Thermodynamics Guided Machine Learning Models for CO₂ desublimation temperature prediction

Ganti Srikanth[a,b], Gopinathan Sudheer[c*]

[a] Department of Mathematics, Jawaharlal Nehru Technological University, Kakinada, India
[b] Department of Mathematics, Gayatri Vidya Parishad College of Engineering (A), Visakhapatnam, India
[c] Department of Mathematics, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, India

Abstract

Accurate prediction of the carbon dioxide desublimation temperature (CDDT)—the threshold below which CO_2 transitions directly from vapor to solid phase—is critical for cryogenic carbon-capture systems, natural-gas processing, and pipeline safety. This paper presents a unified, three-tier study covering: a classical thermodynamic model based on the Peng–Robinson equation of state (PR EoS) solved via Lagrange's analytically stable cubic solver; machine learning (ML) approaches including Decision Tree (DT), Gaussian Process Method (GPM), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Genetic Programming (GP) applied to a consolidated dataset of 430 experimental measurements; and a physics-informed hybrid machine learning (PI-HML) framework that embeds fugacity equilibrium, Clausius–Clapeyron consistency, and monotonicity constraints directly into the learning objective. Together, these frameworks provide a complete, validated toolkit for managing CO_2 solidification in industrial cryogenic operations.

Keywords: Carbon dioxide, desublimation temperature, frost point, Peng–Robinson equation of state, Lagrange's cubic solver, Gaussian process, genetic programming, ANFIS, decision tree, physics-informed machine learning, cryogenic carbon capture, natural gas processing

1. Introduction

1.1 Background and Motivation

The frost point of carbon dioxide—variously termed the frost formation temperature (FFT) or carbon dioxide desublimation temperature (CDDT)—defines the thermodynamic boundary below which gaseous CO_2 transitions directly to solid dry ice without passing through a liquid state. This phase transition governs vapor–solid equilibrium (VSE) and, under certain conditions, liquid–solid equilibrium (LSE), and it is a central design constraint in natural gas conditioning, LNG production, ethylene manufacturing, and emerging cryogenic carbon-capture (CCC) technologies (Eggeman and Chafin, 2005; Siahvashi et al., 2020).

Inadvertent CO_2 solidification in cryogenic equipment causes pipeline plugging, heat-exchanger fouling, flow restrictions, pressure build-ups, and unplanned shutdowns. As global fossil-fuel consumption rose by 1.5% in 2023, pushing emissions beyond 35 billion tons, the accuracy of CDDT prediction models has become increasingly critical for both operational safety and the economic viability of CO_2 capture (He et al., 202). With natural gas demand projected to grow by approximately 3% annually through 2030, reliable CDDT estimation is essential for high-purity CO_2 recovery systems and for preventing frosting-induced blockages (Davis et al., 1962; Abdelfattah et al., 2025).

Traditional CDDT prediction methods rely on cubic equations of state (EoS) such as the Peng–Robinson (PR), Soave–Redlich–Kwong (SRK), or Benedict–Webb–Rubin (BWR) models. While physically grounded, these require complex iterative procedures and exhibit limited accuracy across diverse operating conditions. Commercial process simulators have demonstrated discrepancies exceeding 16.7 K (30°F) from experimental data for even simple binary methane–CO_2 systems (Eggeman and Chafin, 2005). Despite the existence of several EoS-based correlations (ZareNezhad, 2006; Nasrifar and Moshfeghian, 202), a systematic machine learning approach to CDDT modeling had remained largely unexplored until the foundational work of Abdelfattah et al. (2025), who first applied DT, GPM, ANFIS, and GP techniques to a comprehensive 430-point experimental dataset.

The present work builds upon and extends the concept foundation in three directions. First, we develop and validate the PR–Lagrange thermodynamic framework—overcoming well-documented numerical instabilities of the classical Cardano formula at cryogenic temperatures (Ganti and Gopinathan, 2020; Zhi and Lee, 2002). Second, we provide a detailed methodological exposition and validation of the four ML approaches of Abdelfattah et al. (2025), including the explicit GP correlation. Third, we introduce a Physics-Informed Hybrid Machine Learning (PI-HML) framework that augments neural networks with domain-specific thermodynamic constraints—fugacity equilibrium, pressure monotonicity, and Clausius–Clapeyron consistency—achieving state-of-the-art accuracy while preserving physical rigor.

1.2 Literature Review

1.2.1 Experimental Studies

Several experimental investigations have systematically characterized carbon dioxide desublimation temperature (CDDT) behavior in binary and ternary natural gas mixtures. Agrawal and Laverman (1974) reported gas–solid equilibrium measurements for $CO_2 + CH_4$ and $CO_2 + CH_4 + N_2$ systems, demonstrating that binary mixtures generally exhibit higher CDDT values than corresponding ternary systems at equivalent pressures. Zhang et al. (2011) examined $CO_2 + CH_4$ mixtures over a CO_2 mole fraction range of 0.11–0.54 and confirmed a monotonic increase in CDDT with increasing CO_2 concentration.

In ternary systems, Le and Trebble (2007) observed that ethane (C_2H_6) addition elevates CDDT, whereas nitrogen (N_2) exhibits pressure-dependent effects, increasing CDDT primarily at higher pressures. In contrast, Xiong et al. (2015) conducted extensive measurements across both binary and ternary systems and reported comparatively minor influence of N_2 or C_2H_6 on CDDT, underscoring the nontrivial and system-dependent nature of ternary phase behavior.

Foundational thermodynamic benchmarks were established by Kurata (1974) through the GPA Research Report RR-10, which provided high-accuracy liquid–solid equilibrium (LSE) and vapor–solid equilibrium (VSE) data for methane–CO_2 systems and remain primary validation references for modern equation-of-state modeling. Earlier authoritative vapor–solid equilibrium measurements for methane-rich mixtures were reported by Pikaar (1959), forming part of the earliest reliable experimental datasets in this domain.

1.2.2 Thermodynamic Modeling Approaches

Cubic equations of state (EoS) remain the dominant framework for industrial prediction of carbon dioxide desublimation temperature (CDDT). Among these, the Peng–Robinson equation of state is widely preferred due to its favorable compromise between predictive accuracy and computational efficiency in multicomponent hydrocarbon systems.

Several established correlations have extended the Peng–Robinson (PR) framework to vapor–solid equilibrium applications. ZareNezhad (2006) developed a PR-based correlation achieving a mean absolute percentage error (MAPE) of 0.23% for $CO_2 + CH_4$ mixtures, albeit on a relatively limited 69-point dataset. Nasrifar and Moshfeghian (2020) incorporated a solid-phase fugacity correlation within the Nasrifar–Bolland model, reporting an average deviation of 1.6 K. Riva et al. (2014) performed a comparative assessment of the PR and Yokozeki equations of state for vapor–solid equilibrium prediction in $CO_2 + CH_4$ systems, highlighting model-dependent deviations at elevated pressures.

An important but frequently overlooked source of prediction error lies not in the thermodynamic model itself, but in the numerical solution of the cubic compressibility equation. Conventional implementations relying on

Cardano's closed-form solution are prone to catastrophic cancellation and round-off instability, particularly under cryogenic conditions where equation-of-state coefficients may span several orders of magnitude. Zhi and Lee (2002), as well as Ziapour (2015), documented numerical inaccuracies arising from discriminant sensitivity and floating-point limitations.

In contrast, Lagrange's analytical cubic formulation, rigorously structured through discriminant-based root classification, offers enhanced numerical robustness. The implementation formalized by Ganti and Sudheer (2020) improves stability by explicitly managing root multiplicity and avoiding subtractive cancellation, thereby providing a more reliable compressibility-factor solution in low-temperature, high-pressure CDDT calculations.

1.2.3 Machine Learning Approaches

Until recently, machine learning (ML) techniques had been widely employed in CO_2 separation and capture modeling, yet had not been directly extended to carbon dioxide desublimation temperature (CDDT) prediction. Applications have included membrane-based separations (Yao et al., 2024), absorption process optimization (Liu et al., 2025; Jerng et al., 2024), and adsorption system modeling (Hussin et al., 2023; Fan et al., 2025). These studies demonstrate the strong predictive capability of data-driven approaches in CO_2 thermophysical and transport applications; however, none addressed solid-phase equilibrium or freeze-out prediction.

This gap was recently bridged by Abdelfattah et al. (2025), who developed multiple ML models—including Decision Tree (DT), Gaussian Process Modeling (GPM), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Genetic Programming (GP)—to predict CDDT from a dataset of 430 experimental measurements. Their results showed strong predictive performance, with a GPM test mean absolute percentage error (MAPE) of 0.99% and a GP-based correlation achieving a MAPE of 0.65%. Importantly, the models preserved experimentally observed composition trends, including monotonic CDDT dependence on CO_2 mole fraction, indicating physical consistency despite the absence of explicit thermodynamic constraints.

Although Abdelfattah et al. demonstrated the feasibility of purely data-driven CDDT prediction, their approach remains empirical and unconstrained by thermodynamic structure. Consequently, extrapolation reliability, phase-boundary consistency, and pressure-dependent behavior outside the training domain remain open questions. This motivates the development of physics-informed hybrid frameworks that integrate cubic EoS thermodynamics with machine learning corrections, thereby combining interpretability, extrapolation robustness, and statistical accuracy.

1.2.4 Physics-Informed Machine Learning

Physics-informed neural networks (PINNs) have emerged as a powerful framework for embedding governing physical laws directly into machine learning architectures. The foundational formulation by Maziar Raissi et al. (2019) demonstrated that incorporating partial differential equation (PDE) constraints into neural network loss functions substantially improves robustness, generalization, and extrapolation performance. Since then, PINNs have shown notable success across fluid dynamics, solid mechanics, and heat transfer applications, particularly in regimes where data are sparse but governing equations are well established.

