

High-Precision Modeling of Gas-Phase Density for the DME/HFC-134a Binary System Using Machine Learning

Sadegh Sahraei

sahraei.s@lu.ac.ir

Lorestan University

Mohammad Mahdi Tafarroj

Lorestan University

Fatemeh Karami

Lorestan University

Research Article

Keywords: Gas-phase density, Machine Learning, Support Vector Regression, Refrigerant blends, Thermodynamic modeling

Posted Date: September 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10883998/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Chemical Papers on October 5th, 2026.
See the published version at <https://doi.org/10.1007/s11696-026-05708-2>.

High-Precision Modeling of Gas-Phase Density for the DME/HFC-134a Binary System Using Machine Learning

Sadegh Sahraei^{1,*}, Mohammad Mahdi Tafarroj², Fatemeh Karami²

¹ Department of Polymer Engineering, Faculty of Engineering, Lorestan University, Khorramabad, Iran

² Department of Mechanical Engineering, Faculty of Engineering, Lorestan University, Khorramabad, Iran

* Corresponding author, E-mail address: sahraei.s@lu.ac.ir

Abstract

Accurate prediction of gas-phase density is essential for the design and optimization of refrigeration systems employing environmentally benign refrigerant blends. This study evaluates the predictive performance of four machine learning models—Linear Regression, Support Vector Regression (SVR), Random Forest, and Gradient Boosting—for estimating the density of the binary DME/HFC-134a system. A dataset comprising 101 experimental PVT data points, spanning temperatures from 313.53 to 383.49 K, pressures from 541.8 to 3170.2 kPa, and HFC-134a mole fractions from 0.2326 to 0.8258, was employed. The data were acquired through the isochoric method with high precision. Hyperparameters of the non-linear models were optimized via Grid Search combined with 5-fold cross-validation. Among the models examined, the SVR with a radial basis function (RBF) kernel demonstrated superior performance, achieving the highest coefficient of determination ($R^2 = 0.9993$) and the lowest root mean square error (RMSE = $0.0095 \text{ mol}\cdot\text{dm}^{-3}$) on the test set. Furthermore, the predictive capability of the machine learning models was benchmarked against two classical cubic equations of state (Peng–Robinson and Soave–Redlich–Kwong), revealing that the SVR model surpasses both thermodynamic approaches in accuracy. The findings confirm that non-linear machine learning techniques provide a robust, computationally efficient, and accurate alternative to conventional empirical correlations for modeling the thermodynamic properties of DME/HFC-134a mixtures.

Keywords: Gas-phase density, Machine Learning, Support Vector Regression, Refrigerant blends, Thermodynamic modeling.

1. Introduction

The environmental impact of refrigerants has been a major concern since Molina and Rowland's groundbreaking discovery in 1974, which showed that chlorine-containing compounds catalytically destroy stratospheric ozone [1, 2]. This led to the Montreal Protocol and the phase-out of chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs). Hydrofluorocarbons (HFCs) were introduced as zero ozone depletion potential (ODP) alternatives, but many, such as HFC-134a with global warming potential (GWP) 1430, possess high GWP and contribute significantly to climate change [3, 4]. Regulatory measures, including the EU's ban on refrigerants with $\text{GWP} > 150$ in new vehicles, have created an urgent demand for low-GWP alternatives [5, 6]. Recent studies emphasize that the transition to low-GWP refrigerants is not only an environmental necessity but also a technical challenge requiring accurate property data for system design [7-10].

Dimethyl ether (DME) has emerged as a promising candidate due to its excellent thermodynamic properties, zero ODP, zero GWP, low cost, and good compatibility. However, its high flammability limits practical use [11, 12]. Blending DME with HFC-134a has been proposed to reduce flammability while keeping the mixture's GWP below 150, making the DME/HFC-134a binary system an attractive option for next-

generation refrigeration. Recent experimental studies have confirmed that this blend exhibits favorable vapor-liquid equilibrium behavior and reduced flammability compared to pure DME [13, 14].

Reliable design and optimization of refrigeration systems using this blend require accurate thermodynamic properties, particularly gas-phase PVT behavior [15]. Density is essential for calculating heat transfer, pressure drop and compressor work. Although vapor-liquid equilibrium data exist, comprehensive gas-phase PVT data for DME/HFC-134a were scarce until a recent experimental study measured 101 data points using the isochoric method [16]. Traditional modeling approaches, such as cubic equations of state (e.g., Peng-Robinson, Soave-Redlich-Kwong) and virial equations, have been widely used for refrigerant mixtures [17-19]. However, these methods often require empirical mixing rules and binary interaction parameters, which may not generalize well across wide thermodynamic ranges.

In recent years, machine learning (ML) has emerged as a powerful alternative capable of learning non-linear patterns directly from data without strong physical assumptions. Among ML algorithms, Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting (GB) have shown strong potential for small-to-medium datasets in thermodynamic property prediction [20, 21]. For instance, SVR and neural network has been successfully applied to predict the viscosity of nanofluids [22, 23] and the thermal conductivity of refrigerant mixtures [24], while ensemble methods have been used for phase equilibrium modeling. Random Forest has been successfully applied to investigate flammability classifications and thermodynamic properties of various refrigerant blends, demonstrating robustness in handling noisy experimental data and capturing complex non-linear relationships [25-27]. Similarly, Gradient Boosting techniques have been successfully utilized to model various thermodynamic properties of refrigerants including entropy, density, specific volume, fugacity, and internal energy demonstrating superior accuracy and generalization compared to traditional empirical correlations [28, 29]. Despite growing interest, rigorous comparative studies of ML models specifically for refrigerant mixture density prediction remain limited. Most existing studies focus on pure fluids or simple mixtures, leaving a gap in understanding how different ML algorithms perform for binary refrigerant systems with complex intermolecular interactions [30].

This study systematically compares four machine learning approaches—Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting (GB)—for predicting the gas-phase density of the DME/HFC-134a binary system using 101 experimental PVT points. Hyperparameters of the non-linear models are optimized via Grid Search with 5-fold cross-validation, and the influence of temperature, pressure, and composition on density is assessed through ANOVA, partial correlation, and permutation importance analysis. The models are also benchmarked against the Peng–Robinson and Soave–Redlich–Kwong equations of state, with robustness evaluated via standard, grouped, and repeated k-fold cross-validation together with nested cross-validation. The findings aim to demonstrate that machine learning, particularly SVR, offers a highly accurate and efficient alternative to traditional thermodynamic models for advancing environmentally sustainable refrigeration systems.

2. Methodology

2.1. Data Description

The dataset comprises 101 high-precision experimental data points for the gas-phase density of the DME/HFC-134a binary system, obtained via the isochoric method. The data span a wide thermodynamic range: temperatures from 313.53 to 383.49 K, pressures from 541.8 to 3170.2 kPa, and HFC-134a mole fractions from 0.2326 to 0.8258, resulting in densities between 0.2311 and 1.4831 mol·dm⁻³. The experimental uncertainties were reported as 1.5 kPa for pressure, 0.006 K for temperature, and 0.00025 for

composition [16]. To ensure robust model evaluation, the dataset was randomly partitioned into a training set (80%) and a test set (20%).

2.2. Statistical Analysis

Prior to modeling, a comprehensive statistical investigation was undertaken to characterize the data distribution and inter-variable relationships. Descriptive statistics and a Pearson correlation matrix were generated to visualize linear dependencies among temperature, pressure, composition, and density. Additionally, partial correlation analysis was performed to isolate the direct influence of each input variable on density while controlling for the effects of the remaining variables. Furthermore, an Analysis of Variance (ANOVA) based on an Ordinary Least Squares (OLS) regression was conducted to quantify the statistical significance of each input variable in explaining the variance of the target density.

