DL, to tackle complex and diverse problems in thermodynamics. The latter are briefly summarised in Table 4, where we report the main thermodynamic frameworks addressed in recent ML papers. Among others, some topics emerge as central research themes, such as phase transitions and vapour-liquid equilibrium, indicating potentially promising areas for further exploration. We clearly recognise the vast potential scope of ML models for solving a variety of thermodynamic problems.

In thermodynamics, phase transitions are crucial for understanding how substances change state across solid, liquid, and gas phases. In the context of turbomachinery, phase transitions play a vital role in understanding fluid behaviour across components, such as turbines, boilers, condensers, and heat exchangers, where fluids may change states. These changes are often depicted on diagrams that illustrate temperature, pressure, and volume conditions under which different phases coexist in equilibrium. Specifically, phase diagrams are graphical representations that show the different phases of a substance (e.g. solid, liquid, gas) as a function of variables such as temperature, pressure, and sometimes chemical composition (for mixtures). They describe regions of phase stability and the phase transitions that occur between them. For a pure substance, a phase diagram typically plots temperature on the x-axis and pressure on the y-axis, with three regions corresponding to solid, liquid, and gas. The curves separating these regions represent equilibrium between two phases, such as the melting curve (solid-liquid) or the vaporization curve (liquid-gas). Key points on the diagram include the triple point, where all three phases coexist, and the critical point, beyond which the distinction between liquid and gas disappears, a region where supercritical behaviour applies. For binary mixtures, phase diagrams often show temperature vs. composition, where two curves indicate the boundaries between liquid, vapour, and a region where both phases coexist. Ternary systems are represented with triangular diagrams that show how three components mix, with lines indicating phase equilibrium. For mixtures, binary and ternary phase diagrams provide insight into phase boundaries and equilibrium behaviour, useful for multi-component fluids common in industrial processes. In all cases, such tools help engineers and plant designers understand, predict, and optimize fluid behaviour in the conditions encountered in turbomachinery applications. In Fig. 5, a generic phase diagram is presented, illustrating the relationship between pressure and temperature for a given substance. The diagram highlights the triple point, where the solid, liquid, and vapour phases coexist in equilibrium, and the critical point, which represents the end of the liquid-vapour equilibrium curve. The solid green, blue, and red lines delineate the phase boundaries between the different states of matter. The region beyond the critical pressure and temperature is labelled as "supercritical," where the fluid exhibits properties of both liquid and gas phases.

In the following section, we will analyse the main topics and provide details of the different works. In Sect. III-A, we first review the main papers adopting ML architectures for data-driven estimation of phase transition properties. After describing the largest set of papers devoted to phase transitions and vapour-liquid equilibrium estimation, we will focus on ML solutions for the analysis of two relevant cases, namely gas mixtures in Sect. III-B and both subcritical and supercritical fluids in Sect. III-C. Supercritical fluids are above their critical temperature and pressure, where distinct liquid and gas phases no longer exist, and exhibit properties of both liquids and gases, such as the ability to diffuse like a gas while dissolving materials like a liquid.

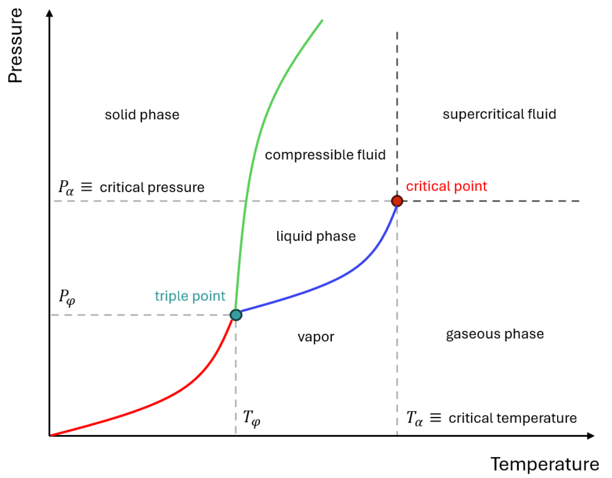


FIGURE 5. **Generic phase diagram illustrating the relationship between pressure and temperature for a substance. The diagram highlights the triple point, where solid, liquid, and vapour phases coexist in equilibrium, represented as a cyan-coloured point in the image, and the critical point, marking the end of the liquid-vapour equilibrium curve, shown as a red-coloured point. The solid green, blue, and red lines represent the phase boundaries between the different states of matter. The area beyond both critical pressure and temperature is labelled as "supercritical", where the fluid exhibits properties from liquid and gas phases. Diagram adapted from general concepts about phase diagrams, as described in [83] or, for particular applications in [84], [85].**

A. PHASE TRANSITION PROBLEMS

Considering phase transition challenges, [42] develops artificial neural networks (ANNs) and RF models to predict vapour-liquid equilibrium in binary mixtures using a new dataset of 4562 experimental data points for 210 mixtures. The RF model outperforms ANNs in predicting vapour fractions based on macroscopic and microscopic features. SHapley Additive exPlanations (SHAP) analysis identifies boiling points and critical temperatures as the most influential factors for vapour-liquid equilibrium prediction. In [70], ANNs predict vapour-liquid equilibrium data for binary systems like ammonia-water and methane-ethane, achieving errors below 1%. Training uses backpropagation on pressure and temperature inputs to predict molar fractions, with better performance when trained on either P-T-x or P-T-y surfaces individually. The work in [72] introduces a thermodynamically consistent NN-based equation of state predicting residual Gibbs free energy, enabling derivation of other properties by differentiation. The network, pretrained on analytical equations of state and fine-tuned on experimental data, accelerates fugacity coefficients calculations up to 400-fold via GPU-based flash calculations, offering superior speed and accuracy. Finally, [59] estimates vaporisation properties of 62 pure substances using NNs optimised with the Levenberg-Marquardt algorithm. Networks with two hidden layers of 40 neurons each outperform traditional models in predicting properties such as volume, internal energy, enthalpy, entropy of vaporisation, and saturation conditions.

Although it is beyond the scope of this survey, some materials science-focused works have been included to explore interdisciplinary insights, particularly how ML advances in materials design and characterization can contribute to more precise and efficient modelling of fluid properties and phase behaviours. In [73], a Transformer-based Deep Neural Network (DNN) accelerates thermodynamic equilibrium calculations and predicts aluminium alloy phase diagrams over 100 times faster than the Calculation of Phase Diagrams (CALPHAD) method, while maintaining high accuracy. Trained on 841,689 data points generated from CALPHAD and experimental measurements, the model predicts phase fractions using cross-entropy and Mean Squared Logarithmic Error (MSLE) losses to better capture discontinuous phase changes. The study in [58] develops an ANN for the interatomic potentials of rubidium under high pressure, trained on molecular dynamics and *ab initio* Density Functional Theory (DFT) calculations of atomic energies and forces. The model accurately reproduces the melting curve, including the temperature maximum, and partially captures the liquid-liquid transition, although limitations remain at high pressures. In [39], the thermodynamic stability of materials and phase transitions is addressed. It offers a comprehensive overview of computational techniques for determining phase diagrams, with a particular focus on integrating traditional methods and ML techniques to predict phase behaviour from existing data, improving the accuracy and speed of phase diagram predictions and detailed molecular simulations. Desgranges et al. [57] also focus on fast and accurate predictions by training ANNs on Expanded Wang-Landau simulations to estimate thermodynamic properties (i.e. partition function, Gibbs and Helmholtz free energies, entropy) of fluids like argon, CO_2, and water across broad temperature and pressure ranges. On the other hand, Tarjomannejad et al. [56] combine a Feed-Forward Neural Network (FFNN) with a group contribution method to predict vapour pressures of 100 pure compounds from molecular and thermodynamic properties such as temperature, critical pressure, critical temperature, and acentric factor. The network achieves an average relative deviation of $\sim$1.18%, outperforming traditional equations of state (PR, SRK, PT), although larger errors occur for alcohols and amines.

Leaving materials science studies aside, [55] employs a NN to predict Gibbs energies, entropies, enthalpies, and heat capacities for unary and binary systems, applying Gibbs energy minimization for equilibrium compositions. However, NNs struggle with systems showing inflexion points in Gibbs energy curves, limiting their applicability in complex scenarios. [38] presents ThermoLearn, a Physics-Informed Neural Network (PINN) predicting Gibbs free energy, total energy, and entropy using composition and crystalline features. Trained on NIST-JANAF and PhononDB datasets, it achieves 43% higher accuracy than XGBoost and standard NNs, especially on out-of-distribution data, though performance depends on data quality and feature representation. At the molecular level, [54] proposes DeepGamma, a DL model using directed Message-Passing Neural Networks (MPNNs) on molecular graphs to predict activity coefficients in binary mixtures directly from molecular structures. Trained on DFT-calculated data, DeepGamma accelerates predictions by over 1900 times compared to quantum simulations. Another work in [53] uses an FFNN to predict the density and temperature of Lennard-Jones fluids from their radial distribution functions. Trained over wide ranges, it attains $\sim$3% average error but shows higher errors near phase boundaries, indicating limitations for extreme states. In [51], thermodynamic equilibrium calculations are accelerated by DL techniques that replace computationally intensive parts, reducing execution time, especially in complex multiphase systems. Thermodynamic algorithms are rewritten and vectorised for GPU parallelisation, and DNNs substitute expensive isothermal flash subroutines. However, challenges remain for scenarios with strong nonlinearities or near-critical points. The study in [37] employs GAEs and RF to predict stable phases and crystal structures of inorganic compounds based on composition, temperature, and pressure data from National Institute of Standard and Technology (NIST) and structural information from the Inorganic Crystal Structure Database. Crystal structures are predicted via classification of Bravais lattices, structure types, and polyhedra, achieving accurate estimations except for underrepresented cases like triclinic lattices.