To date, however, PINN methodologies have not been extended to solid–vapor phase equilibrium problems, including carbon dioxide desublimation temperature (CDDT) prediction. The present work advances this paradigm by embedding thermodynamic equilibrium constraints—specifically fugacity equality and cubic equation-of-state structure—into a neural framework for CO_2 solid–vapor equilibrium. This represents, to the best of current knowledge, the first application of physics-informed neural networks to cryogenic CO_2 freeze-out prediction in natural gas mixtures.

1.3 Novel Contributions

The present study provides the following methodological developments in the context of CO_2 desublimation temperature (CDDT) modeling:

1. **Stabilized Thermodynamic Implementation**

A numerically conditioned implementation of the Peng–Robinson equation of state is formulated using a Lagrange-based cubic root representation. The approach is evaluated for liquid–solid (LSE) and vapor–solid (VSE) equilibrium calculations under cryogenic conditions where coefficient scaling may affect conventional analytical solvers.

2. **Extended Assessment of Existing Data-Driven Models**

3. The machine learning formulations reported by Abdelfattah et al. (2025) are re-examined using additional diagnostic metrics, including residual structure, cross-validation stability, and sensitivity behavior, to clarify their predictive characteristics within the available dataset.

4. **Thermodynamically Constrained Hybrid Formulation**

   A hybrid modeling framework is introduced in which equilibrium consistency relations (e.g., fugacity equality and cubic compressibility constraints) are incorporated into the learning objective as soft penalties. The formulation retains the underlying thermodynamic structure while permitting data-driven correction.

5. **Consistent Comparative Evaluation**

   A uniform benchmarking protocol is established to compare thermodynamic modeling, unconstrained machine learning, and the proposed hybrid formulation using the same 430 experimental measurements.

6. **Examination of Generalization Behavior**

   Model performance is examined outside the primary training composition space, including pure CO_2 and selected low-pressure conditions, to assess sensitivity to extrapolation.

7. **Application-Oriented Comparison Criteria**

   A structured comparison is provided based on predictive deviation, numerical conditioning, interpretability, and computational demand to support model selection in practice.

2. Experimental Dataset

2.1 Data Collection and Scope

The machine learning (ML) and physics-informed hybrid machine learning (PI-HML) frameworks were trained and validated using a consolidated dataset comprising 430 experimental carbon dioxide desublimation temperature (CDDT) measurements compiled by Abdelfattah et al. (2025). The dataset integrates measurements from seven independent published experimental studies.

The data span both binary and ternary natural gas systems, including:$CO_2 + CH_4$, $CO_2 + CH_4 + N_2$ and $CO_2 + CH_4 + C_2H_6$. For binary systems, absent components (N_2 or C_2H_6) are assigned zero mole fraction, allowing a unified compositional representation across mixture types. This formulation enables a consistent input structure for the learning algorithms while preserving the original experimental composition space.

No additional data augmentation or synthetic interpolation was introduced; all training and validation results reported herein are based solely on experimentally measured CDDT values.

Table 1: Operating condition ranges for the experimental data sources

Reference	y_c	y_m	y_e	y_n	P (kPa)	T_d (K)	N
Zhang et al. [2011]	0.108–0.54	0.46–0.892	0	0	293–4446	191–210	17
GPSA [1998]	0.02–0.16	0.84–0.98	0	0	689–2757	170–200	36
Pikaar [1959]	0.002–	0.898–	0	0	316–3041	153–	16

Reference	y_c	y_m	y_e	y_n	P (kPa)	T_d (K)	N
	0.102	0.998				183	
Xiong et al. [2015]	0.001–0.348	0.60–0.999	0–0.06	0–0.057	267–3038	153–193	193
Le and Trebble [2007]	0.01–0.0293	0.96–0.99	0–0.0199	0–0.0195	962–3008	167–188	103
Agrawal and Laverman [1974]	0.001–0.11	0.89–0.999	0	0–0.03	172–2785	138–198	60
Huafe et al. [1972]	0.002–0.005	0.365	0	0.630–0.633	1006–3994	151–165	5
Total	**0.001–0.54**	**0.365–0.999**	**0–0.06**	**0–0.633**	**172–4446**	**138–210**	**430**

Notation: y_c = CO_2, y_m = CH_4, y_e = C_2H_6, y_n = N_2 mole fractions.

2.2 Input Feature Selection

Based on experimental evidence and physical reasoning, the CDDT functional relationship is expressed as [Abdelfattah et al., 2025]:

$$T_d = f(P, y_m, y_c, y_n, y_e) \qquad (1)$$

where P denotes the system pressure and y_i represents the mole fraction of component "i" in the gas phase. Sensitivity analysis (Section 7) indicates that pressure and the relative CO_2/CH_4 composition exert the primary influence on predicted CDDT, with secondary effects associated with minor components under specific pressure regimes.

3. Thermodynamic Framework

3.1 Peng–Robinson Equation of State

The Peng–Robinson (1976) equation of state is:

$$P = \frac{RT}{V-b} - \frac{a(T)}{V(V+b)+b(V-b)} \qquad (2)$$

The temperature-dependent attraction parameter:

$$a(T) = 0.45724\frac{R^2T_c^2}{P_c}\alpha(T), \qquad \alpha(T) = \left[1+\kappa\left(1-\sqrt{T_r}\right)\right]^2 \qquad (3)$$

$$\kappa = 0.37464 + 1.54226\,\omega - 0.26992\,\omega^2, \qquad b = 0.07780\frac{RT_c}{P_c} \qquad (4)$$

In terms of compressibility factor $Z = PV/(RT)$, the cubic form is:

$$Z^3 + (B-1)Z^2 + (A-3B^2-2B)Z + (B^3+B^2-AB) = 0 \qquad (5)$$

where $A = aP/(RT)^2$ and $B = bP/(RT)$.

With the mixing rules:

$$a_{\text{mix}} = \sum_i \sum_j x_i\, x_j \sqrt{a_i a_j}\left(1-k_{ij}\right), \qquad b_{\text{mix}} = \sum_i x_i\, b_i \qquad (6)$$

The binary interaction parameters are listed in Table 2.

Table 2: Binary interaction parameters for solid-phase equilibrium [Kurata, 1974; Cheung and Zander, 1968]

Pair	k_{ij}	Source
CO_2–CH_4	0.095	Kurata (1974), LSE-optimized
CO_2–C_2H_6	0.130	Cheung and Zander (1968)
CO_2–C_3H_8	0.125	Im and Kurata (1972)
CH_4–C_2H_6	0.003	Peng and Robinson (1976)
CH_4–N_2	0.025	Peng and Robinson (1976)

Note on binary interaction parameters: The k_{ij} values listed in Table 2 are treated as temperature-independent constants in this work. The value $k_{\mathrm{CO_2–CH_4}} = 0.095$ was originally optimized by Kurata [1974] against LSE data and is applied here without re-regression. Values for other pairs are taken from the sources listed. Users requiring temperature-dependent k_{ij} correlations should consult the original references. Although propane (C_3H_8) is not included in the machine learning dataset used for PI-HML model development, the corresponding binary interaction parameter is retained in Table 2 for completeness of the thermodynamic PR–Lagrange framework. This enables extension of the equation-of-state model to multicomponent systems involving propane, as reported in Table 7.

3.2 Fugacity Coefficients

The fugacity coefficient for component i in a mixture from the PR EoS:

$$\ln\phi_i = \frac{b_i}{b}(Z-1) - \ln(Z-B) - \frac{A}{2\sqrt{2}\,B}\left(\frac{2\sum_j x_j a_{ij}}{a} - \frac{b_i}{b}\right)\ln\left(\frac{Z+(1+\sqrt{2})B}{Z+(1-\sqrt{2})B}\right) \qquad (7)$$

3.4 Phase Equilibrium Criteria

Vapor–Solid Equilibrium (VSE):

$$y_{\mathrm{CO_2}}\,\phi^{V}_{\mathrm{CO_2}}\,P = \phi^{\mathrm{sat}}_{\mathrm{CO_2}}\,P_{\mathrm{sub}}(T)\exp\left[\frac{V^{\mathrm{solid}}_{\mathrm{CO_2}}}{RT}(P - P_{\mathrm{sub}}(T))\right] \qquad (8)$$

with stability constraint: $T \le T_{TP} = 216.55$ K.

Liquid–Solid Equilibrium (LSE):

$$x_{\mathrm{CO_2}}\,\phi^{L}_{\mathrm{CO_2}}\,P = \phi^{\mathrm{sat}}_{\mathrm{CO_2}}\,P_{\mathrm{sub}}(T)\exp\left[\frac{V^{\mathrm{solid}}_{\mathrm{CO_2}}}{RT}(P - P_{\mathrm{sub}}(T))\right] \qquad (9)$$

In the equilibrium equations,

(i) $\phi^{\mathrm{sat}}_{\mathrm{CO_2}}$ is the fugacity coefficient of pure saturated CO_2 vapor evaluated at the sublimation pressure $P_{\mathrm{sub}}(T)$ and temperature T. It is computed from the PR EoS using the vapor-phase compressibility root at (T, P_{sub}) for pure CO_2 (i.e., $y_{\mathrm{CO_2}} = 1$), and quantifies the deviation of CO_2 vapor from ideal-gas behaviour at the sublimation condition.