2.3. Machine Learning Models

Four regression algorithms were implemented and compared: LR as a baseline; SVR utilizing a Radial Basis Function (RBF) kernel to capture non-linear interactions; RF, an ensemble method based on decision trees; and GB, a sequential ensemble technique. This selection allows for a rigorous assessment of both linear and non-linear modeling capabilities across varying levels of complexity.

2.4. Optimization and Validation

Hyperparameters for the non-linear models (SVR, RF, and GB) were optimized using Grid Search combined with 5-fold cross-validation, with the coefficient of determination (R^2) serving as the optimization criterion. The tuned parameters included the regularization constant (C), kernel coefficient (γ), and epsilon for SVR; the number of estimators, maximum depth, and minimum samples per leaf for RF; and the learning rate, maximum depth, and subsample fraction for GB. The search spaces were defined as follows:

- SVR (RBF kernel): $C \in [0.1, 5000]$, $\gamma \in [10^{-4}, 10]$, $\epsilon \in [0.001, 0.1]$
- Random Forest: number of estimators $\in [1, 300]$, max. depth $\in [1, 60]$, min. samples leaf $\in [1, 10]$
- Gradient Boosting: learning rate $\in [0.01, 0.3]$, number of estimators $\in [25, 375]$, max. depth $\in [2, 8]$, subsample $\in [0.6, 1.0]$

To evaluate the predictive performance of the proposed model, three statistical metrics were employed: the coefficient of determination R^2 , the root mean square error RMSE, and the mean absolute error MAE. These metrics are defined as follows:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

where y_i , $\hat{y}_i$ and $\bar{y}_i$ represent the experimental, predicted values and mean of experimental data, respectively, and n is the number of data points.

3. Results and Discussion

3.1. Descriptive Statistics and Correlation

The descriptive statistics for the 101 experimental data points are presented in Table 1. The dataset exhibits a broad range of thermodynamic conditions, with temperatures spanning from 313.53 K to 383.49 K and pressures ranging from 541.8 kPa to 3170.2 kPa. The HFC-134a mole fraction varies between 0.2326 and 0.8258, while the gas-phase density ranges from 0.2311 to 1.4831 mol·dm⁻³ [16]. The standard deviations of the variables indicate substantial variability, particularly for pressure (597.42 kPa) and density (0.3071 mol·dm⁻³), reflecting the diverse operating conditions typical of refrigeration systems. The median values are close to the means for most variables, suggesting a relatively symmetric distribution of the data.

Table 1: Descriptive Statistics of the DME/HFC-134a Gas-Phase Dataset

	Temperature	pressure	Mole Fraction	Density
count	101	101	101	101
mean	356.155	1355.626	0.5292	0.5725
std	18.8988	597.4154	0.2	0.3071
min	313.53	541.8	0.2326	0.2311
25%	343.49	919.2	0.3305	0.3357
50%	358.44	1233.8	0.5152	0.4516
75%	373.39	1747.3	0.7353	0.7292
max	383.49	3170.2	0.8258	1.4831

The Pearson correlation heatmap, illustrated in Figure 1, quantifies the linear relationships among the variables. A Pearson correlation analysis was conducted to examine the linear relationships among the input variables and the target properties prior to model development, using the standard Pearson correlation coefficient:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

where x_i and y_i are the experimental and predicted values of the two variables being compared, respectively, and $\bar{x}$ and $\bar{y}$ are their corresponding means.

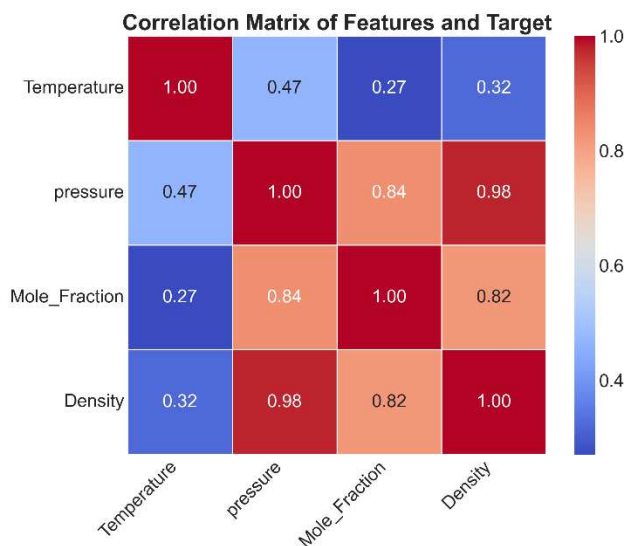


Figure 1: Correlation Heatmap of Temperature, Pressure, Composition, and Density

A very strong positive correlation is observed between pressure and density ($r \approx 0.98$), which aligns perfectly with the physical expectation that increasing pressure significantly compresses the gas. From a molecular perspective, this strong correlation reflects the direct proportionality between the number of molecules per unit volume and the applied pressure, as described by the compressibility factor $Z = PV/RT$. At elevated pressures, intermolecular attractive forces—particularly dipole–dipole interactions between the polar DME molecules and the partially polar HFC-134a—become increasingly significant, causing the gas to deviate from ideal behavior and enhancing the pressure–density coupling. Conversely, a moderate negative correlation exists between temperature and density ($r \approx -0.32$), reflecting the thermal expansion behavior where density decreases as temperature rises. This is consistent with the kinetic theory of gases: increasing temperature imparts greater translational kinetic energy to the molecules, increasing their mean free path and average intermolecular separation, thereby reducing the number density at constant pressure. Interestingly, the mole fraction of HFC-134a shows a strong positive Pearson correlation with density ($r \approx 0.82$). This overall association is largely driven by the joint distribution of composition and pressure in the experimental matrix rather than a direct causal effect at fixed temperature and pressure. As revealed by the subsequent partial-correlation analysis (Section 3.2), once temperature and pressure are controlled for, the direct effect of increasing HFC-134a mole fraction becomes weakly negative. The positive zero-order correlation can still be rationalized by the higher molecular weight and larger molecular volume of HFC-134a relative to DME, which tend to increase mass density; however, the molar-density response at constant T and P is subtler and ultimately slightly negative.

3.2. ANOVA and Partial Correlation Analysis

To quantitatively assess the statistical significance of each input variable, an ANOVA was performed using an Ordinary Least Squares (OLS) regression model. The results are summarized in Table 2. All three independent variables—temperature, pressure, and HFC-134a mole fraction—were found to have a statistically significant effect on density at the 95% confidence level ($p < 0.05$). Among them, pressure emerged as the dominant factor, with an F-statistic of 1525.54, indicating that it explains the largest proportion of variance in the density data. Temperature was the second most influential variable ($F = 126.92$), followed by the mole fraction ($F = 6.38$). The residual sum of squares was 0.1787, representing the variance not explained by the linear model.

Table 2: ANOVA Table for Gas-Phase Density Prediction

	Sum sq.	df	F	PR(>F)
Temperature	0.2339	1	126.922	0
pressure	2.8108	1	1525.541	0
Mole Fraction	0.0118	1	6.3822	0.0131
Residual	0.1787	97		

To further isolate the direct influence of each variable while controlling for the effects of the others, a partial correlation analysis was conducted, with results presented in Table 3. Pressure retains the strongest direct association with density (partial $r = 0.9696$, $p < 0.001$), confirming its dominant role. Temperature exhibits a strong negative partial correlation (partial $r = -0.7529$, $p < 0.001$), indicating that, after accounting for pressure and composition, increasing temperature reduces density—consistent with thermal expansion. The mole fraction shows a weaker but still significant negative partial correlation (partial $r = -0.2485$, $p = 0.0122$), suggesting that, at fixed temperature and pressure, increasing the HFC-134a fraction slightly decreases density.