Notably, a complementary approach in [86] introduces a thermodynamic theory for ML systems, interpreting model training as a phase transition and defining temperature through potential energy and entropy. This framework treats ML systems as complex thermal engines, with global and local temperatures influenced by NN activation functions, enabling deeper insights into the efficiency and performance of NN operations.

B. GAS MIXTURES CHARACTERIZATION

In the field of gas mixtures, [68] employs a multilayer perceptron (MLP) with two hidden layers (24 neurons each) trained on 1,663 experimental data points of reduced temperature, critical molar volume, and the second virial coefficient to predict the Boyle temperature of 16 pure gases. The Boyle temperature, marking ideal gas-like behaviour with unit compressibility, is challenging to measure experimentally. The network, optimised over 10,000 iterations, achieves an RMSE of $2.74\text{cm}^3/\text{mol}$, outperforming traditional correlations and accurately determining Boyle temperatures with minimal deviation from reference values.

In [67], an MLP estimates the compressibility factor of natural gas using temperature, pressure, and Joule-Thomson coefficient as inputs, without requiring gas composition. Trained on 30,000 datasets generated with the GERG2008 equation of state and validated experimentally, the model enables accurate, real-time applications at measurement stations, reducing costs compared to compositional analyses, although accuracy declines near the critical temperature. Similarly, [66] uses an MLP trained on 4,000 biogas samples (75% training, 15% validation, 10% testing) to predict thermodynamic properties (i.e. including density, compressibility, heat capacity, internal energy, and speed of sound) based on pressure, temperature, and molar composition. The Levenberg-Marquardt-trained network shows excellent agreement with experimental and GERG2008-calculated values, offering high accuracy across diverse conditions while reducing computation times and costs compared to traditional methods, thus proving suitable for rapid biogas property estimation in distribution networks.

The work in [46] proposes a novel empirical model using ANNs to estimate the isobaric heat capacity of natural gas based on specific gravity, temperature, and pressure, without requiring gas composition. The NN architecture, optimised via genetic algorithms and particle swarm optimisation, was trained on 847 data sets and validated experimentally, achieving an R^2 of 0.99691 and low RMSE. Particle swarm optimisation notably reduced processing times compared to genetic algorithms. In [65], deep residual learning is applied to model thermodynamic properties of real gases for Computational Fluid Dynamics (CFD) simulations, greatly accelerating calculations versus cubic equations of state. Three ANNs were compared: single hidden layer, five hidden layers, and ResNet. Trained on 32,000 points from Peng-Robinson EoS simulations of nitrogen, ResNet achieved the highest accuracy and computation speed (up to 166 times faster), outperforming simpler networks, especially in predicting complex properties like specific heat capacity. In [71], a physics-informed DL model predicts real-time flow parameters in thermodynamic machines (e.g., aircraft turbofans) by integrating physical laws into LSTM-based NNs for subsystems such as compressors and turbines. Trained on experimental and flight data, the model accurately predicted parameters like exhaust gas temperature and pressure under steady and unstable conditions, outperforming traditional simulations.

However, accuracy decreased under unstable conditions or low engine speeds due to sensor fluctuations and model latency.

The study in [50] employs an MLP trained on 4,158 Standing and Katz data points to predict the compressibility factor of natural gases from reduced temperature and pressure. The ANN achieves significantly lower RMSE and faster convergence than traditional equation-of-state models, though extreme scenarios with complex thermodynamics may require refinement. In [49], a new ANN-based correlation improves gas deviation factor predictions in high-temperature, high-pressure reservoirs, outperforming traditional models with a MAE below 3% in many high-pressure conditions. The network, trained on over 3,000 experimental points, uses a single hidden layer of 10 neurons with sigmoid activation and a linear output, reducing computation time while maintaining high accuracy; however, complex gas compositions may still challenge the model.

Notably, [52] presents an ANN as an educational tool to estimate the specific enthalpy of water vapour from temperature and pressure data generated via the IAPWS IF-97 standard. The network, trained on 97% of the dataset, predicts enthalpy with mean errors under 5%, but shows reduced accuracy in region 3 due to sparse training data, emphasising the need for better data coverage in complex thermodynamic regions.

C. FLUIDS-RELATED PROBLEMS

Fluid thermodynamic problems solved by ML involve predicting properties of fluids in different states under varying temperature and pressure. Most studies focus on subcritical fluids, which exhibit distinct liquid and gas phases with phase transitions, while fewer address supercritical fluids, which combine liquid-like density and gas-like diffusivity beyond the critical point. For example, [63] uses ANNs to predict equations of state (pressure and internal energy) of noble gases and molecular liquids from their radial distribution functions, training on Lennard-Jones simulations and validating with experiments. Predictions show high accuracy, with relative errors below 1%, though slight overestimation occurs at high densities and temperatures. Similarly, [44] employs SVR to estimate thermodynamic properties of pure fluids and mixtures, training on a high-fidelity database with molecular dynamics simulations. After normalization and cross-validated grid search for optimal parameters, the model achieves very low MSE and high computational efficiency.

Supercritical fluids are discussed in [41], where ML is applied to estimate thermodynamic properties like solubility, optimise chemical reactions, support material synthesis, simulate fluid dynamics, and enhance CO_2 geo-extraction. For example, in Carbon Capture, Usage and Storage (CCUS), accurate predictions of supercritical CO_2 properties are critical to leveraging its high density and solubility for effective geological sequestration, reducing atmospheric CO_2.

The review in [41] covers ML methods including non-linear regression, ANNs, and SVR for predicting supercritical fluid properties and optimising related processes, with

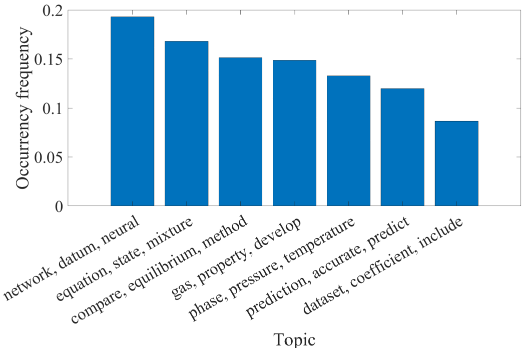


FIGURE 6. Most frequent topics appearing in the papers.

practical case studies highlighting ML's potential for accurate simulation of complex systems.

In [43], ML supports molecular thermodynamics by predicting complex thermodynamic properties, integrating high-throughput simulations for material discovery, and improving molecular force fields combining quantum mechanics and atomic-level simulations. However, model accuracy relies heavily on data quality and appropriate molecular descriptors.

The study in [47] explores using ANNs to predict the performance of the organic Rankine cycle, a thermodynamic process for converting low-temperature heat into electricity with organic fluids. Input features include critical temperature, pressure, acentric factor, and ideal gas heat capacity. Models are trained on data for 106 fluids generated with NIST's REFPROP, which provides accurate thermodynamic properties. The ML models enable precise prediction of basic and regenerative cycles, especially with dry or isentropic fluids.

In [45], ML replaces traditional equations of state like SAFT-VR Mie to model properties such as density, viscosity, and specific heat. Using ANNs and GPR, the models predict critical properties and vapour pressures from pseudodata and molecular descriptors. Results show high accuracy and generalisation within the dataset range, with additional data required near its boundaries.

To sum up, several computational architectures are studied to meet the needs for accurate and fast thermodynamic problem solutions. For the sake of compactness, Fig. 6 provides a summary of the main topics appearing in the analysed papers together with their relative occurrence frequency. In the next Section, we will consider the case of gases and fluids in detail.

D. INDUSTRIAL APPLICATIONS

ML is increasingly being adopted across various industrial sectors to address complex fluid thermodynamics problems, offering significant improvements in efficiency, design, and operational optimization. In turbomachinery, for example, ANNs are being explored for design automation, demonstrating their suitability to define design variables and achieve comparable performance to more computationally demanding methods like CFD [87]. ML-based control systems are also emerging for turbomachinery, emphasizing predictive maintenance, anomaly detection, and real-time optimization to enhance performance, reliability, and efficiency [88]. The transformative impact of AI extends to accelerating materials development and co-optimization of materials and part performance, crucial for advancing turbomachinery applications [89]. ML models are also being utilized in heat exchanger applications for performance prediction, design optimization, and real-time simulations [90]. In the context of waste heat recovery, ML is extensively applied to optimize Organic Rankine Cycle (ORC) systems. This includes using AI to unlock energy savings from waste heat in water treatment plants [91], employing ML for multi-objective optimization of ORC systems [92], and developing ML-based fault detection and diagnosis schemes for ORC systems to ensure operational reliability and efficiency [93]. These industrial applications provide a first overview of the transformative potential of ML in addressing real-world thermodynamic challenges and driving advancements in various engineering domains.

IV. NUMERICAL FRAMEWORKS FOR MACHINE LEARNING IN THERMODYNAMICS

The above-described literature covers a variety of thermodynamic problems, solved using different ML approaches. Herein, we provide an overview of the simulation settings adopted, from different points of view. Firstly, we describe the important GERG2008 framework, which provides a basis for data generation and validation in a number of scientific contributions. Then, we extend our analysis to the main datasets used for training and testing purposes. While we aim to provide a unified overview of the main performance metrics adopted to assess the ML model performances, a summary of the main equations of state models encountered in the surveyed literature, to represent a wide range of thermodynamic, physico-chemical, and molecular properties, is reported in Table 5.