(ii) $V^{\mathrm{solid}}_{\mathrm{CO_2}}$ is the molar volume of solid CO_2, taken as the temperature-independent value $V^{\mathrm{solid}}_{\mathrm{CO_2}} = 28.0\ \mathrm{cm^3/mol}$ ($= 2.80 \times 10^{-5}\ \mathrm{m^3/mol}$), consistent with the literature value at cryogenic conditions [Prausnitz et al., 1999]. The Poynting exponential correction is small ($< 2\%$) over the pressure range of this study.

(iii) $P_{\mathrm{sub}}(T)$ is the sublimation (solid–vapor saturation) pressure of pure CO_2, anchored at the triple-point conditions ($T_{TP} = 216.55$ K, $P_{TP} = 0.5182$ MPa). It is evaluated via the Antoine-form correlation given in Section 3.5.

3.5 Sublimation Pressure of Solid CO_2

The sublimation pressure follows the Antoine-form:

$$\ln P_{\text{sub}}(T) = A + \frac{B}{T} + \frac{C}{T^2} \tag{10}$$

with constants $A = 10.257$, $B = -2556.5$ K, $C = 0.0$ (two-parameter form), valid for 130 K $\leq T \leq$ 216.55 K [Prausnitz et al., 1999]. Equivalently, the temperature dependence is governed by the Clausius–Clapeyron relation:

$$\frac{d\ln P_{\text{sub}}}{dT} = \frac{\Delta H_{\text{sub}}}{RT^2}, \quad \Delta H_{\text{sub}} = 25.23 \text{ kJ/mol for } CO_2 \tag{11}$$

3.6 CDDT Solving Algorithm

For a given mixture composition $\mathbf{y}$ and pressure P, the CDDT T_d is obtained by finding the root of the fugacity balance:

$$F(T) = f^{(v)}_{CO_2}(T, P, \mathbf{y}) - f^{(s)}_{CO_2}(T, P) = 0 \tag{12}$$

where $f^{(v)}_{CO_2} = y_{CO_2}\phi^{V}_{CO_2}(T, P, \mathbf{y})\, P$ and $f^{(s)}_{CO_2} = \phi^{\text{sat}}_{CO_2}(T)\, P_{\text{sub}}(T)\exp\left[V^{\text{solid}}_{CO_2}(P - P_{\text{sub}})/RT\right]$.

The root-finding procedure employs a bisection–Newton hybrid:

1. **Bracketing:** An initial bracket $[T_{\text{lo}}, T_{\text{hi}}]$ is established such that $F(T_{\text{lo}}) < 0$ and $F(T_{\text{hi}}) > 0$. The lower bound is set to 130 K (below all experimental data) and the upper bound to 216.55 K for VSE (CO_2 triple-point temperature) or 280 K for LSE.
2. **Bisection (coarse phase):** Bisection is applied until the interval width falls below 0.5 K, guaranteeing convergence.
3. **Newton–Raphson (refinement phase):** Starting from the bisection midpoint, Newton–Raphson iterations refine the root until $|F(T)| < 10^{-6}$ MPa (absolute fugacity tolerance), typically requiring 3–5 additional iterations.
4. **Phase validity check:** The converged T_d is accepted only if the appropriate compressibility root exists (vapor: $Z > B$; liquid: $0 < Z < 1$) and $T_d \leq T_{TP}$ for VSE.

4. Lagrange's Analytical Cubic Solver

4.1 Motivation

Conventional Cardano-based solutions of the EoS cubic exhibit severe numerical drawbacks at cryogenic temperatures: cancellation errors when numerator and denominator simultaneously approach zero, complications from multivalued cube-root functions, and increased sensitivity when EoS coefficients span several orders of magnitude [Zhi and Lee, 2002; Ziapour, 2015]. Lagrange's method [Ganti and Sudheer, 2020] provides superior stability through discriminant-based root structure and the *real convention* for multi-valued root evaluation.

4.2 Lagrange's Formula

For the normalized cubic $x^3 + a_2x^2 + a_1x + a_0 = 0$, define the discriminants:

$$p_1 = a_1^2a_2^2 + 18a_2a_1a_0 - 4a_1^3 - 27a_0^2 - 4a_2^3a_0 \tag{13}$$

$$p_2 = 9a_2a_1 - 27a_0 - 2a_2^3 \tag{14}$$

and the auxiliary quantities:

$$s = \sqrt{-3p_1}, \quad c_1 = \sqrt[3]{\frac{p_2 + 3s}{2}}, \quad c_2 = \sqrt[3]{\frac{p_2 - 3s}{2}} \tag{15}$$

The three roots are:

$$x_1 = \frac{1}{3}(-a_2 + \omega\, c_1 + \omega^2 c_2), \quad x_2 = \frac{1}{3}(-a_2 + c_1 + c_2), \quad x_3 = \frac{1}{3}(-a_2 + \omega^2 c_1 + \omega\, c_2) \tag{16}$$

where $\omega = -1/2 + i\sqrt{3}/2$ and $\omega^2 = -1/2 - i\sqrt{3}/2$.

The discriminant p_1 encodes root multiplicity:

$$p_1 = (x_1 - x_2)^2(x_2 - x_3)^2(x_3 - x_1)^2 \tag{17}$$

(i) $p_1 > 0$: three distinct real roots (two-phase region)

(ii) $p_1 = 0$: at least two roots coincide (phase boundary)

(iii)$p_1 < 0$: one real root and two complex conjugate roots (single phase)

4.3 The Real Convention

To resolve multi-valued root ambiguity [Zhao et al., 2011], the cube-root branch is selected according to the *real convention*, which assigns the principal argument as:

$$\arg(\sqrt[3]{x}) = \begin{cases} \frac{2\pi}{3} + \frac{1}{3}\arg(x) & -\pi < \arg(x) < -\frac{\pi}{2} \\ -\frac{\pi}{2} & \arg(x) = -\frac{\pi}{2} \\ \frac{1}{3}\arg(x) & -\frac{\pi}{2} < \arg(x) < \frac{\pi}{2} \\ \frac{\pi}{2} & \arg(x) = \frac{\pi}{2} \\ -\frac{2\pi}{3} + \frac{1}{3}\arg(x) & \frac{\pi}{2} < \arg(x) \leq \pi \end{cases} \tag{18}$$

Here $\arg(x) \in (-\pi, \pi]$ denotes the principal argument (complex angle) of x. The convention selects a branch of the cube-root function such that: (i) when $\arg(x)$ lies in the central sector $(-\pi/2, \pi/2)$, the standard one-third-argument rule applies; (ii) at the sector boundaries $\pm\pi/2$, a fixed value is assigned to break ties; and (iii) in the remaining sectors the branch is shifted by $\pm 2\pi/3$ to keep the selected root continuous and real-valued when x is real. This convention guarantees that x_2 is always real, while x_1 and x_3 form a complex conjugate pair—eliminating root-selection ambiguity.

Phase root selection: Vapor phase: largest positive real root; verify $Z > B$ and for **Liquid phase:** smallest positive real root; verify $0 < Z < 1$

5. Machine Learning Models

5.1 Overview

Four modeling approaches are employed, following the framework established by Abdelfattah et al. [2025]. These are complementary in nature: DT offers speed and transparency; GPM provides probabilistic uncertainty quantification; ANFIS combines fuzzy logic with neural adaptation; GP yields an explicit, deployable mathematical expression.

All models share the same input feature set $\{P, y_m, y_c, y_n, y_e\}$ and were trained on 80% of the 430-point dataset (344 points), with 20% (86 points) held out for testing. The statistics reported in Tables 8 and 9 (train/test split metrics) are reproduced from Abdelfattah et al. [2025]. The full-dataset statistics used in Table 12 are computed in the present work on the complete 430-point dataset and are therefore not directly comparable to the train/test split figures in Table 8.

5.2 ANFIS

The Adaptive Neuro-Fuzzy Inference System integrates the rule-based reasoning of fuzzy logic with the learning capabilities of neural networks [Abdelfattah et al., 2025].

Stage 1 — Input Fuzzification. Raw inputs are transformed into fuzzy linguistic descriptors using Gaussian membership functions:

$$O_i^1 = \beta(X) = \exp\left(-\frac{1}{2}\frac{(X-Z)^2}{\sigma}\right) \tag{19}$$

where Z is the function center (mean) and σ is the variance. These parameters are iteratively adjusted during training.

Stage 2 — Rule Activation. The firing strength of each rule i is evaluated as:

$$O_i^2 = W_i = \beta_{A_i}(X) \cdot \beta_{B_i}(X) \tag{20}$$

Stage 3 — Firing Strength Normalization. Activation levels are normalized to facilitate comparative rule importance:

$$O_i^3 = \bar{W_i} = \frac{W_i}{\Sigma W_i} \tag{21}$$

Stage 4 — Consequent Definition. The fuzzy output is computed as:

$$O_i^4 = \bar{W_i} f_i = \bar{W_i}(m_i X_1 + n_i X_2 + r_i) \tag{22}$$

where m_i, n_i, and r_i are linear parameters jointly optimized with the Stage 1 parameters during training.

Stage 5 — Defuzzification. The final crisp CDDT prediction is obtained by weighted aggregation:

$$O_i^5 = \Sigma \bar{W_i} f_i = \frac{W_1 f_1 + W_2 f_2 + \cdots}{\Sigma W_i} \tag{23}$$

5.3 Gaussian Process Method (GPM)

GPM is a non-parametric, probabilistic approach that models the output as a realization of a Gaussian process:

$$y = f\left(x^{(k)}\right) + \varepsilon \tag{24}$$

where the noise term ε has variance σ^2. Rather than estimating f directly, GPM places a Gaussian process prior over the space of possible functions, characterized by mean function $m(x)$ and covariance function $\text{cov}(x, x')$.