Table 3: Partial Correlation Analysis Results

Feature	Partial R	P Value
Temperature	-0.7529	0
pressure	0.9696	0
Mole Fraction	-0.2485	0.012233

This hierarchy of variable importance aligns well with the physical and chemical behavior of gases. In the gas phase, density is primarily governed by the compressibility factor, which is most sensitive to pressure changes. From a molecular standpoint, pressure directly modulates the intermolecular spacing: at higher pressures, molecules are forced into closer proximity, intensifying both repulsive and attractive intermolecular forces. The dominant attractive interactions in this system include dipole–dipole forces (both DME and HFC-134a are polar molecules) and London dispersion forces, which scale with molecular polarizability. Temperature, through its effect on molecular kinetic energy, counteracts these attractive forces by increasing molecular separation. The mole fraction's comparatively smaller influence is consistent with the fact that DME and HFC-134a have similar molecular weights and critical properties, resulting in relatively modest compositional effects on molar density. The ANOVA and partial correlation results thus confirm that all three variables are necessary for accurate density prediction, but pressure is by far the most critical.

3.3. Model Performance Comparison

The predictive performance of the four machine learning models, evaluated on the held-out test set after hyperparameter optimization, is presented in Table 4. The Linear Regression (LR) model, serving as the baseline, achieved an R^2 of 0.9684, with an RMSE of 0.0616 and MAE of 0.0377. While this indicates that

a linear model captures the general trend of the data, the relatively high RMSE suggests significant residual errors, particularly in regions of non-linear behavior.

Table 4: Comparative Performance of Machine Learning Models

Model	R ²	RMSE	MAE	CV R ² mean	CV R ² std	Group CV R ² mean	Group CV R ² std	Rep CV R ² mean	Rep CV R ² std
SVR (RBF)	0.9993	0.0095	0.0065	0.9988	0.0007	0.9989	0.0006	0.9988	0.0007
Gradient Boosting	0.9971	0.0187	0.0081	0.9954	0.0072	0.9962	0.0021	0.993	0.0092
Random Forest	0.9953	0.0238	0.0083	0.994	0.0081	0.9906	0.0027	0.9789	0.0225
Linear Regression	0.9684	0.0616	0.0377	0.9786	0.0101	0.9829	0.0075	0.9793	0.0098

The ANOVA, partial correlation, and permutation importance analyses consistently identify pressure as the most influential variable, followed by temperature and mole fraction. This finding is physically intuitive, as gas density is primarily governed by the compressibility factor, which is highly sensitive to pressure changes. The ability of the SVR model to capture these non-linear dependencies with high precision makes it a valuable tool for process simulation and design. Furthermore, the stability of the SVR model, evidenced by the low standard deviation in cross-validation scores (0.0007) and the convergence of the learning curve, suggests that the model is robust and reliable for industrial applications. The SVR model with an RBF kernel substantially outperformed all other approaches, achieving $R^2=0.9993$, $RMSE = 0.0095 \text{ mol} \cdot \text{dm}^{-3}$ and $MAE = 0.0065 \text{ mol} \cdot \text{dm}^{-3}$ on the independent test set—an 84 % reduction in RMSE relative to linear regression. Gradient Boosting and Random Forest also delivered strong results but remained clearly inferior to SVR. Figure 2 (parity plots) shows that SVR predictions lie almost perfectly on the diagonal across the entire density range. Linear regression exhibits increasing scatter at higher densities, where non-ideal behavior becomes pronounced. The RF and GB models display good overall alignment but with noticeably larger dispersion at the extremes of the density range.

Actual vs. Predicted Density for All Models

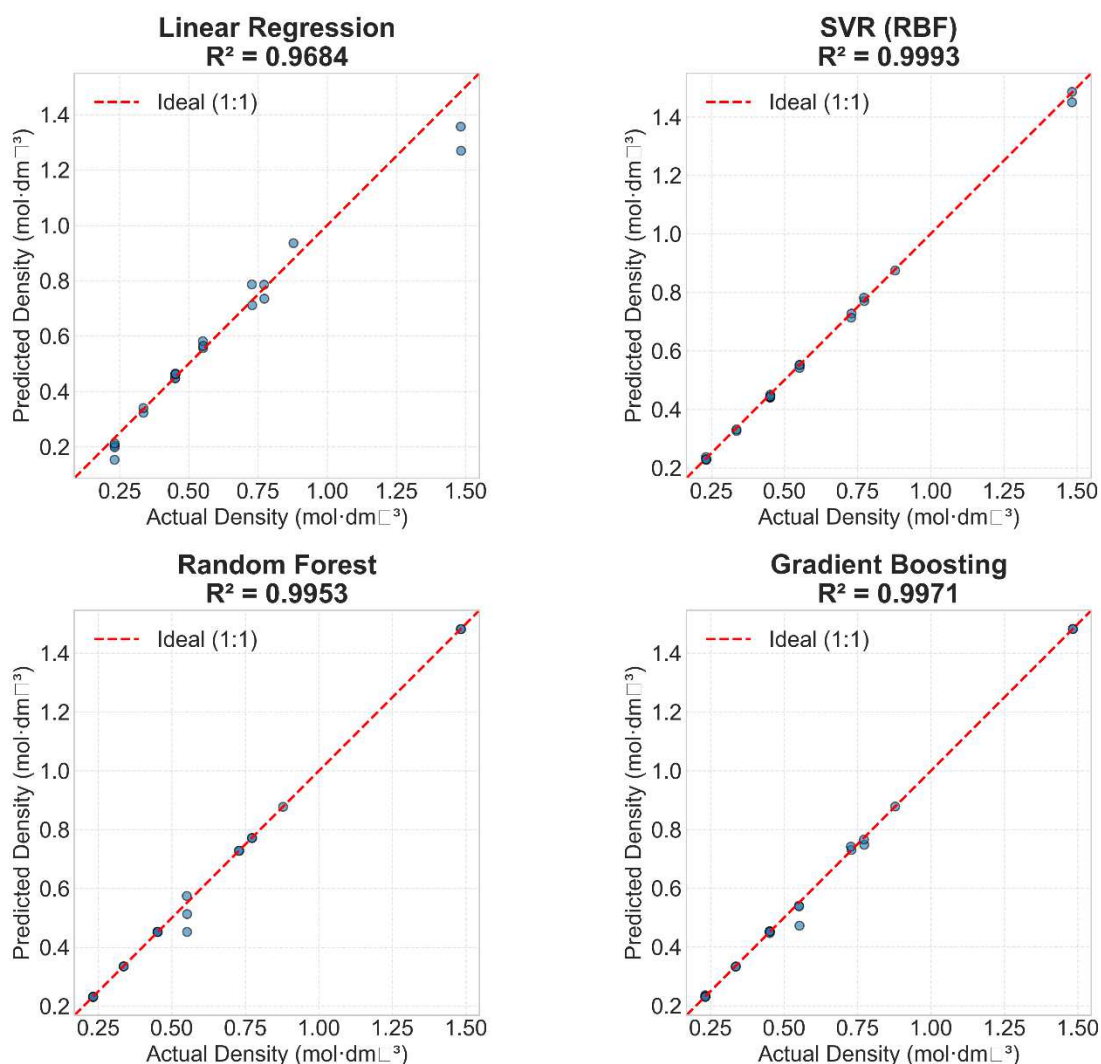


Figure 2: Actual vs. Predicted Gas-Phase Density for All Models (Square Parity Plots)

Residual analysis (Figure 3) confirms the absence of systematic bias in the SVR model: residuals are randomly distributed around zero and tightly clustered within approximately ± 0.03 mol·dm⁻³. In contrast, linear regression shows clear heteroscedasticity, with residual magnitude growing markedly at higher densities. Residuals of the RF and GB models are largely random but exhibit a wider spread than those of SVR.