A. THE GERG2008 MODEL

The GERG2008 framework is a thermodynamic model describing natural gases and mixtures across a broad range of conditions, using a multi-parameter equation of state for accurate calculations of pressure, density, enthalpy, entropy, and phase equilibria in pure gases and complex mixtures. Covering temperatures from 60 K to 700 K and pressures up to 70 MPa, GERG2008 is particularly suited for applications in natural gas and LNG processing.

The model employs Helmholtz free energy formulations for hydrocarbons, inert gases, and polar substances (e.g., CO_2, H_2S), combined with advanced mixing rules to capture binary and higher-order interactions in non-ideal mixtures. Based on extensive experimental data, it has been optimised for pipeline flow, LNG, and phase equilibrium calculations, adopted as an ISO standard (ISO 2076-2) [94], and serves as a reference model for accurate property predictions. Comparisons with Peng-Robinson are discussed in [74], and detailed

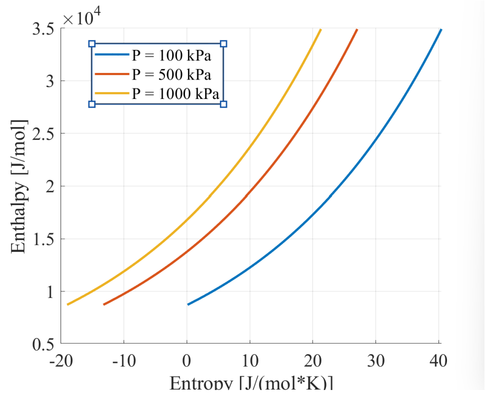


FIGURE 7. Example of enthalpy-entropy diagram for an ideal gas by calculating enthalpy and entropy using the GERG2008 model over a range of temperatures and pressures.

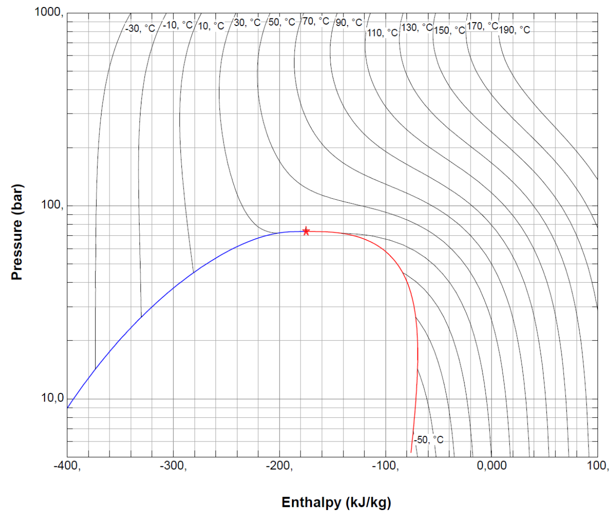


FIGURE 8. Example of pressure-enthalpy diagram for a pure compound by calculating pressure and enthalpy using the GERG2008 model over a range of temperatures.

studies on compressibility and Joule-Thomson factors can be found in [80], [95], [96].

Fig. 7 shows an enthalpy-entropy diagram for an ideal gas, with enthalpy modelled as $H = C_p \cdot T$ and entropy as $S = C_p \cdot \ln(T/T_0) - R \cdot \ln(P/P_0)$, illustrating variations over temperature and pressure ranges.

Fig. 8 depicts a pressure-enthalpy diagram for CO_2, highlighting phase boundaries, the Saturated Liquid (blue) and Saturated Vapour (red) lines, forming the Condensation Bell, essential for sizing heat exchange systems. The critical point is marked with a red star, beyond which supercritical behaviour occurs. Black iso-temperature lines trace temperature evolution across the diagram.

The GERG2008 equation of state, presented in [80], models the thermodynamic properties of natural gases and mixtures of 21 components over wide ranges of temperature, pressure, and composition, accurately covering gas, liquid, and supercritical regions. It relies on experimental data for pure components and binary mixtures, employing multi-fluid approximations to optimize structure and coefficients. All thermodynamic properties are expressed via the Helmholtz free energy, split into ideal and residual contributions:

$$A(T,\rho,\mathbf{x}) = A^{\text{ideal}}(T,\rho,\mathbf{x}) + A^{\text{residual}}(T,\rho,\mathbf{x})$$

where: $A^{\text{ideal}}(T,\rho,\mathbf{x})$ represents the ideal gas part of the Helmholtz free energy, $A^{\text{residual}}(T,\rho,\mathbf{x})$ accounts for the interactions between molecules (non-ideal part), T is the temperature, ρ is the molar density, $\mathbf{x}$ is the composition vector (mole fractions of the gas components). From the Helmholtz energy function, the main thermodynamic properties can be derived from A as $P = \rho^2 \frac{\partial A}{\partial \rho}$, $U = A + T\frac{\partial A}{\partial T}$, $S = -\frac{\partial A}{\partial T}$, $H = U + PV$.

GERG2008 incorporates complex temperature- and density-dependent terms in the residual part and uses interaction parameters between components. Validated against over 125,000 experimental data points, it ensures high accuracy across applications in natural gas processing, LNG, and energy production. It surpasses previous EoS models in component range and operating conditions but remains limited by the quality of experimental data, particularly for mixtures containing secondary components like n-nonane or hydrogen sulfide. GERG2008 is extensively used for vapor-liquid equilibrium predictions, critical point determination, and thermophysical property calculations, serving as a reference model in both industry and research.

In [95], GERG2008 demonstrates superior accuracy over the traditional PR model in predicting the compressibility factor (Z) at 40°C, showing excellent agreement with experimental data across all pressures, while PR shows significant discrepancies. The GERG2008 model is therefore adopted as a framework to develop in a number of ML-based papers. In the following, we describe these papers to provide an idea of the manifold applications of the model.

The work in [69] employs MLP neural networks trained with thermodynamic data generated using REFPROP 6.0 (240–340 K, up to 20 MPa) to predict Z as a function of temperature and pressure for refrigerants like R-12 and R-134a. Properties such as enthalpy and entropy are calculated by numerical derivatives. Optimal models, selected via the Wilcoxon test, achieve R^2 values close to 1 and low MSE, outperforming traditional equations of state except near the critical point [62].

In [61], NNs trained on REFPROP-generated PVT data predict enthalpy, entropy, heat capacity, expansion coefficients, and speed of sound for six refrigerants. The models show low relative deviations, but accuracy decreases near the critical point, especially for heat capacity and speed of sound. Another study in [60] uses ANNs with the scaled conjugate gradient and Levenberg-Marquardt algorithms to predict volume, enthalpy, and entropy of R508b in saturated and superheated vapour regions, achieving R^2 values near 1, with the Levenberg-Marquardt method performing best.

A recent study in [40] applies GB, k-Nearest Neighbors (kNN), Decision Tree, and RF algorithms to predict self-diffusion coefficients of dense fluids, trained on 7,931 experimental points from 223 substances. The GB model with recursive feature elimination reduces relevant variables to 5–8 and surpasses phenomenological models like Zhu et al. [97] and Lennard-Jones equations, without requiring substance-specific parameters.

In another recent study [81], DNNs estimate the vapour-liquid equilibrium of binary systems using pure component parameters and temperature as inputs. The model, trained on experimental binary refrigerant data, predicts cross-interaction parameters for the Peng-Robinson EoS with van der Waals mixing, achieving errors below 3.5% for equilibrium pressures and vapour mole fractions, though further refinement is needed for complex systems.

In [82], an end-to-end ML approach using K-nearest neighbours regression predicts thermodynamic properties of refrigerant R134-a across subcooled, superheated, and supercritical regions, based on augmented experimental data from the NIST database. The model demonstrates high speed and accuracy, though precision decreases near phase transitions where properties change rapidly. In contrast, [75] computes phase equilibria for multi-component mixtures using multiparameter Helmholtz energy equations of state, enhancing accuracy over cubic EoS. Algorithms adapted from Michelsen minimise the tangent plane distance function and solve for densities, enabling accurate and stable isothermal and isenthalpic flash calculations. However, Helmholtz models introduce computational challenges requiring robust solvers. Another study [78] integrates the GERG2008 equation of state for natural gas mixtures into simulation tools via the CAPE-OPEN standard, improving accuracy in properties like density and heat capacity compared to Peng-Robinson and Redlich-Kwong-Soave models, especially in the liquid phase. The approach enables precise simulations in common environments (e.g., Aspen Plus, ProSim, UniSim) but depends on CAPE-OPEN compatibility, which may limit broader adoption.

The goal of [76] is to use Monte Carlo simulations in the Gibbs ensemble to accurately compute critical parameters and phase equilibria of industrially relevant gases and mixtures, enhancing understanding of complex phase behaviour. The method employs the Transferable Potentials for Phase Equilibria force field, applicable across diverse compounds to predict properties like phase equilibria and adsorption by modelling intermolecular interactions. Accuracy is assessed by varying cutoff radii and applying tail corrections, with results compared to experimental data and equations of state (Peng-Robinson, GERG2008), showing good agreement, especially for CO_2 and CH_4. The study in [79] determines critical points of complex natural gas mixtures containing linear alkanes, N_2, and CO_2, using GERG2008 combined with Helmholtz energy models. It applies the Heidemann and Khalil method, solving non-linear equations involving second and third derivatives of Helmholtz energy. Stability checks and comparisons with Peng-Robinson and SAFT models are performed. GERG2008 predicts critical temperatures and pressures accurately for most binary mixtures with hydrocarbons, N_2, and CO_2, outperforming Peng-Robinson in many

TABLE 5. Main Equation of State Models in the Surveyed Literature.