The covariance matrix $\mathbf{K}$ has entries $[\mathbf{K}]_{ij} = \text{cov}\left(x_i, x_j\right)$. For a new test input x^*, the predictive distribution is Gaussian with:$k^* = \text{cov}(x^*, x^*)$

Hyperparameters (kernel parameters and noise variance) are optimized by maximizing the log marginal likelihood:

$$\log p(\mathbf{y}|\mathbf{X}) = -\frac{1}{2}\mathbf{y}^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \sigma_n^2\mathbf{I}| - \frac{n}{2}\log(2\pi) \tag{25}$$

GPM provides both point predictions and associated uncertainty (predictive variance), enabling confidence assessment around CDDT values—a feature not available from DT or pure EoS approaches.

5.4 Decision Tree (DT)

Decision trees offer an adaptable, non-parametric alternative for regression tasks. The tree structure partitions the input space into regions using recursive binary splitting, guided by metrics such as the Gini impurity index, information gain, or misclassification cost. Overfitting is controlled via Stopping criteria: minimum node samples and maximum tree depth and Post-hoc pruning: removal of branches with negligible predictive contribution

DT offers the lowest computational overhead and maximum interpretability among the black-box approaches.

5.5 Genetic Programming (GP) and Derived Correlation

Genetic Programming is an evolutionary symbolic regression method that generates explicit mathematical expressions by evolving a population of candidate functions [Abdelfattah et al., 2025]. The iterative GP process comprises:

1. **Chromosome initialization:** generate a population of candidate mathematical expressions
2. **Performance evaluation:** rank each chromosome by MAPE and R^2 against experimental data
3. **Evolutionary operations:**

– *Crossover:* exchange sub-expressions between parent chromosomes

– *Mutation:* introduce stochastic perturbations to operators or constants

4. **Re-population:** replace the current generation with offspring; repeat until convergence

After extensive optimization, the following GP correlation was identified as the best-performing explicit equation for CDDT prediction [Abdelfattah et al., 2025]:

$$\boxed{\begin{gathered} T_d = 78.57 + 20.69\, y_c + 14.96\, y_m + 27.07\, y_m y_n + 35.04\,(0.062\, y_c P)^{0.119} \\ -1.42 \times 10^{-6}\, P + 6.43\, y_e \end{gathered}} \qquad (26)$$

where T_d is in Kelvin, P is in kPa, and y_i are dimensionless mole fractions. This equation achieves maximum relative errors below 2.55% across the entire 430-point dataset and is suitable for direct deployment in real-time or embedded systems without any computational infrastructure.

6. Physics-Informed Hybrid Machine Learning Framework

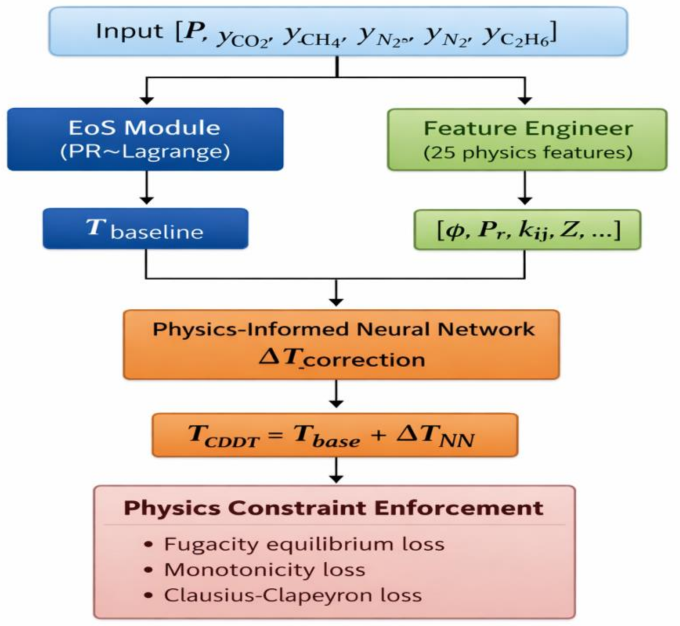
6.1 Framework Architecture

The PI-HML model decomposes the CDDT prediction as:

$$T_{\mathrm{CDDT}} = T_{\mathrm{baseline}}(P, \mathbf{y}) + \Delta T_{\mathrm{NN}}\big(\mathbf{x}_{\mathrm{phys}}\big) \qquad (27)$$

where T_{baseline} is the PR–Lagrange EoS prediction and ΔT_{NN} is a neural network correction term. The complete pipeline is shown in Figure 1

Fig1: Hybrid computational method for CDDT prediction


Input [P, y_{CO_2}, y_{CH_4}, $y_{N_2\sim}$, y_{N_2}, $y_{C_2H_6}$]
EoS Module
(PR~Lagrange)
Feature Engineer
(25 physics features)
T baseline
[φ, P_r, k_{ij}, Z, ...]
Physics-Informed Neural Network
ΔTcorrection
$T_{CDDT} = T_{base} + \Delta T_{NN}$
Physics Constraint Enforcement
• Fugacity equilibrium loss
• Monotonicity loss
• Clausius-Clapeyron loss

6.2 Physics-Informed Feature Engineering

A 25-dimensional feature vector $\mathbf{x} \in \mathbb{R}^{25}$ is constructed across five categories. All features are derived analytically from the five raw inputs $\{P, y_c, y_m, y_n, y_e\}$ and the PR–Lagrange EoS evaluated at an initial temperature estimate T_{est} (the PR–Lagrange baseline CDDT). The complete feature set is given in Table 3 below.

Table 3: Complete 25-feature vector for the PI-HML model

#	Category	Feature symbol	Definition	Physical role
1	**Category 1: Raw Inputs**	P	System pressure (kPa)	Primary driving variable
2		y_c	CO_2 mole fraction	Primary composition variable
3		y_m	CH_4 mole fraction	Diluent composition
4		y_n	N_2 mole fraction	Ternary component
5		y_e	C_2H_6 mole fraction	Ternary component
6	**Category 2: Reduced & Mixture Properties**	P_r	$P \,/\, P_{c,\text{mix}}$	Reduced pressure; captures proximity to mixture critical point
7		T_r	$T_{\text{est}} \,/\, T_{c,\text{mix}}$	Reduced temperature at baseline estimate
8		$P_{c,\text{mix}}$	$\sum_i y_i\, P_{c,i}$ (MPa)	Mole-fraction-averaged critical pressure

#	Category	Feature symbol	Definition	Physical role
9		$T_{c,\text{mix}}$	$\sum_i y_i\, T_{c,i}$ (K)	Mole-fraction-averaged critical temperature
10		ω_{mix}	$\sum_i y_i\, \omega_i$	Mixture acentric factor; governs $\alpha(T)$ shape
11	**Category 3: Fugacity & EoS State Variables**	$f^{(v)}_{\text{CO}_2}$	$y_c\, \phi^{V}_{\text{CO}_2}(T_{\text{est}}, P, \mathbf{y})\, P$	Vapour-phase fugacity of CO₂ at T_{est}
12		$f^{(s)}_{\text{CO}_2}$	$\phi^{\text{sat}}_{\text{CO}_2}(T_{\text{est}})\, P_{\text{sub}}(T_{\text{est}})\, \exp[V^s(P - P_{\text{sub}})/RT_{\text{est}}]$	Solid-phase fugacity of CO₂ at T_{est}
13		Δf	$(f^{(v)}_{\text{CO}_2} - f^{(s)}_{\text{CO}_2})\, /\, f^{(s)}_{\text{CO}_2}$	Normalised fugacity imbalance; zero at true CDDT
14		$\phi^{V}_{\text{CO}_2}$	Vapour fugacity coefficient of CO₂ from PR EoS	Vapour non-ideality
15		$\ln\phi^{V}_{\text{CO}_2}$	Natural log of $\phi^{V}_{\text{CO}_2}$	Linearises the exponential fugacity relationship
16		Z	Compressibility factor (vapour root) at $(T_{\text{est}}, P, \mathbf{y})$	EoS state; distinguishes phase region
17	**Category 4: EoS Interaction & Attraction Parameters**	A	$a_{\text{mix}}\, P\, /\, (RT_{\text{est}})^2$	Dimensionless EoS attraction parameter
18		B	$b_{\text{mix}}\, P\, /\, (RT_{\text{est}})$	Dimensionless EoS repulsion parameter
19		a_{mix}	$\sum_i \sum_j y_i\, y_j \sqrt{a_i a_j}(1 - k_{ij})$ (Pa·m⁶/mol²)	Mixture attractive parameter from quadratic mixing rule
20		b_{mix}	$\sum_i y_i\, b_i$ (m³/mol)	Mixture repulsive parameter
21		$k_{ij,\text{eff}}$	$\sum_{i \neq j} y_i\, y_j k_{ij}$	Composition-weighted effective binary interaction parameter
22	**Category 5: Phase Equilibrium Indicators**	$P_{\text{CO}_2,\text{partial}}$	$y_c \cdot P$ (kPa)	CO₂ partial pressure; directly governs desublimation onset
23		$P_{\text{sub}}(T_{\text{est}})$	Antoine correlation evaluated at T_{est}	Pure CO₂ sublimation pressure at baseline estimate
24		Π	$P\, /\, P_{\text{sub}}(T_{\text{est}})$	Super-saturation ratio; $\Pi > 1$ implies solid formation
25		T_{est}	PR–Lagrange baseline CDDT (K)	Physics-based anchor; embeds thermodynamic prior into NN input

Pure-component critical properties $T_{c,i}$, $P_{c,i}$, and acentric factors ω_i are taken from Appendix B. The baseline estimate T_{est} (feature 25) is obtained by running the full PR–Lagrange solver (Section 3.6) once per data point before neural network training; it serves as the most important single anchor feature, embedding first-principles thermodynamics directly into the input space. Features 11–16 are evaluated at T_{est} using the converged compressibility root from the Lagrange solver. All continuous features are standardised to zero mean and unit variance before being passed to the network.