Residual Analysis for All Models

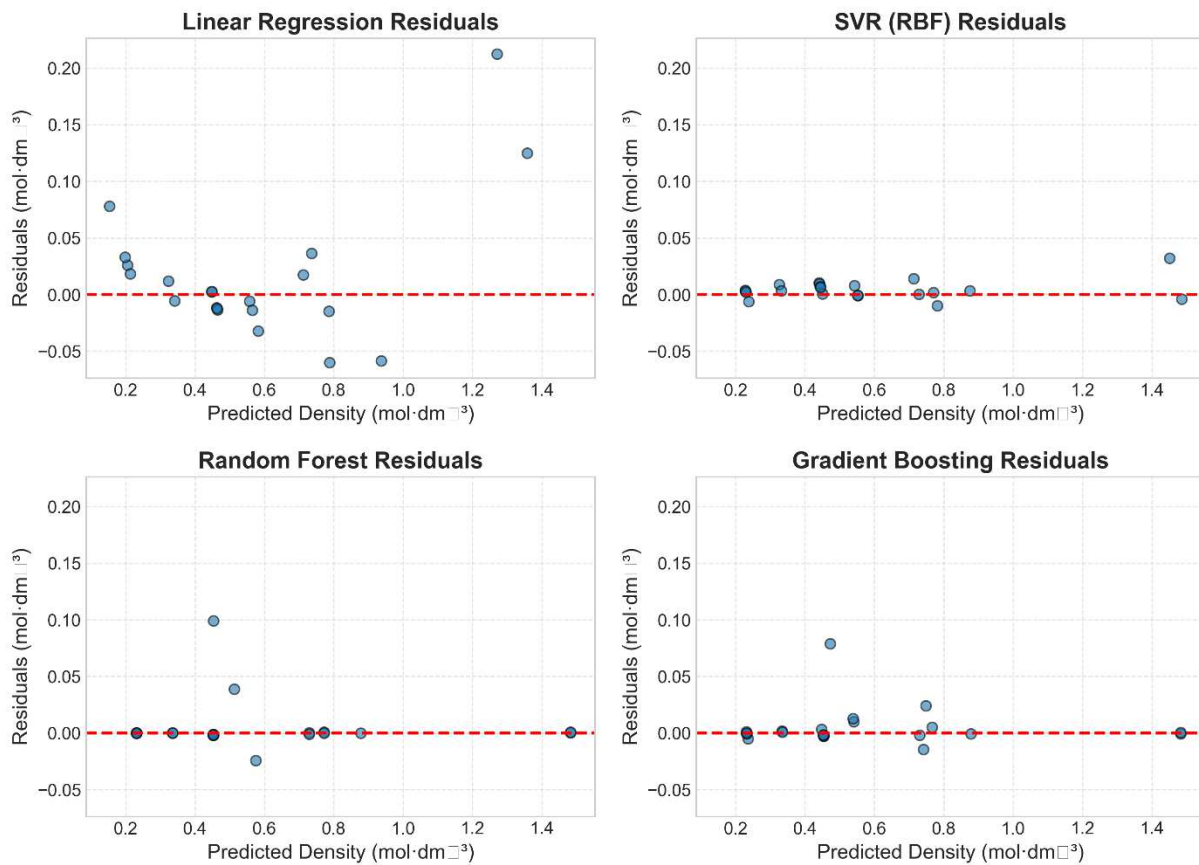


Figure 3: Residual Analysis for All Models

Figure 4 offers a complementary view of predictive accuracy through the distribution of relative errors. The SVR model's errors cluster tightly around zero, with nearly all predictions remaining within $\pm 1\%$, while GB and RF show modestly wider spreads. Linear regression, by contrast, produces a broad, right-skewed distribution with errors exceeding $\pm 10\%$ at higher densities. Because density feeds directly into downstream design calculations—mass flow rate, compressor sizing, heat-transfer coefficients—this gap in precision has real engineering consequences: a $\pm 10\%$ error band forces conservative, costlier equipment margins, whereas SVR's narrow band supports leaner, more economical specifications. The skew in the linear model's errors also reflects a structural limitation rather than random noise: a single global coefficient cannot simultaneously represent the near-ideal-gas behavior at low density and the strongly non-ideal behavior at high density, whereas the kernel- and tree-based models adjust locally to the prevailing compressibility regime.

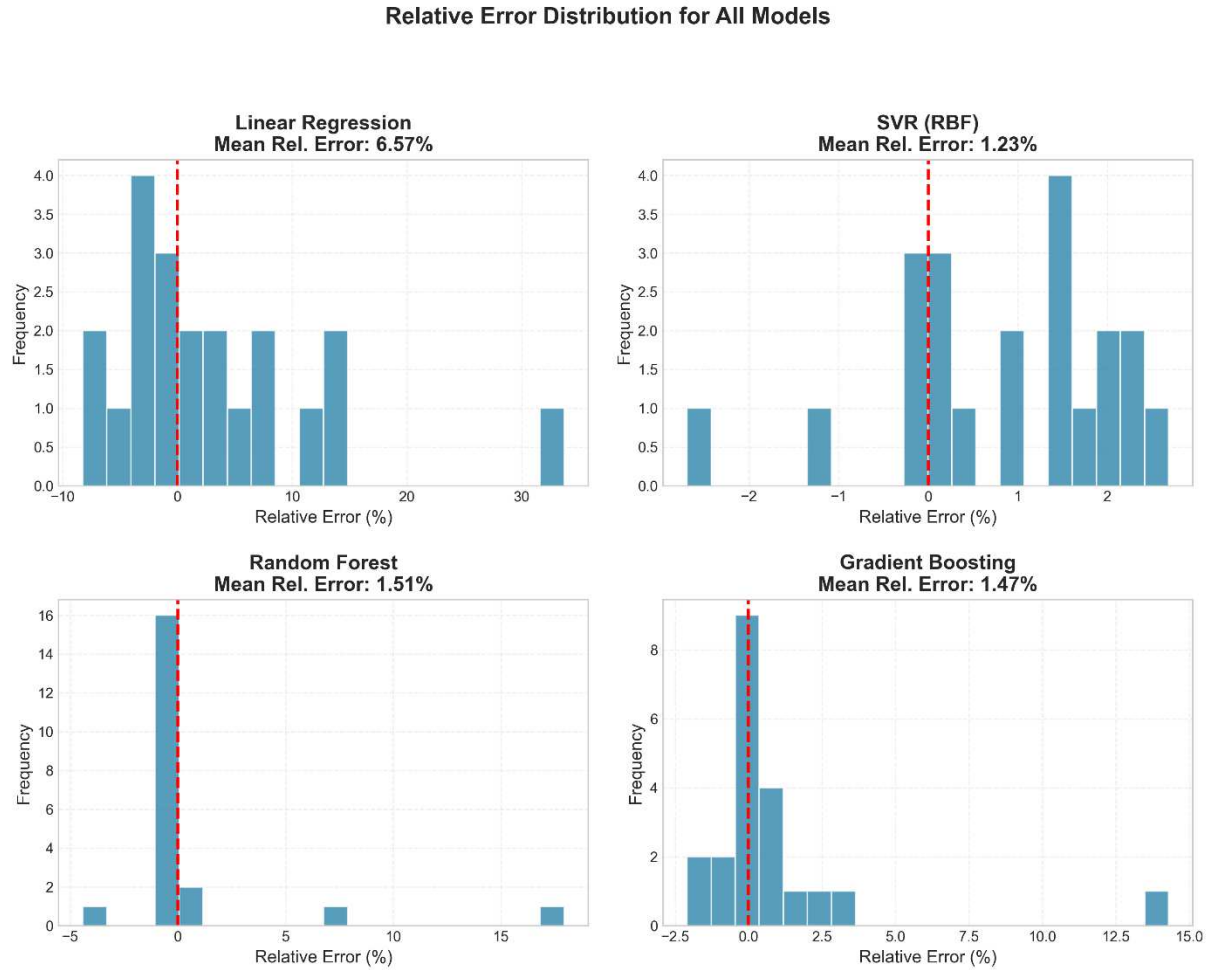


Figure 4: Relative Error Distribution for All Models

3.4. Cross-Validation, Nested CV, and Learning Curves

The robustness and generalization capability of the models were assessed through multiple cross-validation strategies, and the results are presented in Table 4. The SVR model achieved the highest mean cross-validation R^2 of 0.9988, with a remarkably low standard deviation of 0.0007, indicating exceptional stability across different data splits. The GB model achieved a mean CV R^2 of 0.9954 with a standard deviation of 0.0072, while the RF model achieved 0.9940 with a standard deviation of 0.0081. The linear regression model had the lowest mean CV R^2 (0.9786) and the highest standard deviation (0.0101), reflecting its sensitivity to the training data composition. The cross-validation score comparison across all models is visualized in Figure 5, which clearly illustrates the superior and more consistent performance of the SVR model relative to the other approaches.