Equation of State Model	Description
GERG 2008 [80]	The GERG-2008 model is an advanced equation of state used to predict thermodynamic properties of natural gas mixtures. It accurately models gases, including hydrocarbons, nitrogen, carbon dioxide, and hydrogen, across a wide range of temperatures and pressures. GERG-2008 incorporates detailed interactions between gas components, making it suitable for complex gas systems. It is widely used in the natural gas industry for applications like custody transfer and pipeline transportation. The model provides more precise results than traditional cubic equations of state, especially for non-ideal gas mixtures.
Peng-Robinson [98]	The Peng-Robinson equation of state is a widely used model for predicting the behavior of real gases. It improves upon the van der Waals equation by incorporating temperature-dependent parameters and accounting for attractive forces between molecules. This model accurately predicts phase behavior, especially for hydrocarbons and non-polar substances, over a wide range of conditions. It is frequently used in chemical and petroleum engineering to model vapor-liquid equilibria. The Peng-Robinson equation is known for balancing simplicity and accuracy in industrial applications.
van der Waals [99]	The van der Waals equation of state is an early model that improves the ideal gas law by accounting for molecular volume and intermolecular forces. It introduces two parameters: one for the volume occupied by gas molecules and another for attractive forces between them. This equation provides a better description of real gas behavior, especially near condensation points, compared to the ideal gas law. However, it is less accurate at high pressures and temperatures. It is mainly used for qualitative insights into gas-liquid transitions and real gas behavior.
Redlich-Kwong [100]	The Redlich-Kwong equation of state is an early improvement on the van der Waals equation, designed to better predict the behavior of real gases. It introduces a temperature-dependent correction to the attractive term, enhancing its accuracy, particularly for gases at moderately high pressures and temperatures. While less accurate at very high pressures or near the critical point, it is simpler and more computationally efficient than later models. The Redlich-Kwong equation is commonly used in engineering applications for non-ideal gas behavior. It laid the groundwork for further refined equations of state like Peng-Robinson.
Redlich-Kwong-Soave [101]	The Redlich-Kwong-Soave equation of state is a modified version of the Redlich-Kwong equation, incorporating a temperature-dependent correction factor to improve accuracy. It is primarily used to model the behavior of real gases, especially non-polar and weakly polar compounds. The Soave modification enhances the prediction of liquid-phase properties and vapor-liquid equilibria. It is commonly applied in the petroleum and chemical industries for processes like gas separation and distillation. The equation provides better accuracy than simpler models for a wide range of temperatures and pressures.
SAFT [102]	The SAFT (Statistical Associating Fluid Theory) model is an advanced equation of state used to predict the thermodynamic properties of complex fluids, including phase behavior, densities, and equilibria. It is based on statistical mechanics and models fluids by considering molecules as chains of interacting segments, which can associate or bond with each other. The SAFT model incorporates molecular shape, size, and association effects, making it particularly useful for systems with hydrogen bonding, such as water, alcohols, and polymers. It is widely applied in chemical engineering for the design of processes involving complex fluids and mixtures.

cases, though larger errors arise when experimental data or interaction parameters are lacking.

In [74], GERG2008 and Peng-Robinson equations of state are compared for predicting thermodynamic properties of natural gas mixtures. The study assesses pressure, density, and vapour-liquid equilibria predictions for binary and multicomponent systems, highlighting GERG2008's multi-fluid Helmholtz model versus Peng-Robinson's cubic equation. GERG2008 shows superior accuracy, especially for liquid phases (0.1–5% deviations), and better vapour-liquid equilibrium predictions. However, it incurs 2–6 times higher computational cost due to iterative molar density calculations.

The study [77] compares the GERG-2008 EoS with simpler cubic and non-cubic models (SRK, PR, Soave-BWR, PC-SAFT) for predicting phase equilibria (bubble point pressure, vapour composition) and physical properties (density, compressibility) of natural gas mixtures across wide temperature and pressure ranges. GERG2008 shows superior accuracy, especially under high-pressure and temperature conditions, but underperforms in predicting bubble point pressures for heavy asymmetric mixtures (e.g., n-butane, n-nonane), where Soave-BWR provides better results. The work [48] presents a proof of concept for rapidly developing EoS by integrating symbolic regression (a ML technique), molecular dynamics simulations, and optimal experimental design. This approach enables accurate models with significantly fewer experiments, reducing cost and time compared to traditional methods, though accuracy depends on the initial simulation and data quality. Finally, the study [64] adopts ANNs to estimate the compressibility factor of natural gas for precise volumetric flow calculations from temperature, pressure, heating value, and optionally density, without prior gas composition analysis. Trained on 50,000 samples from GERG2008 and validated experimentally, the networks achieve high accuracy, particularly with four inputs, with minor accuracy loss at extreme conditions.

B. DATASETS

The collection of datasets appearing in Table 7 encompasses a wide range of thermodynamic, physico-chemical, and molecular properties for various substances and mixtures, as reported in the literature. These include the vaporisation properties of pure fluids, vapour-liquid equilibrium data for binary mixtures, and diffusion coefficients for liquids and gases, all obtained under diverse temperature and pressure conditions. Additionally, data on the condensed phases of pure elements, thermophysical properties of refrigerants, and phase diagrams for inorganic compounds are provided. Some datasets focus on energy-related metrics, such as solvation energy and activity coefficients, offering valuable resources for studying material properties and molecular behaviour across different phases and systems. While the use of established datasets and EoS is fundamental for developing ML models in thermodynamics, a critical examination reveals significant limitations that present challenges and opportunities for future research. The datasets frequently referenced, such as those from the NIST and the GERG-2008, are invaluable but not without their shortcomings. A major limitation of these resources is their fluid-specific coverage. For instance, GERG-2008 was originally developed for natural gas mixtures and, while it has been expanded to include 21 components, its accuracy can diminish for mixtures outside this primary scope, particularly for those containing fluids at high concentrations or with significant impurities that were not the main focus of its development. This specificity means that ML models trained on these datasets may not generalise well to novel fluids or mixtures of industrial interest that are not well-represented. Furthermore, even within their intended scope, models like GERG-2008 can exhibit unphysical behaviours, such as predicting unreasonable critical curves for certain asymmetric binary mixtures, which can introduce artefacts into the training data.

The quality and uncertainty of the underlying experimental data are another critical concern. Databases like NIST's REFPROP are based on the most accurate experimental data available, but they explicitly caution that uncertainties vary considerably depending on the fluid, property, and thermodynamic state. These uncertainties are inevitably propagated into any ML model trained on this data. The scarcity of high-quality experimental data, especially for multi-component mixtures and across wide temperature and pressure ranges, is a major bottleneck. The experimental process is expensive, time-consuming, and technically challenging, leading to a sparse data landscape. This sparsity poses a serious challenge for data-hungry ML techniques, particularly DL.

To propel the field forward, several improvements in data collection and accessibility are necessary. First, there is a clear need for broader and more diverse experimental data collection. Efforts should be focused on under-represented fluids and mixtures, especially those relevant to emerging technologies like carbon capture and hydrogen energy. High-throughput experimental setups combined with automated data acquisition could help to populate these sparse regions of the thermodynamic landscape. Second, fostering data accessibility and standardisation is crucial. The development of centralised, open-access repositories that adhere to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) would be a significant step forward. These platforms should not only host the data but also provide comprehensive metadata, including detailed uncertainty quantification for each data point. This would allow researchers to make more informed decisions when selecting training data and to build more robust models that are aware of data quality. Finally, data augmentation through simulation can play a pivotal role. Leveraging molecular simulations, such as Monte Carlo and Molecular Dynamics, can generate large, internally consistent datasets. While computationally intensive, these methods can fill gaps where experimental data is lacking and help in optimising and validating EoS, thereby providing higher quality data for training advanced ML models. The integration of data from experiments, simulations, and established physical models will be key to building the next generation of predictive thermodynamic tools.

V. OPEN CHALLENGES AND FUTURE DEVELOPMENTS

The field of machine learning applied to fluid thermodynamics, while promising, is at a crucial juncture where several open challenges must be addressed to unlock its full potential. A primary and persistent challenge revolves around the inherent trade-off between the accuracy of machine learning models and their computational cost. This is particularly critical when modelling complex phenomena such as critical points and phase transitions, where high precision is paramount. While DL models have shown a remarkable ability to capture the complex, non-linear behaviour of thermodynamic systems, their predictive power under extreme conditions remains an area of active investigation and requires significant improvement. To address these limitations, the integration of physics-based models with ML has emerged as a key area for future development. PINNs, which embed physical laws directly into the learning process, represent a significant step forward. This approach not only enhances model performance, especially in data-scarce or critical condition scenarios, but also improves data efficiency and the stability of predictions. However, the application of PINNs is not without its own set of challenges. Issues such as the proper handling of partial differential equation discretisations, the imposition of boundary conditions, and the potential for unstable training processes highlight the need for more robust NN architectures and advanced training algorithms. A rigorous mathematical framework to fully understand the capabilities and limitations of PINNs is still under development.

The quality and availability of data remain a fundamental bottleneck for the entire field. The development of robust and generalizable models is often hampered by the scarcity of extensive, high-quality, and publicly accessible datasets for training and validation. This underscores the need for a community-wide effort to establish standardised benchmarking datasets and organise competitions. Such initiatives would allow for the systematic evaluation and comparison of different machine learning methods, fostering a culture of reproducibility and accelerating progress, similar to what has been achieved in more mature fields like computer vision. Furthermore, developing models that can generalise to new substances and conditions without extensive retraining is a crucial frontier. This leads to the broader challenge of

TABLE 6. Summary of Open Challenges in Machine Learning Applied to Fluid Thermodynamics.