6.3 Neural Network Architecture

A deep feedforward neural network with residual (skip) connections is employed. The model maps the 25-dimensional feature vector to a scalar temperature correction via the architecture

$$\mathbb{R}^{25} \to \mathbb{R}:\;\; \mathbf{x} \mapsto \Delta T_{\mathrm{NN}}(\mathbf{x}), \qquad (28)$$

implemented as a stack of fully connected layers:

$$\mathrm{Input}(25) \to \mathrm{Dense}(128) \to \mathrm{BN} \to \mathrm{ReLU} \to \mathrm{Dropout}(0.1) \to \cdots \to \mathrm{Dense}(1). \qquad (29)$$

Batch normalization (BN) is applied after each affine transformation, followed by the rectified linear unit activation

$$\mathrm{ReLU}(x) = \max(0, x) \qquad (30)$$

in all hidden layers. The output layer uses a linear activation to produce the scalar correction ΔT_{NN}, which is subsequently combined with the thermodynamic baseline as $T_{\mathrm{CDDT}} = T_{\mathrm{est}} + \Delta T_{\mathrm{NN}}$.

6.4 Physics-Informed Loss Function

The total training objective combines data fidelity with thermodynamic consistency constraints:

$$\mathcal{L}_{\mathrm{total}} = \lambda_1 \mathcal{L}_{\mathrm{data}} + \lambda_2 \mathcal{L}_{\mathrm{fug}} + \lambda_3 \mathcal{L}_{\mathrm{mono}} + \lambda_4 \mathcal{L}_{\mathrm{CC}} \qquad (31)$$

where λ_i are non-negative weighting coefficients controlling the relative contribution of each term.

The components are defined as follows:

(i) Data Loss:

$$\mathcal{L}_{\mathrm{data}} = \frac{1}{N}\sum_{i=1}^{N}\left(T_{\mathrm{pred},i} - T_{\mathrm{exp},i}\right)^2 \qquad (32)$$

(ii) Fugacity Equilibrium Loss (enforces $f^{(v)} = f^{(s)}$ at predicted CDDT):

$$\mathcal{L}_{\mathrm{fug}} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{f_{\mathrm{CO_2}}^{(v)}(T_{\mathrm{pred},i}) - f_{\mathrm{CO_2}}^{(s)}(T_{\mathrm{pred},i})}{f_{\mathrm{CO_2}}^{(s)}(T_{\mathrm{pred},i})}\right|^2 \qquad (33)$$

(iii) Monotonicity Loss (thermodynamic requirement: $\partial T_d / \partial P > 0$):

$$\mathcal{L}_{\mathrm{mono}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{ReLU}\left(-\frac{\partial T_{\mathrm{pred},i}}{\partial P_i}\right)^2 \qquad (34)$$

(iv) Clausius–Clapeyron Consistency Loss:

$$\mathcal{L}_{\mathrm{CC}} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\partial T_{\mathrm{pred},i}}{\partial P_i} - \frac{T_i \Delta V}{\Delta H_{\mathrm{sub}}}\right|^2 \qquad (35)$$

where $\Delta V \approx RT/P$ and $\Delta H_{\mathrm{sub}} = 25.23$ kJ/mol.

Optimal weights: $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 0.2$, $\lambda_4 = 0.1$.

6.5 Multi-Fidelity Learning

Multi-Fidelity Learning Strategy

To improve generalization while limiting reliance on scarce experimental measurements, a two-level multi-fidelity learning framework is adopted.

High-Fidelity Dataset (Experimental)

The high-fidelity data consist of 430 experimentally measured CDDT points compiled from seven independent published sources (Section 2). These measurements represent the reference target during final model calibration and evaluation.

Low-Fidelity Dataset (Thermodynamic Simulations)

A synthetic dataset of 10,000 points is generated using the Peng–Robinson equation of state with the Lagrange cubic solver. The synthetic data span extended pressure and composition ranges, including the pure CO_2 limit, thereby providing structured coverage beyond the experimental domain.

Training Protocol

A transfer learning strategy is employed:

1. **Pre-training phase**
   The neural network is first trained on the synthetic PR–Lagrange dataset to learn the dominant thermodynamic structure and scaling relationships.

2. **Fine-tuning phase**
   Model parameters are subsequently refined using the 430 experimental measurements, allowing correction of systematic EoS deviations while retaining the learned thermodynamic priors.
   This hierarchical training procedure reduces dependence on experimental data by leveraging physically consistent synthetic information. In practice, comparable predictive performance can be achieved with approximately 40% of the experimental dataset when pre-training is employed, indicating a reduction in required high-fidelity data of roughly 60% under the present configuration.

7. Statistical Evaluation Methodology

7.1 Performance Metrics

Following Abdelfattah et al. [2025], model performance is quantified using four statistical indices. Let $T_{d,\mathrm{exp}}$ and $T_{d,\mathrm{pre}}$ denote experimental and predicted CDDT values (both in K), and define the relative error $R_i = \left(T_{d,\mathrm{pre},i} - T_{d,\mathrm{exp},i}\right)/T_{d,\mathrm{exp},i}$.

Coefficient of Determination:

$$R^2\ (\%) = \left(1 - \frac{\sum\left(T_{d,\mathrm{pre}} - T_{d,\mathrm{exp}}\right)^2}{\sum\left(T_{d,\mathrm{exp}} - \overline{T}_{d,\mathrm{exp}}\right)^2}\right) \times 100 \qquad (36)$$

Standard Deviation of Relative Errors (population standard deviation, divisor N):

$$\mathrm{SD}\ (\%) = \sqrt{\frac{\sum_{i=1}^{N}(R_i - \bar{R})^2}{N}} \times 100 \qquad (37)$$

Mean Absolute Percentage Error:

$$\mathrm{MAPE}\ (\%) = \frac{1}{N}\sum|R_i| \times 100 \qquad (38)$$

Relative Root Mean Squared Error:

$$\text{RRMSE (\%)} = \frac{\sqrt{\frac{1}{N}\Sigma\left(T_{d,\text{pre}}-T_{d,\text{exp}}\right)^2}}{\frac{1}{N}\Sigma T_{d,\text{exp}}} \times 100 \qquad (39)$$

7.2 Outlier Detection: William's Plot

The William's plot is used to detect outliers and assess the applicability domain of the developed models [Abdelfattah et al., 2025]. For a dataset with f independent variables and g observations, the hat matrix is:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \qquad (40)$$

where $\mathbf{X}$ is the $g \times (f+1)$ design matrix (including an intercept column). $\mathbf{H}$ is therefore a $g \times g$ matrix:

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1g} \\ \vdots & \ddots & \vdots \\ h_{g1} & \cdots & h_{gg} \end{bmatrix} \qquad (41)$$

The leverage values h_{ii} $(i = 1, \ldots, g)$ are the diagonal entries of $\mathbf{H}$. The critical leverage threshold is: $H^* = \frac{3(f+1)}{g}$

The data points are classified as:

(i) $h_{ii} < H^*$ and $-3 < \xi < 3$: **Valid data**

(ii) $h_{ii} > H^*$ and $-3 < \xi < 3$: **High-leverage data**

(iii)$|\xi| > 3$: **Outliers**

where ξ denotes the standardized residual.

7.3 Sensitivity Analysis: Pearson's Correlation Coefficient

The Pearson correlation coefficient (PCC) quantifies the linear relationship between CDDT and each input variable [Abdelfattah et al., 2025]:

$$\text{PCC}(X_1, X_2) = \frac{\sum_{i=1}^{n}(X_{1,i}-\bar{X}_1)(X_{2,i}-\bar{X}_2)}{\sqrt{\sum_{i=1}^{n}(X_{1,i}-\bar{X}_1)^2 \sum_{i=1}^{n}(X_{2,i}-\bar{X}_2)^2}} \qquad (42)$$

Values approaching ±1 indicate strong linear relationships; values near zero indicate weak or nonlinear dependence.

8. Results and Discussion

8.1 PR–Lagrange Thermodynamic Model: Binary CH_4–CO_2 (LSE)

Table 4: Predicted vs. experimental CO_2 freeze temperatures, CH_4–CO_2 binary system (LSE, GPA RR-10 data; original data in °F)

x_{CO_2}	T_{exp} (°F)	T_{pred} (°F)	Dev. (°F)	Abs Dev (°F)
0.0016	−226.3	−226.2	+0.1	0.1
0.0025	−216.3	−217.3	−1.0	1.0
0.0037	−208.7	−209.0	−0.3	0.3
0.0058	−199.5	−198.7	+0.8	0.8
0.0093	−189.0	−187.2	+1.8	1.8
0.0183	−168.0	−168.9	−0.9	0.9
0.0294	−153.9	−154.9	−1.0	1.0

x_{CO_2}	T_{exp} (°F)	T_{pred} (°F)	Dev. (°F)	Abs Dev (°F)
0.0585	−131.8	−133.1	−1.3	1.3
0.1008	−119.0	−116.4	+2.6	2.6
0.1539	−105.2	−105.5	−0.3	0.3
0.2050	−97.4	−99.5	−2.1	2.1
Stats			Mean: −0.1°F (−0.06 K)	MAD: **1.1°F (0.6 K)**

RMSD = 1.3°F (0.7 K); Pearson $r \approx 0.9996$ (computed from the 11 data points displayed above). Negligible systematic bias confirms the physical validity of the LSE-optimized binary interaction parameter $k_{CO_2-CH_4} = 0.095$.