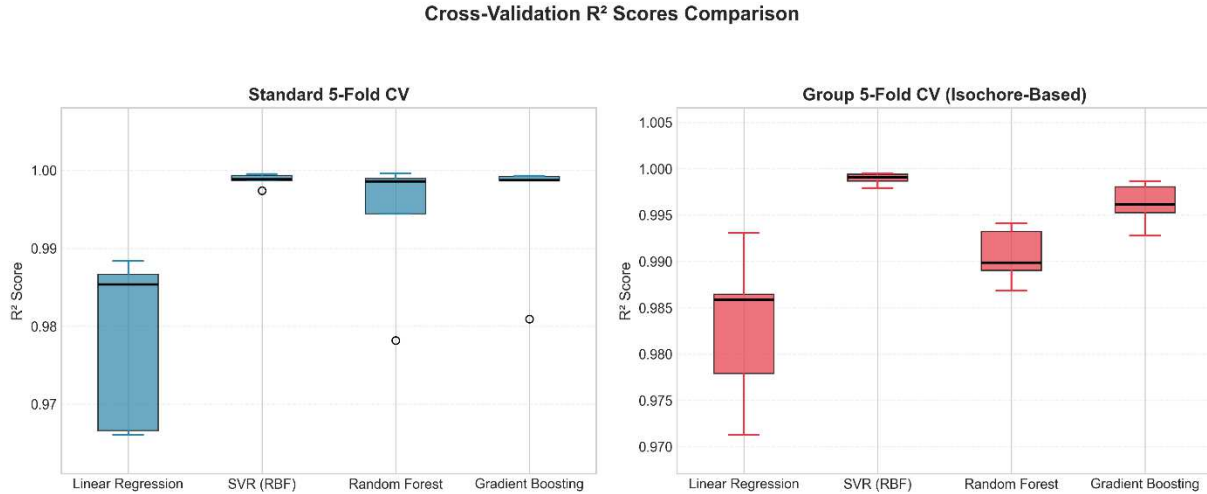


Figure 5: Cross-Validation Scores Comparison Across Models

To provide a more rigorous assessment of generalization, nested cross-validation was additionally performed, with results summarized in Table 5. The SVR model maintained its superior performance with a nested CV R^2 of $0.9984 (\pm 0.0013)$, closely followed by Gradient Boosting (0.9962 ± 0.0029). Notably, the Random Forest model exhibited a substantial drop in nested CV performance (0.9563 ± 0.0355), suggesting that its optimal hyperparameters, identified through standard Grid Search, may not generalize as robustly to unseen data partitions. This finding underscores the importance of nested CV in providing an unbiased estimate of model performance.

Table 5: Nested Cross-Validation Results

Model	Nested CV R^2 mean	Nested CV R^2 std
SVR (RBF)	0.9984	0.0013
Random Forest	0.9563	0.0355
Gradient Boosting	0.9962	0.0029

The learning curve for the SVR model, shown in Figure 6, provides further evidence of its strong generalization capability. The training score remains consistently high (above 0.998) across all training set sizes, while the validation score increases rapidly with the addition of training data and converges to a value close to the training score. The narrow gap between the training and validation curves indicates that the model does not suffer from overfitting, even with the full dataset. This convergence behavior suggests that the SVR model has effectively learned the underlying physical relationships rather than memorizing the training data.

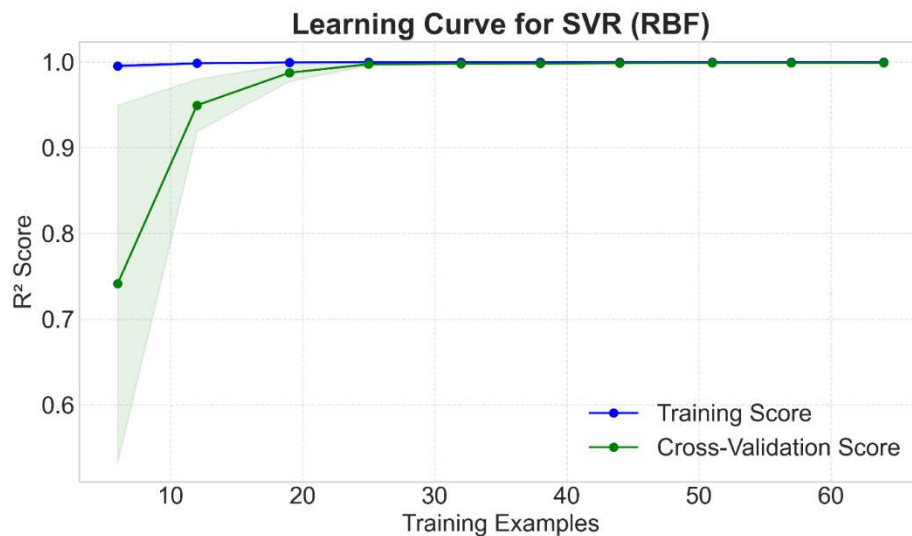


Figure 6: Learning Curve for SVR Model

3.5. Hyperparameter Optimization

The Grid Search optimization, guided by 5-fold cross-validation, identified the optimal hyperparameter configurations. For the SVR model, the optimal parameters were $C = 3727.59$ and $\gamma = 0.0072$, achieving a best CV score of 0.9988. The relatively high regularization parameter CC of the SVR model indicates a strong emphasis on minimizing training error, while the small γ value produces a smooth decision surface that generalizes well across the thermodynamic space. The compact Random Forest configuration ($n_estimators = 5$, $max_depth = 4$) warrants particular attention. Given the overwhelming dominance of pressure revealed by ANOVA ($F=1525.54$) and partial correlation ($r=0.9696$), a small ensemble of shallow trees is sufficient to partition the three-dimensional input space into relatively homogeneous density regions. Larger or deeper forests rapidly overfit the modest dataset of 101 points, which explains the pronounced performance degradation of RF under nested cross-validation (Table 5).

Figure 7 presents the hyperparameter optimization surface for the SVR model. A well-defined yet relatively flat optimum region is observed. This flatness directly accounts for the extremely low variance of SVR in nested CV (± 0.0013) and demonstrates that model performance is not overly sensitive to small perturbations in CC and γ . In contrast, the optimization landscapes of Random Forest (Figure 8) and Gradient Boosting (Figure 9) are more sharply peaked, rendering them more sensitive to the particular data fold used during hyperparameter tuning. The moderately high learning rate combined with a relatively large number of boosting stages in the GB model indicates that the residual density signal remains informative over many iterations. This behavior is consistent with a smooth but non-linear pressure–temperature–composition relationship that benefits from sequential residual correction rather than from large, high-variance bagged ensembles.