Challenge Category	Issues	Description
Methodological	Accuracy vs. Computational Cost	Addressing the inherent trade-off between model accuracy and computational expense, especially for complex phenomena like critical points and phase transitions.
	Predictive Power under Extreme Conditions	Improving the predictive capability of deep learning (DL) models under extreme thermodynamic conditions.
	PINN Implementation	Proper handling of partial differential equation (PDE) discretizations, imposition of boundary conditions, and managing unstable training processes in Physics-Informed Neural Networks (PINNs).
	Robust Architectures	Developing more robust neural network (NN) architectures and advanced training algorithms for PINNs.
	Mathematical Framework for PINNs	Establishing a rigorous mathematical framework to fully understand the capabilities and limitations of PINNs.
	Data Scarcity and Quality	Overcoming the scarcity of extensive, high-quality, and publicly accessible datasets for model training and validation.
Data-related	Standardization and Benchmarking	Establishing standardized benchmarking datasets and organizing community-wide competitions to foster reproducibility and accelerate progress.
	Generalization	Developing models that can generalize effectively to new substances and conditions without requiring extensive retraining.
Application-specific	Interpretability (Explainable AI)	Addressing the black-box nature of complex models and enhancing interpretability (Explainable AI) for increased trust and adoption in safety-critical engineering applications.
	Sustainable AI Development	Creating more energy-efficient and sustainable learning systems for AI, drawing inspiration from the efficiency of biological intelligence.

interpretability. As models become more complex, understanding the reasoning behind their predictions becomes vital. The black-box nature of many models is a significant barrier to their adoption in safety-critical engineering applications. Consequently, the development of Explainable AI is paramount. Recent advancements have introduced novel concepts like Thermodynamics-inspired Explainable Representations of AI (TERP), which aim to provide human-interpretable explanations by introducing concepts such as interpretation entropy. The ultimate goal is to move beyond post-hoc explanations and design models that are inherently interpretable, allowing researchers and engineers to gain clearer insights into the underlying thermodynamics and build trust in the model's outputs.

Looking ahead, the synergy between traditional physical models and data-driven ML approaches, often named hybrid modelling, holds high promise. This is especially true for navigating the challenges posed by sparse and uncertain data. Emerging areas, such as the use of reinforcement learning for real-time optimisation of thermodynamic systems, have already demonstrated the potential for significant gains in efficiency. Finally, on a more conceptual level, a future challenge involves viewing the development of AI itself through the lens of thermodynamics, aiming to create more energy-efficient and sustainable learning systems, drawing inspiration from the remarkable efficiency of biological intelligence. Addressing these multifaceted challenges will require a collaborative, interdisciplinary effort, paving the way for the next generation of intelligent thermodynamic modelling. Table 6 provides a comprehensive overview of the open challenges currently facing the application of ML in fluid thermodynamics, organised in three different groups: Methodological Challenges, Data-related Challenges, and Application-specific Challenges.

To sum up, this survey paves the way for future developments such as i) inclusion of hybrid models that combine physical principles with advanced DL for better prediction under critical conditions; ii) generalization to predict new conditions without extensive retraining; iii) introducing explainable AI to offer clearer insights of the impact of physical informed model into predictions. These developments provide a basis for introducing DL to solve thermodynamic problems for robust industrial applications.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3615454

TABLE 7: Main Datasets Adopted in the Literature for ML-based Thermodynamical Studies.

Dataset	Description	Data type	Open
[59]	The dataset consist of vaporization properties of 62 pure fluids, including haloalkanes, hydrocarbons (saturated, unsaturated, aliphatic, and aromatic), inorganic substances, and one alcohol. The vaporization properties collected include saturation temperature (T_{sat}), saturation pressure (P_{sat}), internal energy (ΔU_{vap}), enthalpy (ΔH_{vap}), entropy (ΔS_{vap}), and volume (ΔV_{vap}) of vaporization. The data was sourced from the NIST web-book and includes a range of physicochemical descriptors such as critical temperature (T_c), critical pressure (P_c), and acentric factor (ω).	excel	yes
[42]	The dataset contains VLE experimental data of 210 binary mixtures with a total of 4562 data points, covering compounds with boiling points ranging from 293.8 K to 470.5 K. The mixtures encompass water, sulfoxides, ketones, esters, diols, cycloalkanes, carboxylic acids, aromatics, aldehydes, and alcohols. Key descriptors in the data include macroscopic properties like boiling point, critical temperature, and pressure, as well as microscopic properties such as zero-point vibrational energy, electronic energy, and Gibbs free energy.	xlsx	yes
[40]	The database consists of 7931 experimental data points from 223 different substances, covering liquids, compressed gases, and supercritical fluids. These substances span a wide range of temperatures (25.0 K to 973.2 K) and pressures (0.2 bar to 10133.0 bar). The primary focus is on the self-diffusion coefficient (D11), with the dataset including various properties like density, acentric factor, temperature, critical temperature, critical volume, number of NH/OH bonds, pressure, and the number of rotatable bonds.	-	no
[55]	The dataset contains condensed phases of 78 pure elements, compiled by the Scientific Group Thermodata Europe (SGTE). The data include the Gibbs energy variations as a function of temperature for different phases of each element, relative to a standard reference phase at 298.15 K.	tdb	yes
[38]	*JANAF Dataset*: Sourced from the NIST-JANAF database, this experimental dataset comprises data on free energy, total energy, and entropy for 694 materials in the gas phase at a temperature of 1200 K. This dataset is feature-engineered based on the compositions of the materials. *PhononDB Dataset*: Extracted from the PhononDB database, this dataset focuses on metal oxide compounds and contains information on free energy, total energy, and entropy across varying temperatures. It includes a total of 873 materials, and the features are derived using both crystal and elemental properties, along with graph-based features using CGCNN.	xlsx	yes
[82]	The dataset is related to refrigerant R134a, covering various thermodynamic and thermophysical properties provided by the National Institute of Standards and Technology (NIST). The dataset includes properties across different regions such as subcooled, superheated, supercritical, and liquid-vapor mixture phases. Key properties in the dataset are temperature, pressure, specific volume, specific internal energy, specific enthalpy, specific entropy, speed of sound, viscosity, and thermal conductivity. The data spans a temperature range from -40°C to 150°C and a pressure range from 60 kPa to 9 MPa.	csv,txt	yes
[37]	The research paper uses data extracted from phase diagrams available in the NIST database and structural information from the Inorganic Crystal Structure Database (ICSD). The NIST dataset contains 14464 phase diagrams, including 8649 binary and 5815 ternary diagrams, while the ICSD provides records on chemical formulas, crystal structures, Bravais lattices, and local atomic environments for inorganic compounds.	-	no
[54]	*Combisolv Solvation Energy Dataset*: This dataset contains one million binary pairs of molecules, calculated using the COSMO-RS model. It provides solvation energies, which describe the change in free energy when a solute molecule is placed into a solvent. *COSMO-RS Activity Coefficient Dataset*: This dataset contains over 18 million calculated activity coefficients for 460 common solvent molecules. These coefficients are computed using density functional theory (DFT) at various temperature increments and compositions, all at atmospheric pressure.	-	yes,no
[103]	The REFPROP dataset provides highly accurate thermodynamic and transport properties for a wide range of pure substances and mixtures, sourced from the NIST Reference Fluid Thermodynamic and Transport Properties Database (REFPROP). It includes data for industrial refrigerants, hydrocarbons, natural gases, and cryogenic fluids, covering key thermodynamic properties such as pressure, temperature, density, enthalpy, entropy, internal energy, and specific heat capacities. Additionally, the dataset contains transport properties, including viscosity and thermal conductivity, across a broad range of conditions. The data are derived from equations of state and experimental measurements, ensuring high accuracy for applications in process engineering, refrigeration, energy systems, and chemical thermodynamics.	FLD, HMX, xlsx, csv	no
[104]	The Dortmund Data Bank (DDB) dataset is a comprehensive collection of experimental thermophysical and thermodynamic properties for a wide range of pure substances, binary and multicomponent mixtures. It includes data for phase equilibria (vapor-liquid, liquid-liquid, and solid-liquid equilibria), excess properties, caloric properties, transport properties, and critical phenomena. The dataset spans thousands of chemical compounds, covering hydrocarbons, alcohols, ketones, esters, gases, and refrigerants across a wide range of temperatures, pressures, and compositions.	xlsx, csv	no

Dataset	Description	Data type	Open
[105]	The AGA8DC92 dataset provides high-accuracy thermodynamic and volumetric properties for natural gas mixtures based on the AGA8 Detailed Characterization Equation of State (AGA8-DC92). It includes data for pressure, temperature, density, compressibility factor, speed of sound, and heat capacities across a wide range of temperatures, pressures, and gas compositions. The dataset primarily covers methane-rich mixtures, incorporating hydrocarbons (C1-C10), nitrogen, carbon dioxide, hydrogen sulfide, and other trace components relevant to pipeline gas transport and custody transfer applications.	xlsx, csv, json	no

VI. CONCLUSION

In conclusion, this paper has provided a comprehensive survey of the application of ML in addressing thermodynamics problems. We have reviewed the key computational schemes employed in this domain and analyzed current research activities, with a particular focus on estimating vapour-liquid equilibrium, phase diagrams, and supercritical fluids, especially for pure gases. Additionally, we have introduced the main datasets available for thermodynamic property computations, including notable contributions from the GERG2008 project.