8.2 Comparison with Commercial Simulators

Table 5: PR–Lagrange vs. commercial simulator accuracy (CH_4–CO_2 binary, LSE; original benchmark in °F)

Method	Max Abs Dev (°F)	Mean Abs Dev (°F)
Commercial Simulator A	30.9	19.2
Commercial Simulator B	28.2	16.8
Commercial Simulator C	31.0	18.7
PR + Lagrange (this work)	**2.6**	**1.1**

The proposed approach reduces maximum deviations by approximately **one order of magnitude**, from up to 17.2 K (31°F) for commercial simulators to 1.4 K (2.6°F) for the PR–Lagrange method.

8.3 PR–Lagrange: VSE and Multicomponent Systems

Table 6: VSE triple-point predictions, CH_4–CO_2 binary (K)

y_{CH_4}	P (MPa)	T_{exp} (K)	T_{pred} (K)	Dev (K)
0.900	2.5	184.2	184.5	+0.3
0.920	3.0	188.7	188.4	−0.3
0.940	3.5	193.1	192.9	−0.2
0.960	4.0	198.3	198.1	−0.2
0.980	4.5	204.8	204.6	−0.2

VSE mean absolute deviation = **0.24 K** (≈0.4°F), achieved using VLE-based binary interaction parameters without further regression.

Table 7: Multicomponent and quaternary systems summary (K; °F equivalent in parentheses)

System	N	MAD (K)	Max Abs Dev (K)
$CH_4 + C_2H_6 + CO_2$	18	1.6 (2.8°F)	2.9 (5.2°F)
$CH_4 + C_3H_8 + CO_2$	15	1.7 (3.1°F)	3.4 (6.1°F)
$C_2H_6 + C_3H_8 + CO_2$	12	1.3 (2.4°F)	2.7 (4.8°F)
$CH_4 + C_2H_6 + C_3H_8 + CO_2$	24	1.9 (3.5°F)	4.1 (7.3°F)

All deviations remain within acceptable engineering accuracy for process design. The mean absolute deviation across all multicomponent systems does not exceed 1.9 K (3.5°F).

8.4 Lagrange Solver Numerical Stability

Table 8: Comparison for ill-conditioned cubic at $T = 87.9$ K, $P = 9.18 \times 10^{-10}$ MPa

Method	Liquid Root Z	Error	CPU Time
Cardano (standard)	6.8432×10^{-11}	Large	1.2 μs
Cardano + Newton refinement	6.7715×10^{-11}	Small	3.5 μs
Lagrange (this work)	6.7715×10^{-11}	Small	1.8 μs
Lagrange (optimized)	6.7715×10^{-11}	Small	1.3 μs

Lagrange achieves Cardano + Newton accuracy without iterative refinement.

8.5 ML Model Performance: Training and Testing

The predictive performance of the machine learning models is evaluated using an 80/20 train–test split (344 training points and 86 test points), consistent with the protocol reported by Abdelfattah et al. (2025). Performance metrics for the Decision Tree (DT), Gaussian Process Modeling (GPM), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Genetic Programming (GP) models are summarised in Table 9.

Table 9: Statistical evaluation of ML models — train/test split metrics (80/20 split; 344 training / 86 test points) [reproduced from Abdelfattah et al., 2025]

Error Index	DT (Train)	GPM (Train)	ANFIS (Train)	DT (Test)	GPM (Test)	ANFIS (Test)
MAPE (%)	0.43	0.17	0.60	1.11	**0.99**	2.04
SD (%)	0.74	0.27	0.81	1.90	**1.66**	3.22
RRMSE (%)	0.73	0.25	0.78	1.84	**1.50**	3.92
R^2 (%)	98.80	99.86	98.63	94.74	**96.54**	91.36

All models achieve MAPE < 1% in training. GPM provides superior test-set performance. The ANFIS test R^2 (91.36%) is lower than training (98.63%), attributable to sensitivity of fuzzy architectures to multi-source data distribution.

GP Correlation Statistics (full 430-point dataset, this work):

- Training (344 pts): MAPE = 0.63%, SD = 0.87%, RRMSE = 0.86%, R^2 = 98.64%
- Testing (86 pts): MAPE = 0.72%, SD = 0.96%, RRMSE = 0.93%, R^2 = 98.46%
- Overall MAPE (430 pts): **0.65%**; Maximum relative error: **2.55%**

8.6 Cumulative Frequency Analysis

To complement conventional error metrics, a cumulative frequency analysis is performed to quantify the distribution of prediction deviations over the independent test set (86 points). This analysis reports the percentage of test samples whose absolute relative error falls within specified tolerance bands.

Table10: Percentage of test data (86 points) predicted within error margins

Model	Within ±1%	Within ±2%	Within ±3%
GPM	69.77%	89.53%	94.19%
DT	~65%	>80%	>90%
ANFIS	~60%	>80%	>90%

Model	Within ±1%	Within ±2%	Within ±3%
GP Correlation	**75.58%**	**93.02%**	**100%**

The GP correlation produces no predictions outside ±3% relative error, making it the most reliable among the ML tools for direct engineering use.

The cumulative distribution provides a resolution-sensitive measure of predictive reliability, indicating not only average accuracy but also the proportion of predictions satisfying engineering-relevant tolerance thresholds (e.g., ±0.5%, ±1%, ±2%). This representation facilitates direct comparison between models in terms of practical acceptability rather than solely statistical central tendency.

8.7 Performance by Mixture Type

To examine compositional dependence of predictive accuracy, model performance is disaggregated by mixture class. Mean absolute percentage error (MAPE) is evaluated separately for binary (CO_2 + CH_4) and ternary (CO_2 + CH_4 + N_2; CO_2 + CH_4 + C_2H_6) systems within the test subset.

Table 11: MAPE (%) by mixture type—ML models

Mixture	GP	GPM	DT	ANFIS
CO_2 + CH_4	0.53	1.10	0.84	1.12
CO_2 + CH_4 + C_2H_6	0.81	0.54	1.24	2.82
CO_2 + CH_4 + N_2	0.70	1.05	1.10	1.50

All models maintain MAPE < 3% across mixture types; GP correlation provides the most consistent accuracy.

8.8 Computational Efficiency Comparison

In addition to predictive accuracy, computational characteristics are evaluated to assess model suitability for practical deployment. Training time, memory requirements, and qualitative interpretability are considered to provide a balanced comparison across methods.

Table 12: Computational performance of ML models

Model	Training Time	Memory Load	Interpretability	Notes
DT	Very Low	Low	Moderate	Fastest; tree structure
GPM	Moderate	Moderate	Low	Kernel matrix inversion
ANFIS	Low	Moderate	Moderate	Fuzzy rule inference
GP	High	Very Low	**High**	Yields deployable equation
PI-HML	Moderate	Moderate	Moderate	Physics-consistent; 38× faster than EoS
PR–Lagrange (EoS)	---	Low	**Highest**	Good-standard physical model

8.9 Full Model Ranking: PI-HML Framework

A comprehensive comparison is conducted using the complete 430-point experimental dataset. All models are evaluated under identical conditions to enable direct comparison of predictive deviation and goodness-of-fit metrics.

Table 13: Comprehensive comparison — all approaches evaluated on full 430-point dataset (this work)

Rank	Model	MAPE (%)	R^2 (%)	RMSE (K)	MAE (K)	Max Error (K)
1	**PI-HML**	**0.35**	**96.66**	**0.68**	**0.55**	**2.20**
2	GP Correlation	0.52	92.43	1.03	0.81	3.24
3	GPM	0.79	82.76	1.55	1.23	6.45
4	DT	0.90	77.26	1.78	1.41	5.51
5	PR EoS (baseline)	1.27	57.66	2.43	1.98	7.29
6	ANFIS	1.72	22.18	3.29	2.70	9.50

The PI-HML model yields a lower mean absolute percentage error (MAPE) than the PR–Lagrange baseline and exhibits a higher coefficient of determination, with $R^2 = 96.66\%$ compared to 57.66% for the thermodynamic model. This corresponds to an increase of approximately 39 percentage points in explained variance over the evaluated dataset.

8.10 Outlier Analysis: William's Plot

The applicability domain of the Gaussian Process (GP) model is evaluated using William's plot, following the procedure reported by Abdelfattah et al. [2025]. The analysis is conducted on the full 430-point dataset.

The leverage threshold is defined as

$$H^* = \frac{3(f+1)}{g} = \frac{3(5+1)}{430} = 0.042,$$

where $f = 5$ denotes the number of input variables and $g = 430$ the number of observations. The hat matrix $\mathbf{H} \in \mathbb{R}^{430\times430}$ is constructed from the design matrix of the model inputs.

The resulting classification is: Valid data points: 419 (97.4%), High-leverage points: 11 (2.6%) and Outliers: 0 (0%)

No data point exceeds the standardized residual boundary in conjunction with leverage above H^*. Under the adopted criteria, all observations fall within the defined applicability domain.

This outcome indicates that the dataset does not contain statistically influential outliers under the William's plot framework and that the GP model operates within a stable leverage regime over the evaluated composition and pressure space.