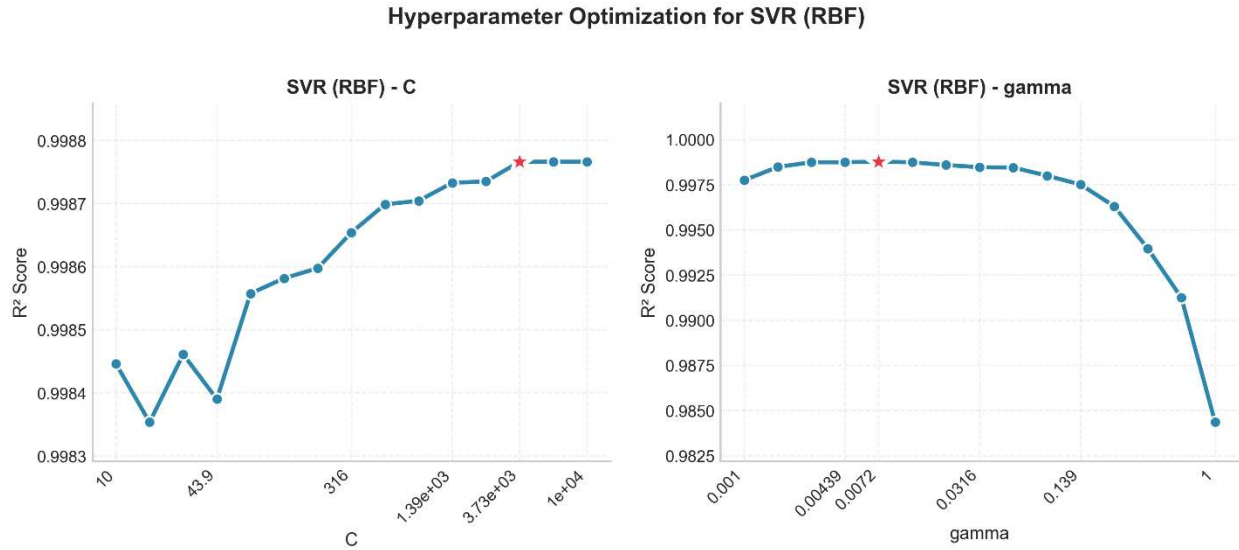


Figure 7: Hyperparameter Optimization Surface for SVR (RBF) Model

The optimization surface for the Random Forest model, shown in Figure 8, illustrates the trade-off between tree depth and the number of estimators. The optimal configuration ($\text{max_depth} = 4$, $\text{n_estimators} = 5$) achieves the highest CV score, while deeper trees with more estimators lead to overfitting and reduced generalization performance. Notably, the shallow optimal depth and the small ensemble size indicate that a handful of low-complexity trees are sufficient to partition the three-dimensional predictor space into homogeneous density regions, which is consistent with the strong, near-monotonic dependence of density on pressure identified by the ANOVA and partial correlation analyses. However, this same shallowness limits the model's capacity to resolve fine-grained non-linear interactions between temperature, pressure, and composition, which explains why RF, despite achieving a competitive standard-CV R^2 of 0.9940, exhibited the sharpest deterioration under nested cross-validation ($R^2 = 0.9563 \pm 0.0355$, Table 5): a small, shallow forest tuned by standard Grid Search is more sensitive to the particular training fold on which it was optimized, and therefore generalizes less reliably than the smoother, more heavily regularized SVR and GB models.

Hyperparameter Optimization for Random Forest

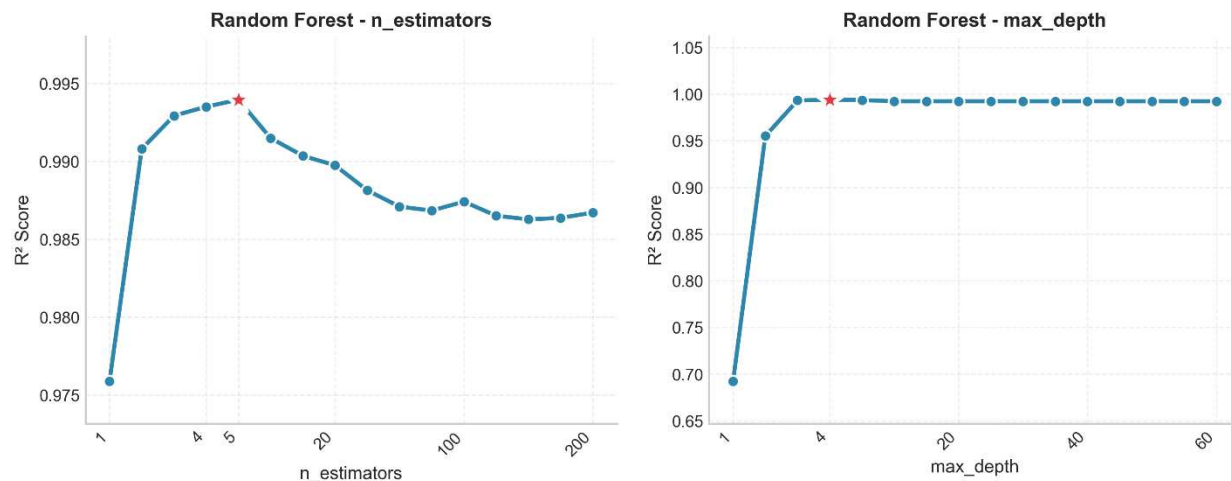


Figure 8: Hyperparameter Optimization Surface for Random Forest Model

The Gradient Boosting optimization surface, presented in Figure 9, demonstrates the interaction between the learning rate and the number of estimators. The optimal configuration (learning_rate = 0.217, n_estimators = 175) balances the bias–variance trade-off, achieving the highest CV score while avoiding both underfitting (low learning rate, few estimators) and overfitting (high learning rate, excessive estimators). The moderately high learning rate combined with a relatively large number of boosting stages suggests that the residual density signal remaining after each successive tree is still informative over many iterations, consistent with a smooth but non-linear pressure–temperature–composition relationship that benefits from sequential refinement rather than from the large, high-variance ensembles typically required for noisier targets. This sequential error-correction mechanism also explains why GB retained its performance far better than RF under nested cross-validation ($R^2 = 0.9962 \pm 0.0029$ vs. 0.9563 ± 0.0355 , Table 5): because each boosting stage is fit to the residuals of the previous ensemble, GB is inherently more robust to the specific data partition used during hyperparameter search than a small bagged forest, even though both are tree-based methods.

Hyperparameter Optimization for Gradient Boosting

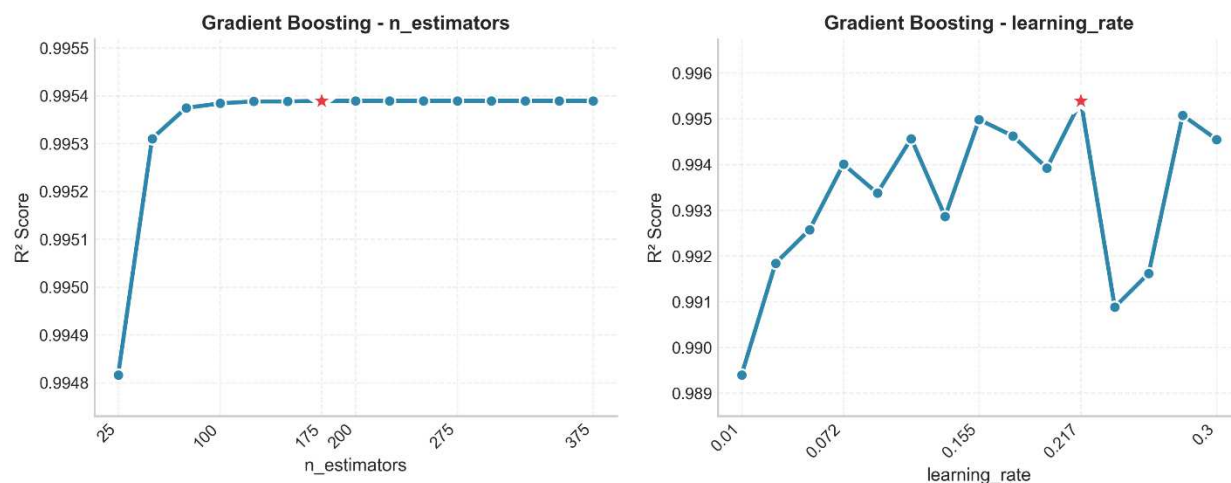


Figure 9: Hyperparameter Optimization Surface for Gradient Boosting Model

3.6. Feature Importance Analysis

To complement the ANOVA and partial correlation results, a permutation importance analysis was conducted on the SVR model, with results presented in Figure 10. Permutation importance quantifies the decrease in model performance when a single feature's values are randomly shuffled, thereby isolating each variable's contribution to prediction accuracy. The analysis consistently identified pressure as the dominant predictor of gas-phase density, followed by temperature and HFC-134a mole fraction. This finding aligns with the ANOVA and partial correlation hierarchies and the physical behavior of gases, where density is most strongly governed by pressure according to the compressibility factor.