Future research should prioritise addressing the persistent trade-off between model accuracy and computational efficiency in ML applications for fluid thermodynamics. A significant focus must also be placed on developing more sophisticated models capable of accurately capturing highly non-linear fluid behaviour, particularly near critical points or under extreme conditions. This involves specific approaches aimed at improving accuracy in such challenging environments, potentially through the integration of hybrid models that combine physics-based principles with ML techniques. Furthermore, concrete steps are imperative to tackle the pervasive problem of data scarcity, including concerted efforts to improve data collection methodologies and enhance data accessibility. For PINNs, ongoing research should strive for the development of more robust NN architectures and advanced training algorithms to effectively manage PDE discretisations, accurately impose boundary conditions, and mitigate potential training instabilities. This also necessitates the establishment of a more rigorous mathematical framework for PINNs. Overall, this survey offers a unified perspective on the ML frameworks used in thermodynamic problem-solving, while also identifying existing challenges and potential future directions in the field.

ACKNOWLEDGMENT

This paper was developed under the Research contract "*Development of an Efficient Physics-Guided Hybrid Algorithm for Solving a System of Thermodynamic Equations*" between Nuovo Pignone Tecnologie s.r.l. and the Department of Information Engineering, Electronics and Telecommunications of the University of Rome "La Sapienza".

REFERENCES

[1] G. Auti, S. Paul, W.-L. Hsu, S. Chiashi, S. Maruyama, and H. Daiguji, "Statistical modeling of equilibrium phase transition in confined fluids," *Physical Review Research*, vol. 6, no. 2, p. 023333, 2024.

[2] A. Neophytou, D. Chakrabarti, and F. Sciortino, "Topological nature of the liquid–liquid phase transition in tetrahedral liquids," *Nature Physics*, vol. 18, no. 10, pp. 1248–1253, 2022.

[3] D. Fraggedakis, M. R. Hasyim, and K. K. Mandadapu, "Inherent-state melting and the onset of glassy dynamics in two-dimensional supercooled liquids," *Proceedings of the National Academy of Sciences*, vol. 120, no. 14, p. e2209144120, 2023.

[4] X. Li and Y. Jin, "Thermodynamic crossovers in supercritical fluids," *Proceedings of the National Academy of Sciences*, vol. 121, no. 18, p. e2400313121, 2024.

[5] Q. Tang, L. Zhong, C. Tang, and Y. Huang, "Unified theoretical framework for temperature regulation via phase transition," *Physical Review E*, vol. 110, no. 1, p. 014112, 2024.

[6] J. Simão, L. Emmanuel, A. João, E. Manuel, E. Nzinga, F. Cangue, and A. Barros, "Analysis of the thermodynamic behavior of gaseous mixtures using equations of state," *South African Journal of Chemical Engineering*, vol. 49, p. 339 – 347, 2024. Cited by: 0; All Open Access.

[7] M. L. Huber, E. W. Lemmon, I. H. Bell, and M. O. McLinden, "The nist refprop database for highly accurate properties of industrially important fluids," *Industrial and Engineering Chemistry Research*, vol. 61, no. 42, p. 15449 – 15472, 2022. Cited by: 123; All Open Access, Green Open Access, Hybrid Gold Open Access.

[8] I. Khmelinskii and L. V. Woodcock, "Supercritical fluid gaseous and liquid states: A review of experimental results," *Entropy*, vol. 22, no. 4, 2020. Cited by: 10; All Open Access, Gold Open Access.

[9] J. V. Madana Gopal, R. Morgan, G. De Sercey, and K. Vogiatzaki, "Overview of common thermophysical property modelling approaches for cryogenic fluid simulations at supercritical conditions," *Energies*, vol. 16, no. 2, 2023. Cited by: 7; All Open Access, Gold Open Access, Green Open Access.

[10] M. König and M. Weber Sutter, "Applying common equations of state to three reference fluids: Water, carbon dioxide, and helium," *Chemie-Ingenieur-Technik*, vol. 94, no. 4, p. 493 – 510, 2022. Cited by: 3; All Open Access, Hybrid Gold Open Access.

[11] C. M. Invernizzi and G. Di Marcoberardino, "An overview of real gas brayton power cycles: Working fluids selection and thermodynamic implications," *Energies*, vol. 16, no. 10, 2023. Cited by: 0; All Open Access, Gold Open Access, Green Open Access.

[12] I. Abdulagatov and P. Skripov, "Thermodynamic and transport properties of supercritical fluids: review of thermodynamic properties (part 1)," *Russian Journal of Physical Chemistry B*, vol. 14, pp. 1178–1216, 2020.

[13] K. N. Marsh and M. E. Kandil, "Review of thermodynamic properties of refrigerants+ lubricant oils," *Fluid Phase Equilibria*, vol. 199, no. 1-2, pp. 319–334, 2002.

[14] D. Han, S. Kwon, J. Kim, W. Jin, and H. Son, "Comprehensive analysis for production prediction of hydraulic fractured shale reservoirs using proxy model based on deep neural network," 10 2020.

[15] S. Ma, B. Xiao, R. Hong, B. Addissie, Z. Drikas, T. Antonsen, E. Ott, and S. Anlage, "Classification and prediction of wave chaotic systems with machine learning techniques," 08 2019.

[16] A. Alblawi, "Fault diagnosis of an industrial gas turbine based on the thermodynamic model coupled with a multi feedforward artificial neural networks," *Energy Reports*, vol. 6, pp. 1083–1096, 2020.

[17] S. Rezaei, A. Moeineddin, and A. Harandi, "Learning solutions of thermodynamics-based nonlinear constitutive material models using physics-informed neural networks," *Computational Mechanics*, vol. 74, no. 2, pp. 333–366, 2024.

[18] Q. Hernández, A. Badías, F. Chinesta, and E. Cueto, "Thermodynamics-informed graph neural networks," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 3, pp. 967–976, 2022.

[19] I. A. Luchnikov, A. Ryzhov, P.-J. Stas, S. N. Filippov, and H. Ouerdane, "Variational autoencoder reconstruction of complex many-body physics," *Entropy*, vol. 21, no. 11, p. 1091, 2019.

[20] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, "Physics-informed neural networks (pinns) for fluid mechanics: A review," *Acta Mechanica Sinica*, vol. 37, no. 12, pp. 1727–1738, 2021.

[21] A. Farea, O. Yli-Harja, and F. Emmert-Streib, "Understanding physics-informed neural networks: techniques, applications, trends, and challenges," *AI*, vol. 5, no. 3, pp. 1534–1557, 2024.

[22] L. Menicali, A. Grace, D. H. Richter, and S. Castruccio, "A physics-informed convolutional long short term memory statistical model for fluid thermodynamics simulations," *arXiv preprint arXiv:2505.10919*, 2025.

[23] S. Cai, Z. Wang, S. Wang, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks for heat transfer problems," *Journal of Heat Transfer*, vol. 143, no. 6, p. 060801, 2021.

[24] J. D. Toscano, V. Oommen, A. J. Varghese, Z. Zou, N. Ahmadi Daryakenari, C. Wu, and G. E. Karniadakis, "From pinns to pikans: Recent advances in physics-informed machine learning," *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, pp. 1–43, 2025.

[25] C. Zhao, F. Zhang, W. Lou, X. Wang, and J. Yang, "A comprehensive review of advances in physics-informed neural networks and their applications in complex fluid dynamics," *Physics of Fluids*, vol. 36, no. 10, 2024.

[26] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, and M. A. Marques, "Predicting the thermodynamic stability of solids combining density functional theory and machine learning," *Chemistry of Materials*, vol. 29, no. 12, pp. 5090–5103, 2017.

[27] S. Yao, A. Kronenburg, A. Shamooni, O. T. Stein, and W. Zhang, "Gradient boosted decision trees for combustion chemistry integration," *Applications in Energy and Combustion Science*, vol. 11, p. 100077, 2022.

[28] V. Shetty, S. Shedthi, and J. Kumaraswamy, "Predicting the thermodynamic stability of perovskite oxides using multiple machine learning techniques," *Materials Today: Proceedings*, vol. 52, pp. 457–461, 2022.

[29] P. Kumar, M. Alruqi, H. Hanafi, P. Sharma, and V. V. Wanatasanappan, "Effect of particle size on second law of thermodynamics analysis of al2o3 nanofluid: application of xgboost and gradient boosting regression for prognostic analysis," *International Journal of Thermal Sciences*, vol. 197, p. 108825, 2024.

[30] Z. Yu, E. R. Batista, P. Yang, and D. Perez, "Acceleration of solvation free energy calculation via thermodynamic integration coupled with gaussian process regression and improved gelman–rubin convergence diagnostics," *Journal of Chemical Theory and Computation*, vol. 20, no. 6, pp. 2570–2581, 2024.

[31] C. Desgranges and J. Delhommelle, "Towards a machine learned thermodynamics: Exploration of free energy landscapes in molecular fluids, biological systems and for gas storage and separation in metal–organic frameworks," *Molecular Systems Design & Engineering*, vol. 6, no. 1, pp. 52–65, 2021.

[32] H. Wang, Y. Cao, Z. Huang, Y. Liu, P. Hu, X. Luo, Z. Song, W. Zhao, J. Liu, J. Sun, *et al.*, "Recent advances on machine learning for computational fluid dynamics: A survey," *arXiv preprint arXiv:2408.12171*, 2024.

[33] F. Sofos, C. Stavrogiannis, K. K. Exarchou-Kouveli, D. Akabua, G. Charilas, and T. E. Karakasidis, "Current trends in fluid research in the era of artificial intelligence: A review," *Fluids*, vol. 7, no. 3, p. 116, 2022.

[34] Z. Guo, "Machine learning in the prediction of thermo-fluid behavior of nanofluids," in *Thermopedia*, Begel House Inc., 2023.

[35] G. Chaparro and E. A. Müller, "Development of thermodynamically consistent machine-learning equations of state: Application to the mie fluid," *The Journal of Chemical Physics*, vol. 158, no. 18, 2023.