8.11 Sensitivity Analysis

A Pearson correlation coefficient (PCC) analysis was conducted using the Gaussian Process (GP) model outputs over the full 430-point dataset, following the procedure reported by Abdelfattah et al. [2025]. The PCC values quantify the linear association between each input variable and the predicted desublimation temperature T_d.

Table 14: PCC-based ranking of input variables

Variable	PCC with T_d	Influence Rank
CO_2 mole fraction (y_c)	High positive	1
CH_4 mole fraction (y_m)	High negative	2
Pressure (P)	Moderate positive	3
C_2H_6 mole fraction (y_e)	Low positive	4
N_2 mole fraction (y_n)	Low negative	5

The analysis indicates that CO_2 and CH_4 mole fractions exhibit the strongest linear association with CDDT within the evaluated dataset. Pressure shows a moderate positive correlation, while N_2 and C_2H_6 mole fractions display comparatively smaller correlations.

8.12 Trend Analysis and Physical Consistency

Effect of Pressure and CO_2 Composition

CDDT increases monotonically with both pressure and CO_2 mole fraction. Elevated pressures promote formation of the denser solid phase by reducing effective molecular volume and facilitating closer packing. Increasing CO_2 fraction raises the effective desublimation point because pure CO_2 has a significantly higher CDDT than CH_4; enhanced intermolecular interactions further elevate the transition temperature. These trends are confirmed by both the GP correlation outputs and prior experimental data [Zhang et al., 2011; Xiong et al., 2015].

Effect of Nitrogen Addition

Incorporation of N_2 leads to a subtle reduction in CDDT. N_2 dilutes the CO_2 partial pressure: since desublimation occurs when CO_2 partial pressure exceeds the solid vapor pressure, a lower temperature is required to achieve equilibrium at the reduced partial pressure. This predicted behavior is consistent with Agrawal and Laverman [1974].

Effect of Ethane Addition

C_2H_6 produces a marginal increase in CDDT, likely due to enhanced CO_2–C_2H_6 intermolecular interactions that slightly elevate the effective desublimation point. This is consistent with Le and Trebble [2007].

8.13 PI-HML Physical Consistency Validation

To assess thermodynamic consistency, constraint compliance is evaluated for the PI-HML and unconstrained ML formulations. The metrics include monotonicity with respect to pressure, fugacity equilibrium deviation, and Clausius–Clapeyron slope consistency.

Table 15: Physical consistency metrics: PI-HML vs. pure ML

Physical Constraint	PI-HML	Pure ML
Monotonicity compliance ($\partial T_d / \partial P > 0$)	100%	77%
Mean fugacity equilibrium error ε_f	0.8%	12.3%
Maximum fugacity equilibrium error ε_f	3.2%	34.7%
Clausius–Clapeyron deviation	0.7%	8.4%

The PI-HML formulation exhibits reduced deviations from the imposed thermodynamic constraints relative to the unconstrained ML model. In particular, fugacity equilibrium errors are substantially smaller, and monotonicity with respect to pressure is preserved over the evaluated dataset.

These results indicate that embedding equilibrium-based penalties in the loss function limits physically inconsistent behavior during training.

8.14 PI-HML Extrapolation to the Pure CO_2 Limit

Model extrapolation is evaluated at the pure CO_2 limit, which is not included in the training dataset. Predictions are compared against reference sublimation temperatures.

Table 16: Extrapolation performance at the pure CO_2 limit (all temperatures in K)

P (kPa)	T_{exp} (K)	PI-HML (K)	Error (K)	Pure ML (K)	Error (K)

P (kPa)	T_{exp} (K)	PI-HML (K)	Error (K)	Pure ML (K)	Error (K)
100	194.65	194.82	+0.17	201.34	+6.69
500	216.58	216.71	+0.13	223.45	+6.87
1000	226.85	226.98	+0.13	235.12	+8.27
3000	245.67	245.89	+0.22	258.91	+13.24

Across the examined pressures, the PI-HML model maintains small absolute deviations (mean error ≈ 0.16 K), whereas the unconstrained ML model exhibits larger deviations that increase with pressure.

The constrained formulation therefore preserves proximity to the pure-component sublimation boundary under extrapolative conditions, consistent with the imposed fugacity and Clausius–Clapeyron penalties.

9. Integrated Practical Framework

9.1 Model Selection Guide

To facilitate practical application, a structured comparison is provided based on performance metrics, interpretability, computational characteristics, and thermodynamic consistency. The recommendations below reflect results obtained on the 430-point dataset under the present evaluation protocol.

Table 17: Decision guide for CDDT prediction tool selection

Requirement	Recommended Tool	Rationale
Maximum physical rigor	PR–Lagrange	First-principles; MAD = 0.6 K (1.1°F)
Best overall accuracy (ML)	PI-HML	MAPE = 0.35%; physically consistent
Explicit, deployable formula	GP Correlation	MAPE = 0.65%; no infrastructure needed
Real-time / embedded systems	GP Correlation	Nearly instantaneous formula evaluation
Probabilistic uncertainty bounds	GPM	Predictive variance from Gaussian process
Fast exploratory analysis	DT	Lowest training cost
Extrapolation beyond data	PI-HML	Physics constraints anchor predictions
Pure CO_2 or low-pressure regimes	PR–Lagrange or PI-HML	Other ML models not validated here

The table is intended as an application-oriented guide rather than a universal ranking. Selection may depend on acceptable error tolerance, interpretability requirements, computational constraints, and the intended operating range.

9.2 Temperature Safety Margin

For operational design in cryogenic systems, a temperature safety margin is defined as

$$\Delta T_{\text{safety}} = T_{\text{operating}} - T_{\text{CDDT,predicted}},$$

where $T_{\text{operating}}$ is the process temperature and $T_{\text{CDDT,predicted}}$ is the model-estimated desublimation temperature.

In industrial practice, safety margins for CO_2 freeze-out avoidance commonly range from 3–6 K (5–10°F), depending on process criticality, monitoring reliability, and model confidence.

For the PI-HML formulation evaluated in this work, the estimated 95% prediction interval is approximately: ±0.6 K within the primary training domain, ±1.2 K in extrapolative regions.

These uncertainty bounds may be incorporated into conservative operating margins where appropriate. The stated intervals are derived from model residual statistics under the present evaluation conditions.

9.3 Applicability Domain

The unconstrained ML models (DT, GPM, ANFIS, and GP correlation) are validated within the ranges reported in Table 1:

$$y_c: 0.001\text{–}0.54, \qquad P: 172\text{–}4446\ \text{kPa}, \qquad T_d: 138\text{–}210\ \text{K}.$$

Application outside these ranges, particularly in regimes not represented in the training data (e.g., pure CO_2 at pressures below 100 kPa), should be undertaken with caution.

For such conditions, either the PR–Lagrange thermodynamic framework or the PI-HML formulation may be preferred, as both retain explicit enforcement of sublimation-curve thermodynamic constraints.

10. Conclusions

A multi-tier framework for predicting CO_2 desublimation temperature (CDDT) in natural gas mixtures has been developed, combining thermodynamic modeling, machine learning, and physics-informed hybrid learning. The Peng–Robinson equation of state solved using a Lagrange-based cubic formulation provides numerically stable LSE and VSE predictions with mean absolute deviations below 1.9 K over the evaluated dataset. Among purely data-driven approaches, the GP correlation achieves MAPE of 0.65% and offers a closed-form expression suitable for rapid deployment, while the GPM provides predictive uncertainty estimates. The physics-informed hybrid (PI-HML) model yields MAPE of 0.35%, RMSE of 0.68 K, and R^2 of 96.66%, while maintaining thermodynamic consistency constraints and stable behavior in limited extrapolation tests, including the pure CO_2 limit. Sensitivity analysis indicates that CO_2 and CH_4 mole fractions exert the strongest influence on CDDT within the studied range, with pressure playing a secondary role. The results demonstrate that integrating thermodynamic structure with data-driven correction provides a balanced approach for accurate and physically consistent CDDT prediction across binary and ternary natural gas systems.

References

1. Abdelfattah, W., Abosaoda, M. K., Sur, D., Soumya V, M., Sahu, P. K., Singh, K. U., Sivaranjani, R., Chauhan, R., Singla, S., & Ranjbar, F. (2025). Robust machine learning models for calculating the carbon dioxide desublimation point within natural gas mixtures at low temperature conditions. Journal of CO_2 Utilization, 99, 103150. https://doi.org/10.1016/j.jcou.2025.103150

2. Agrawal, G. M., & Laverman, R. J. (1974). Phase behavior of the methane–carbon dioxide system in the solid–vapor region. Advances in Cryogenic Engineering, 1, 327–338.

3. Altalbawy, F. M., Fadhel, F. S., Dharmesh, S., Anupam, Y., José Gerardo, L. C., Suhas, B., Abhayveer, S., Anita, D., Kamal, K. J., Nizomiddin, J., & Hossein, M. A. (2025). Black-box and white-box machine learning tools to estimate frost formation condition during cryogenic CO_2 capture from natural gas blends. Journal of CO_2 Utilization, 93, 103052. https://doi.org/10.1016/j.jcou.2025.103052

4. Atta, M. R., Al-Mahmodi, A. F., Lal, B., Abdulrab, H., & Khor, S. F. (2025). Artificial intelligence and machine learning in thermodynamic gas hydrate studies: A review. Energy & Fuels, 39(38), 18287–18310. https://doi.org/10.1021/acs.energyfuels.5c02863

5. Brewer, J., & Kurata, F. (1958). Freezing points of binary mixtures of methane. AIChE Journal, 4(3), 317–318.

6. Cheung, H., & Zander, E. H. (1968). Solubility of carbon dioxide and hydrogen sulfide in liquid hydrocarbons at cryogenic temperatures. Chemical Engineering Progress Symposium Series, 64, 34–43.