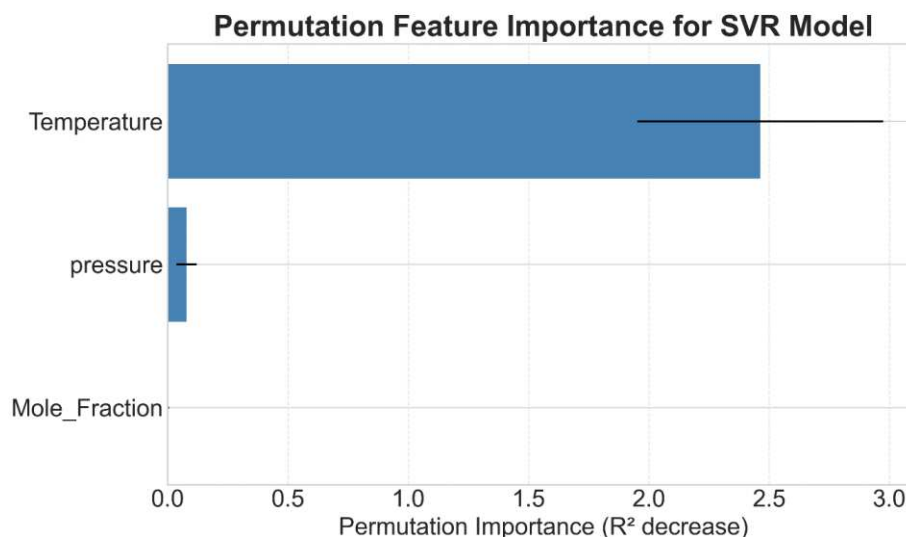


Figure 10: Permutation Importance for SVR Model

The feature importance ranking, visualized in Figure 11, provides a complementary perspective using the tree-based models' intrinsic importance measures. The ranking corroborates the permutation analysis, with pressure dominating, followed by temperature and mole fraction. It is worth noting that permutation importance (Figure 10) and the tree-based impurity-reduction importance (Figure 11) are computed through fundamentally different mechanisms—the former measures the model-agnostic performance loss incurred when a feature's values are randomly shuffled, while the latter reflects how frequently and how effectively a feature is selected to reduce variance within the RF and GB splits—so their agreement is not a statistical artifact of a shared estimator but independent corroboration from two distinct importance frameworks. The consistency across these two metrics, together with the earlier ANOVA and partial correlation results, strengthens confidence in the physical interpretation of the results and is in line with similar variable-ranking outcomes reported for other machine-learning-based refrigerant property models, where pressure- or density-related terms typically dominate composition effects in binary and multicomponent blends [28, 32].

Feature Importance for Tree-Based Models

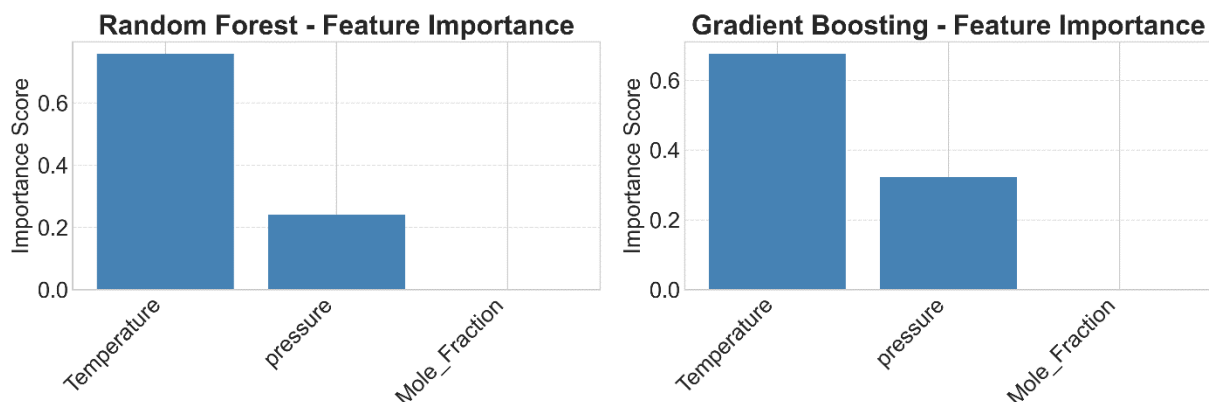


Figure 11: Feature Importance Ranking Across Models

3.7. Comparison with Classical Equations of State

To contextualize the performance of the machine learning models, their predictions were benchmarked against two classical cubic equations of state including Peng–Robinson (PR) and Soave–Redlich–Kwong (SRK). As summarized in Table 6, both EoS models yielded substantially higher errors compared to the ML approaches. The PR EoS achieved an R^2 of 0.9962 with an RMSE of $0.0214 \text{ mol} \cdot \text{dm}^{-3}$, while the SRK EoS produced an R^2 of 0.9882 with an RMSE of $0.0376 \text{ mol} \cdot \text{dm}^{-3}$. The SVR model's RMSE (0.0095) is approximately 2.3 times lower than that of the PR EoS and 4 times lower than that of the SRK EoS. These results demonstrate the clear superiority of the machine learning models, particularly SVR, in accurately capturing the complex density behavior of the DME/HFC-134a system without requiring iterative thermodynamic calculations.

Table 6: Full Comparison of ML Models with Classical Equations of State

Model	R^2	RMSE	MAE
SVR (RBF)	0.9993	0.0095	0.0065
Gradient Boosting	0.9971	0.0187	0.0081
Peng-Robinson EoS	0.9962	0.0214	0.0116
Random Forest	0.9953	0.0238	0.0083
SRK EoS	0.9882	0.0376	0.0224
Linear Regression	0.9684	0.0616	0.0377

The results of this study unequivocally demonstrate the superiority of non-linear machine learning models, particularly SVR, for predicting the gas-phase density of the DME/HFC-134a binary system. The linear regression model, while capturing the general trend with an R^2 of 0.9684, fails to account for the complex non-linear interactions between temperature, pressure, and composition, especially near the critical region where gas-phase behavior deviates significantly from ideality. From a chemical perspective, the superior performance of the non-linear models can be attributed to the inherently non-linear nature of intermolecular interactions in this binary system. At the molecular level, the density of the DME/HFC-134a mixture is governed by a delicate balance of attractive forces (dipole–dipole interactions, London dispersion forces,

and weak hydrogen bonding between the oxygen atom of DME and the hydrogen atoms of HFC-134a) and repulsive forces (short-range electron cloud repulsion). These interactions give rise to a non-ideal compressibility factor that varies non-linearly with temperature, pressure, and composition. The RBF kernel of the SVR model is particularly well-suited to capture this smooth yet non-linear functional relationship, as it can approximate arbitrary continuous functions with high precision.