[36] V. Jayaprakash and H. Li, "Deep-learning-based acceleration of critical point calculations," *Chemical Engineering Science*, vol. 298, p. 120371, 2024.

[37] F. Zipoli, V. Viterbo, O. Schilter, L. Kahle, and T. Laino, "Prediction of phase diagrams and associated phase structural properties," *Industrial & Engineering Chemistry Research*, vol. 61, no. 24, pp. 8378–8389, 2022.

[38] R. Hammad and S. Mondal, "Advancements in thermochemical predictions: A multi-output thermodynamics-informed neural network approach," *ChemRxiv (preprint)*, 2024.

[39] P. Y. Chew and A. Reinhardt, "Phase diagrams—why they matter and how to predict them," *The Journal of Chemical Physics*, vol. 158, no. 3, 2023.

[40] A. F. Dias, I. Portugal, J. P. Aniceto, and C. M. Silva, "Machine learning model for the prediction of self-diffusion coefficients in liquids, compressed gases and supercritical fluids," *Chemical Engineering Journal*, p. 153274, 2024.

[41] L. Roach, G.-M. Rignanese, A. Erriguible, and C. Aymonier, "Applications of machine learning in supercritical fluids research," *The Journal of Supercritical Fluids*, p. 106051, 2023.

[42] G. Sun, Z. Zhao, S. Sun, Y. Ma, H. Li, and X. Gao, "Vapor-liquid phase equilibria behavior prediction of binary mixtures using machine learning," *Chemical Engineering Science*, vol. 282, p. 119358, 2023.

[43] J. Ding, N. Xu, M. T. Nguyen, Q. Qiao, Y. Shi, Y. He, and Q. Shao, "Machine learning for molecular thermodynamics," *Chinese Journal of Chemical Engineering*, vol. 31, pp. 227–239, 2021.

[44] Y. Liu, W. Hong, and B. Cao, "Machine learning for predicting thermodynamic properties of pure fluids and their mixtures," *Energy*, vol. 188, p. 116091, 2019.

[45] K. Zhu and E. A. M uller, "Generating a machine-learned equation of state for fluid properties," *The Journal of Physical Chemistry B*, vol. 124, no. 39, pp. 8628–8639, 2020.

[46] M. Esmaeili, M. R. Moradi, and H. R. Afshoun, "A new empirical model and neural network-based approach for evaluation of isobaric heat capacity of natural gas," *Journal of Natural Gas Science and Engineering*, vol. 102, p. 104575, 2022.

[47] Y. Peng, X. Lin, J. Liu, W. Su, and N. Zhou, "Machine learning prediction of orc performance based on properties of working fluid," *Applied Thermal Engineering*, vol. 195, p. 117184, 2021.

[48] O. Frotscher, V. Martinek, R. Fingerhut, X. Yang, J. Vrabec, R. Herzog, and M. Richter, "Proof of concept for fast equation of state development using an integrated experimental–computational approach," *International Journal of Thermophysics*, vol. 44, no. 7, p. 105, 2023.

[49] Z. Tariq and M. Mahmoud, "New correlation for the gas deviation factor for high-temperature and high-pressure gas reservoirs using neural networks," *Energy & Fuels*, vol. 33, no. 3, pp. 2426–2436, 2019.

[50] N. Azizi, M. Rezakazemi, and M. M. Zarei, "An intelligent approach to predict gas compressibility factor using neural network model," *Neural Computing and Applications*, vol. 31, pp. 55–64, 2019.

[51] J. Qu, *Acceleration of Numerical Simulations with Deep Learning: Application to Thermodynamic Equilibrium Calculations*. PhD thesis, Sorbonne Université, 2023.

[52] A. D. Narendranath, "Introducing artificial neural networks as a specific enthalpy approximator for a course on introductory thermodynamics," in *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1–5, IEEE, 2020.

[53] U. Que-Salinas, P. E. Ramírez-González, and A. Torres-Carbajal, "Determination of thermodynamic state variables of liquids from their microscopic structures using an artificial neural network," *Soft Matter*, vol. 17, no. 7, pp. 1975–1984, 2021.

[54] K. C. Felton, H. Ben-Safar, and A. Alexei, "Deepgamma: A deep learning model for activity coefficient prediction," in *1st Annual AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE)*, 2022.

[55] D. Schatzl, "Computerassisted thermodynamics: From gibbs energy minimizers to neural networks," Master's thesis, Montanuniversitaet Leoben (000), 2022.

[56] A. Tarjomannejad, "Prediction of the liquid vapor pressure using the artificial neural network-group contribution method," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 34, no. 4, pp. 97–111, 2015.

[57] C. Desgranges and J. Delhommelle, "A new approach for the prediction of partition functions using machine learning techniques," *The Journal of Chemical Physics*, vol. 149, no. 4, 2018.

[58] E. Oren, D. Kartoon, and G. Makov, "Machine learning-based modeling of high-pressure phase diagrams: Anomalous melting of rb," *The Journal of Chemical Physics*, vol. 157, no. 1, 2022.

[59] G. Y. Ottaiano, I. N. da Cruz, H. S. da Cruz, and T. D. Martins, "Estimation of vaporization properties of pure substances using artificial neural networks," *Chemical Engineering Science*, vol. 231, p. 116324, 2021.

[60] A. Sözen, M. Özalp, and E. Arcaklioğlu, "Calculation for the thermodynamic properties of an alternative refrigerant (r508b) using artificial neural network," *Applied Thermal Engineering*, vol. 27, no. 2-3, pp. 551–559, 2007.

[61] S. Laugier and D. Richon, "Use of artificial neural networks for calculating derived thermodynamic quantities from volumetric property data," *Fluid phase equilibria*, vol. 210, no. 2, pp. 247–255, 2003.

[62] A. Chouai, S. Laugier, and D. Richon, "Modeling of thermodynamic properties using neural networks: Application to refrigerants," *Fluid phase equilibria*, vol. 199, no. 1-2, pp. 53–62, 2002.

[63] A. Torres-Carbajal, U. Que-Salinas, and P. E. Ramírez-González, "Prediction of equations of state of molecular liquids by an artificial neural network," *Revista mexicana de física*, vol. 68, no. 6, pp. 0–0, 2022.

[64] M. Farzaneh-Gord and H. R. Rahbari, "The precise designation of natural gas volumetric flow by measuring simple thermodynamic properties and using artificial intelligence methods," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 44, no. 8, p. 370, 2022.

[65] Y. Ge, M. Hansinger, C. Traxinger, and M. Pfitzner, "Deep residual learning applied to real-gas thermodynamics," in *AIP Conference Proceedings*, vol. 2040, AIP Publishing, 2018.

[66] M. Farzaneh-Gord, B. Mohseni-Gharyehsafa, A. Arabkoohsar, M. H. Ahmadi, and M. A. Sheremet, "Precise prediction of biogas thermodynamic properties by using ann algorithm," *Renewable energy*, vol. 147, pp. 179–191, 2020.

[67] M. Farzaneh-Gord, H. R. Rahbari, B. Mohseni-Gharesafa, A. Toikka, and I. Zvereva, "Accurate determination of natural gas compressibility factor by measuring temperature, pressure and joule-thomson coefficient: Artificial neural network approach," *Journal of Petroleum Science and Engineering*, vol. 202, p. 108427, 2021.

[68] G. Coccia, G. Di Nicola, S. Tomassetti, M. Pierantozzi, and G. Passerini, "Determination of the boyle temperature of pure gases using artificial neural networks," *Fluid Phase Equilibria*, vol. 493, pp. 36–42, 2019.

[69] C. Pérez *et al.*, "Thermodynamic properties of refrigerants using artificial neural networks," *International journal of refrigeration*, vol. 46, pp. 9–16, 2014.

[70] R. Sharma, D. Singhal, R. Ghosh, and A. Dwivedi, "Potential applications of artificial neural networks to thermodynamics: vapour-liquid equilib-

rium predictions,'' *Computers & chemical engineering*, vol. 23, no. 3, pp. 385–390, 1999.

[71] Z. Lin, D. Xiao, and H. Xiao, ''Physics-embedded deep learning to predict real-time flow parameters in complex thermodynamic machinery,'' *Aerospace*, vol. 11, no. 2, p. 140, 2024.

[72] J. Qu, S. Yousef, T. Faney, J.-C. de Hemptinne, and P. Gallinari, ''Nneos: Neural network-based thermodynamically consistent equation of state for fast and accurate flash calculations,'' *Applied Energy*, vol. 374, p. 124025, 2024.

[73] Y. Okano, T. Kaneshita, K. Okuno, S. Takemoto, and Y. Okuno, ''Thermodynamic equilibrium state prediction by deep learning method,'' *PriMera Scientific Engineering*, vol. 4, 2023.

[74] L. Baladão and P. Fernandes, ''Comparison of the gerg-2008 and peng-robinson equations of state for natural gas mixtures,'' *Int J Eng Res Afr*, vol. 8, no. 10, 2018.

[75] J. Gernert, A. Jäger, and R. Span, ''Calculation of phase equilibria for multi-component mixtures using highly accurate helmholtz energy equations of state,'' *Fluid Phase Equilibria*, vol. 375, pp. 209–218, 2014.

[76] M. Ramdin, T. M. Becker, S. H. Jamali, M. Wang, and T. J. Vlugt, ''Computing equation of state parameters of gases from monte carlo simulations,'' *Fluid Phase Equilibria*, vol. 428, pp. 174–181, 2016.