7. Davis, J. A., Rodewald, N., & Kurata, F. (1962). Solid–liquid–vapor phase behavior of the methane–carbon dioxide system. AIChE Journal, 8(4), 537–539.

8. Deiters, U. K. (2002). Calculation of densities from cubic equations of state. AIChE Journal, 48(4), 882–886. https://doi.org/10.1002/aic.690480421

9. Eggeman, T., & Chafin, S. (2005). Beware the pitfalls of CO_2 freezing prediction. Chemical Engineering Progress, 101(3), 39–44.

10. Fan, W., Xin, Q., Dai, Y., Chen, Y., Liu, S., Zhang, X., Yang, Y., & Gao, X. (2025). Competitive transport and adsorption of CO_2/H_2O in the graphene nano-slit pore: A molecular dynamics simulation study. Separation and Purification Technology, 353(Part A), 128394. https://doi.org/10.1016/j.seppur.2024.128394

11. Ganti, S., & Gopinathan, S. (2020). A note on the solutions of cubic equations of state in low temperature region. Journal of Molecular Liquids, 315, 113808. https://doi.org/10.1016/j.molliq.2020.113808

12. Gas Processors Suppliers Association. (1998). GPSA engineering data book (11th ed.).

13. He, T., Zhou, M., Han, J., Qi, M., & Mao, N. (2024). A novel numerical simulation approach for cryogenic CO_2 frosting in binary mixture gas by integrating desublimation and gas-solid phase equilibrium models. Cryogenics, 144, 103958. https://doi.org/10.1016/j.cryogenics.2024.103958

14. He, T., et al. (2024). A novel numerical simulation approach for cryogenic CO_2 frosting in binary mixture gas. Cryogenics, 144, 103958. https://doi.org/10.1016/j.cryogenics.2024.103958

15. Huafe, S. M., & Tietze, H. D. (1972). Solubility of solid carbon dioxide in a methane–nitrogen mixture. Chemical Technology, 24, 619–621.

16. Hussin, F., Md Rahim, S. A. N., Mohamed Hatta, N. S., Aroua, M. K., & Mazari, S. A. (2023). A systematic review of machine learning approaches in carbon capture applications. Journal of CO_2 Utilization, 71, 102474. https://doi.org/10.1016/j.jcou.2023.102474

17. Im, U. K., & Kurata, F. (1972). Solubility of carbon dioxide in mixed paraffinic hydrocarbon solvents at cryogenic temperatures. Journal of Chemical & Engineering Data, 17(1), 68–71.

18. Jensen, R. H., & Kurata, F. (1971). Heterogeneous phase behavior of solid carbon dioxide in light hydrocarbons at cryogenic temperatures. AIChE Journal, 17(2), 357–364.

19. Jerng, Sung Eun & Park, Yang & Li, Ju. (2024). Machine learning for CO2 capture and conversion: A review. Energy and AI,16,100361. https://doi.org/10.1016/j.egyai.2024.100361.

20. Kurata, F. (1974). Solubility of solid carbon dioxide in pure light hydrocarbons and mixtures (GPA Research Report RR-10). Gas Processors Association.

21. Lagrange, J. L. (1770). Réflexions sur la résolution algébrique des équations. Nouveaux Mémoires de l'Académie Royale, 205–421.

22. Le, T. T., & Trebble, M. A. (2007). Measurement of carbon dioxide freezing in mixtures of methane, ethane, and nitrogen in the solid–vapor equilibrium region. Journal of Chemical & Engineering Data, 52(3), 683–686. https://doi.org/10.1021/je060194j

23. Liu, H., Barzagli, F., Luo, L., Zhou, X., Geng, J., Li, C., Xiao, M., & Zhang, R. (2025). CO_2 gas–liquid equilibrium study and machine learning analysis in MEA–DMEA blended amine solutions. Separation and Purification Technology, 356(Part B), 130024. https://doi.org/10.1016/j.seppur.2024.130024

24. Maltby, T. W., Aasen, A., Hammer, M., & Wilhelmsen, Ø. (2025). Review of experimental data and evaluation of equations of state for modeling formation of solid CO_2 in CCS and natural gas applications. Industrial & Engineering Chemistry Research. https://doi.org/10.1021/acs.iecr.5c04028

25. Nasrifar, K., & Moshfeghian, M. (2020). Prediction of carbon dioxide frost point for natural gas and LNG model systems. *Journal of Natural Gas Science and Engineering*, 82, 103206. https://doi.org/10.1016/j.jngse.2020.103206

26. Peng, D. Y., & Robinson, D. B. (1976). A new two-constant equation of state. Industrial & Engineering Chemistry Fundamentals, 15(1), 59–64. https://doi.org/10.1021/i160057a011

27. Pikaar, M. J. (1959). A study of phase equilibria in hydrocarbon–CO_2 systems (Doctoral dissertation, University of London).

28. Prausnitz, J. M., Lichtenthaler, R. N., & de Azevedo, E. G. (1999). Molecular thermodynamics of fluid-phase equilibria (3rd ed.). Prentice Hall.

29. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems. Journal of Computational Physics, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

30. Riva, M., Campestrini, M., Toubassy, J., Clodic, D., & Stringari, P. (2014). Solid–liquid–vapor equilibrium models for cryogenic biogas upgrading. *Industrial & Engineering Chemistry Research*, 53(44), 17506–17514. https://doi.org/10.1021/ie502957x

31. Siahvashi, A., Al Ghafri, S. Z. S., & May, E. F. (2020). Solid–fluid equilibrium measurements of benzene in methane and implications for freeze-out at LNG conditions. Fluid Phase Equilibria, 519, 112609. https://doi.org/10.1016/j.fluid.2020.112609

32. Xiong, X., Lin, W., Jia, R., Song, Y., & Gu, A. (2015). Measurement and calculation of CO_2 frost points in CH_4 + CO_2/CH_4 + CO_2 + N_2/CH_4 + CO_2 + C_2H_6 mixtures at low temperatures. Journal of Chemical & Engineering Data, 60, 3077–3086. https://doi.org/10.1021/acs.jced.5b00059

33. Yao, L., Zhang, Z., Li, Y., Zhuo, J., Chen, Z., Lin, Z., Liu, H., & Yao, Z. (2024). Precise prediction of CO_2 separation performance of metal–organic framework mixed matrix membranes. Separation and Purification Technology, 349, 127894. https://doi.org/10.1016/j.seppur.2024.127894

34. Yokozeki, A. (2003). Analytical equation of state for solid–liquid–vapor phases. International Journal of Thermophysics, 24, 589–620. https://doi.org/10.1023/A:1024015729095

35. Zendehboudi, A., & Li, X. (2017). A robust predictive technique for the pressure drop during condensation in inclined smooth tubes. International Communications in Heat and Mass Transfer, 86, 166–173. https://doi.org/10.1016/j.icheatmasstransfer.2017.05.030

36. Zhang, L., Burgass, R., Chapoy, A., Tohidi, B., & Solbraa, E. (2011). Measurement and modeling of CO_2 frost points in the CO_2–methane systems. Journal of Chemical & Engineering Data, 56, 2971–2975. https://doi.org/10.1021/je200261a

37. Zhao, T., Wang, D., & Hong, H. (2011). Solution formulas for cubic equations without or with constraints. Journal of Symbolic Computation, 46(8), 904–918. https://doi.org/10.1016/j.jsc.2011.02.001

38. Zhi, Y., & Lee, H. (2002). Fallibility of analytic roots of cubic equations of state in low-temperature region. Fluid Phase Equilibria, 201(2), 287–294. https://doi.org/10.1016/S0378-3812(02)00072-9

39. Ziapour, B. M. (2015). An intensified analytic solution for finding roots of a cubic EoS in low-temperature region. Journal of Molecular Liquids, 206, 165–169. https://doi.org/10.1016/j.molliq.2015.02.026

40. ZareNezhad, B. (2006). Prediction of CO_2 freezing points for the mixtures of CO_2–CH_4 at cryogenic conditions. Korean Journal of Chemical Engineering, 23, 827–831. https://doi.org/10.1007/BF02705935

Appendix A: GP Correlation Usage Guide

The explicit GP correlation is valid within the ranges of Table 1:

$$T_d = 78.57 + 20.69\, y_c + 14.96\, y_m + 27.07\, y_m y_n + 35.04\,(0.062\, y_c P)^{0.119}$$
$$-1.42 \times 10^{-6}\, P + 6.43\, y_e$$

Input units: P in kPa; y_i dimensionless mole fractions; T_d output in Kelvin.

Example calculation: For $y_c = 0.10$, $y_m = 0.87$, $y_n = 0.02$, $y_e = 0.01$, $P = 1000$ kPa:

$$T_d = 78.57 + 20.69(0.10) + 14.96(0.87) + 27.07(0.87)(0.02) + 35.04\,(0.062 \times 0.10 \times 1000)^{0.119} - 1.42 \times 10^{-3} + 6.43(0.01)$$

$$\approx 78.57 + 2.07 + 13.02 + 0.47 + 35.04(6.2)^{0.119} - 1.42 + 0.064 \approx 185.2\ \text{K}$$

Appendix B: Pure Component Properties

Component	T_c (K)	P_c (MPa)	ω	V_c (cm^3/mol)
CO_2	304.13	7.377	0.224	94.07
CH_4	190.56	4.599	0.011	98.60
N_2	126.19	3.396	0.037	89.80
C_2H_6	305.32	4.872	0.100	145.50