[77] F. Varzandeh, E. H. Stenby, and W. Yan, ''Comparison of gerg-2008 and simpler eos models in calculation of phase equilibrium and physical properties of natural gas related systems,'' *Fluid Phase Equilibria*, vol. 434, pp. 21–43, 2017.

[78] F. Dauber and R. Span, ''Achieving higher accuracies for process simulations by implementing the new reference equation for natural gases,'' *Computers & chemical engineering*, vol. 37, pp. 15–21, 2012.

[79] D. Guerrero-Zárate, A. Estrada-Baltazar, and G. A. Iglesias-Silva, ''Calculation of critical points for natural gas mixtures with the gerg-2008 equation of state,'' *Fluid Phase Equilibria*, vol. 437, pp. 69–82, 2017.

[80] O. Kunz and W. Wagner, ''The gerg-2008 wide-range equation of state for natural gases and other mixtures: an expansion of gerg-2004,'' *Journal of chemical & engineering data*, vol. 57, no. 11, pp. 3032–3091, 2012.

[81] S. Bekri, D. Özmen, and A. Özmen, ''Deep learning based combining rule for the estimation of vapor–liquid equilibrium,'' *Brazilian Journal of Chemical Engineering*, vol. 41, no. 1, pp. 613–629, 2024.

[82] M. Samawal and M. Mohamed Mousa, ''A diversified end-to-end machine learning approach for thermo-fluids' properties prediction,'' *A Diversified End-to-End Machine Learning Approach for Thermo-Fluids' Properties Prediction (June 1, 2023)*, 2023.

[83] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*. John wiley & sons, 1991.

[84] S. Davoodi, M. Al-Shargabi, D. A. Wood, M. Mehrad, and V. S. Rukavishnikov, ''Carbon dioxide sequestration through enhanced oil recovery: A review of storage mechanisms and technological applications,'' *Fuel*, vol. 366, p. 131313, 2024.

[85] M. Wensing, T. Vogel, and G. Götz, ''Transition of diesel spray to a supercritical state under engine conditions,'' *International Journal of Engine Research*, vol. 17, no. 1, pp. 108–119, 2016.

[86] D. Zhang, ''On the temperature of machine learning systems,'' *arXiv preprint arXiv:2404.13218*, 2024.

[87] N. Becattini, G. Cascini, *et al.*, ''Turbomachinery design: Checking artificial neural networks suitability for design automation,'' *Proceedings of the Design Society*, vol. 3, pp. 3651–3660, 2023.

[88] M. Song, ''Machine learning-based control systems for turbomachinery,'' 10 2024.

[89] V. Michelassi and J. Ling, ''Challenges and opportunities for artificial intelligence and high-fidelity simulations in turbomachinery applications: A perspective,'' *Journal of the Global Power and Propulsion Society*, vol. 2021, no. May, pp. 1–14, 2021.

[90] J. Zou, T. Hirokawa, J. An, L. Huang, and J. Camm, ''Recent advances in the applications of machine learning methods for heat exchanger modeling—a review,'' *Frontiers in Energy Research*, vol. 11, p. 1294531, 2023.

[91] S. Al-Dahidi, M. Alrbai, L. Al-Ghussain, and A. Alahmer, ''Maximizing energy efficiency in wastewater treatment plants: A data-driven approach for waste heat recovery and an economic analysis using organic rankine cycle and thermal energy storage,'' *Applied Energy*, vol. 362, p. 123008, 2024.

[92] M. A. Z. Rafe, ''Multi-objective optimization of organic rankine cycle based waste heat recovery systems using artificial intelligence,'' 2022.

[93] J. Wang, Q. Zuo, G. Liao, F. Luo, P. Zhao, W. Wu, Z. He, and Y. Dai, ''Machine learning–based fault detection and diagnosis of organic rankine cycle system for waste-heat recovery,'' *Journal of Energy Engineering*, vol. 147, no. 4, p. 04021016, 2021.

[94] ISO-Standard 20765-2, ''Natural gas-calculation of thermodynamic properties – part 2: Single-phase properties (gas, liquid, and dense fluid) for extended ranges of application,'' tech. rep., ISO/TC 193/SC 1, https://www.iso.org/standard/59222.html, Jan. 2015.

[95] J. Mollerup and P. Angelo, ''Measurement and correlation of the volumetric properties of a synthetic natural gas mixture,'' *Fluid phase equilibria*, vol. 19, no. 3, pp. 259–271, 1985.

[96] B. Sage, D. Botkin, and W. Lacey, ''Joule-thomson coefficients for two natural gases,'' *Transactions of the AIME*, vol. 151, no. 01, pp. 216–220, 1943.

[97] Y. Zhu, X. Lu, J. Zhou, Y. Wang, and J. Shi, ''Prediction of diffusion coefficients for gas, liquid and supercritical fluid: application to pure real fluids and infinite dilute binary solutions based on the simulation of lennard–jones fluid,'' *Fluid Phase Equilibria*, vol. 194, pp. 1141–1159, 2002.

[98] D. B. Robinson, D.-Y. Peng, and H.-J. NG, ''Applications of the peng-robinson equation of state,'' ACS Publications, 1977.

[99] G. Soave, ''Improvement of the van der waals equation of state,'' *Chemical engineering science*, vol. 39, no. 2, pp. 357–369, 1984.

[100] A. Horvath, ''Redlich-kwong equation of state: review for chemical engineering calculations,'' *Chemical Engineering Science*, vol. 29, no. 5, pp. 1334–1340, 1974.

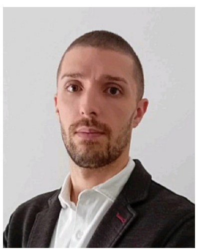
[101] G. Soave, ''Rigorous and simplified procedures for determining the pure-component parameters in the redlich—kwong—soave equation of state,'' *Chemical Engineering Science*, vol. 35, no. 8, pp. 1725–1730, 1980.

[102] S. P. Tan, H. Adidharma, and M. Radosz, ''Recent advances and applications of statistical associating fluid theory,'' *Industrial & Engineering Chemistry Research*, vol. 47, no. 21, pp. 8063–8082, 2008.

[103] E. W. Lemmon, I. H. Bell, M. L. Huber, and M. O. McLinden, ''NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 10.0, National Institute of Standards and Technology,'' 2018.

[104] D. D. B. D. GmbH), ''Dortmund data bank (ddb): Experimental thermophysical and thermodynamic property database,'' 2024.

[105] A. G. A. (AGA), ''Aga8 detailed characterization equation of state (aga8-dc92) dataset,'' 1992.



FEDERICO SUCCETTI was born in 1992. In 2019, he received the Master's Degree in Electronic Engineering from the University of Rome ''La Sapienza'', Italy. From November 2019 to October 2021 he worked as a Research Fellow and in November 2021 he started the Ph.D. in Information and Communication Technologies. He is currently a Research Fellow at the same university. His main task is the development of deep learning algorithms for the management and control of complex systems. He has gained specific experience in the development of deep neural networks for time series forecasting in the field of renewable energy.

PAOLO GIANNITRAPANI received a Master's degree in Telecommunications Engineering in 2005 and a PhD in Information and Communication Technologies (ICT) in 2022, both from the University of Rome "La Sapienza," Italy. He worked at IBM (2006–2007) on SSL/TLS architectures and digital signature algorithms and at General Electric (2007–2017) as Scientific Lead on optimization and prediction algorithms. Since 2017, he has been Scientific Director at Baker Hughes, collaborating with universities and research centers. His research interests include parallel algorithms for signal processing, spectral estimation, and array processing, estimation and detection theories. Main activities are in signal and image processing using Pattern Recognition and Deep Learning, linear and nonlinear optimization techniques.

JEAN-CHRISTOPHE RIGO earned a Master's degree in Engineering (*Diplôme d'Ingénieur*) in Energy, Aerodynamics, Combustion, and Propulsion from École Nationale Supérieure de Mécanique et Aérotechnique (ENSMA), Poitiers, France, in 2009, an Executive MBA with Honors from Quantic School of Business & Technology in 2022, and a Master's degree in Global Blockchain Technologies at Zigurat Institute of Technology and University of Barcelona in 2023. He has extensive experience in turbomachinery design and optimization processes, with a particular focus on real gas thermodynamics. He has worked as an aeromechanical R&D engineer, specializing in compressor performance computation and risk assessment, and has developed software tools for turbines and compressors. Currently, he is a Staff Software Engineer at Baker Hughes. His research interests include digital transformation in engineering, concurrent engineering systems, aeromechanical optimization, and blockchain technologies.

MASSIMO PANELLA (S'01–M'02–SM'16) was born in 1971. In 2002, he received the Ph.D. in Information and Communication Engineering from the University of Rome "La Sapienza", Italy. He is currently a Full Professor of electrical engineering, quantum computing and applied machine learning with research interests on the modeling, optimization, and control of real-world systems and for solving both supervised and unsupervised learning problems on real data. The application areas mainly focus on energy, ICT, bioengineering, economy, aerospace and security, especially considering complex systems and networked services such as smart grids, IoT, logistics, smart sensor networks, etc. He has held several editorial roles including Associate Editor of the *IEEE Transactions on Neural Networks and Learning Systems* and Associate Editor of the *IEEE Transactions on Fuzzy Systems*.

STEFANIA COLONNESE (Senior Member, IEEE) is Associate Professor with the Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Italy. She has coauthored more than a hundred journal articles and conference papers, two book chapters, and several ISO MPEG-4 contributing documents. Her research interests include statistical signal processing, image deconvolution and restoration, and biomedical signal processing to video encoding, processing, and networking. She serves as Associate Editor for IEEE Transactions on Multimedia, devoted to the topics of multimedia broadcasting, standardization, and quality of experience.