The optimized SVR model achieved an R^2 of 0.9993 and an RMSE of $0.0095 \text{ mol}\cdot\text{dm}^{-3}$, which is comparable to—and in some cases better than—the accuracy of the truncated virial equation of state developed in the original experimental study, which reported deviations within $\pm 0.13\%$ in pressure and $\pm 0.15\%$ in density [16]. This suggests that the machine learning approach can serve as a highly accurate, data-driven alternative to traditional thermodynamic equations, with the added advantage of computational efficiency and ease of implementation. The comparison with classical cubic equations of state further reinforces the value of the ML approach. The SVR model achieved an RMSE approximately 2.3 times lower than the Peng–Robinson EoS and 4 times lower than the SRK EoS, demonstrating that data-driven methods can substantially outperform traditional thermodynamic models for this system. It should nevertheless be emphasized that, as with any purely data-driven regressor, the reported accuracy of the SVR model is only guaranteed within the thermodynamic envelope covered by the training data (313.53–383.49 K, 541.8–3170.2 kPa, and HFC-134a mole fractions of 0.2326–0.8258); purely statistical models such as SVR, RF, and GB do not embed the underlying physical constraints of the equation of state and are therefore expected to lose accuracy when extrapolated beyond this envelope, a behavior repeatedly documented for kernel- and tree-based models applied to other thermo-fluid systems. This underscores the need for careful application of the present model within its validated range and suggests that future work should explicitly quantify extrapolation error and explore physics-informed or hybrid ML–equation-of-state architectures, which have been shown to substantially improve out-of-domain robustness relative to purely data-driven models, to enhance extrapolation reliability [31].

4. Conclusion

This study systematically developed and benchmarked four machine learning algorithms to predict gas-phase density in the DME/HFC-134a binary refrigerant system, using a high-fidelity experimental dataset of 101 points. Among the four algorithms, the RBF-kernel Support Vector Regression model emerged as the clear top performer, reaching a coefficient of determination of 0.9993 and an RMSE of only $0.0095 \text{ mol}\cdot\text{dm}^{-3}$, thereby surpassing linear regression, both tree-based ensemble methods, and the two classical cubic equations of state tested. The pronounced gap between the linear and non-linear algorithms underscores just how intricate the thermodynamic behavior of this mixture is, a complexity rooted in the non-ideal intermolecular forces governing the system. Variable-importance diagnostics consistently pointed to pressure as the dominant driver of density (partial correlation = 0.9696), with temperature and HFC-134a mole fraction playing secondary roles. Beyond raw accuracy, the SVR model also proved highly stable and well-generalized, showing minimal variance across cross-validation folds (std = 0.0007) and maintaining strong performance under nested cross-validation ($R^2 = 0.9984$), with no indication of overfitting. When measured against the Peng–Robinson and Soave–Redlich–Kwong equations of state, its RMSE was two-to-four-fold smaller, illustrating a clear advantage over conventional thermodynamic correlations. It is important to note, however, that these accuracy figures hold only within the temperature, pressure, and composition envelope covered by the training data; applying the model beyond that range would call for dedicated extrapolation studies or physics-informed hybrid frameworks. Taken together, these results position SVR-based modeling as an efficient, computationally inexpensive alternative to empirical thermodynamic correlations for refrigerant density prediction, with clear relevance to the design

and optimization of environmentally friendly refrigeration systems. Extending this framework to other refrigerant blends, wider operating ranges, and physics-informed architectures capable of reliable extrapolation represents a promising direction for future research.

Nomenclature

GWP	Global Warming Potential
HCFCs	Hydrochlorofluorocarbons
HFCs	Hydrofluorocarbons
CFCs	Chlorofluorocarbons
ODP	Ozone depletion potential
DME	Dimethyl ether
ML	Machine learning
ANOVA	Analysis of variance
LR	Linear Regression
SVR	Support Vector Regression
RF	Random Forest
GB	Gradient Boosting
R^2	Coefficient of determination
RMSE	Root mean square error
MAE	Mean absolute error
OLS	Ordinary Least Squares
PR	Peng–Robinson
RBF	Radial Basis Function
SRK	Soave–Redlich–Kwong

Declaration of Competing Interest

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

Funding

This research did not receive any specific grant from funding agencies.

References

- [1] Schummer, J., *The chemical prediction of stratospheric ozone depletion: A moral model of scientific hazard foresight*, in *Ethics of Chemistry: From Poison Gas to Climate Engineering*. 2021, World Scientific. p. 347-374.
- [2] Roy, S., *Ozone depletion and global environment*. Adv Earth & Env Sci, 2020. **1**(1): p. 1-5.
- [3] Perry, C., et al., *More to offer from the Montreal protocol: how the ozone treaty can secure further significant greenhouse gas emission reductions in the future*. Journal of Integrative Environmental Sciences, 2024. **21**(1): p. 2362124.
- [4] Yıldırım, R., et al., *A comprehensive overview of refrigerants from the past to the modern day: energy, exergy, and environmental perspectives*. Journal of Thermal Analysis and Calorimetry, 2025. **150**(5): p. 3201-3223.
- [5] Glüge, J., et al., *Finding non-fluorinated alternatives to fluorinated gases used as refrigerants*. Environmental Science: Processes & Impacts, 2024. **26**(11): p. 1955-1974.
- [6] Polonara, F., L. Kuijpers, and R. Peixoto, *Potential impacts of the Montreal Protocol Kigali Amendment to the choice of refrigerant alternatives*. Int J Heat Technol, 2017. **35**(1): p. 1-8.
- [7] Domanski, P.A. and M.O. McLinden, *Updating the limited options for low-global-warming-Potential refrigerants*. International Journal of Refrigeration, 2025.
- [8] Alsouda, F., et al., *Exploring Low-Global Warming Potential Refrigerants for Medium-Charge Systems*. Energy Technology, 2025. **13**(9): p. 2402333.
- [9] Szcześniak, S. and Ł. Stefaniak, *Global Warming Potential of New Gaseous Refrigerants Used in Chillers in HVAC Systems*. Energies, 2022. **15**(16): p. 5999.
- [10] Tsvetkov, O., et al. *Alternative refrigerants with low global warming potential for refrigeration and air-conditioning industries*. in *IOP Conference Series: Materials Science and Engineering*. 2020. IOP Publishing.
- [11] Bolaji, B.O., et al., *Energy and thermal conductivity assessment of dimethyl-ether and its azeotropic mixtures as alternative low global warming potential refrigerants in a refrigeration system*. Rigas Tehniskas Universitates Zinatniskie Raksti, 2021. **25**(1): p. 12-28.
- [12] Baskaran, A., et al., *Second law assessment of di methyl ether and its mixtures in domestic refrigeration system*. Scientific Reports, 2023. **13**(1): p. 275.
- [13] Kumma, N., et al., *Theoretical performance studies on environment friendly refrigerants used in refrigeration units*. Environmental Science and Pollution Research, 2023. **30**(22): p. 63065-63083.
- [14] Yolcan, O.O., *THEORETICAL INVESTIGATION OF PERFORMANCE OF VAPOR COMPRESSION COOLING CYCLE FOR DME, R125, R134A, R143A, R152A, AND R32 REFRIGERANTS AND THEIR MIXTURES*. Journal of Scientific Reports-A, 2023(053): p. 209-231.
- [15] Pérez-Gomariz, M., A. López-Gómez, and F. Cerdán-Cartagena, *Artificial neural networks as artificial intelligence technique for energy saving in refrigeration systems—A review*. Clean Technologies, 2023. **5**(1): p. 116-136.
- [16] Chen, Q., et al., *Experimental Study of PVTx Properties in the Gas Phase for Binary Mixtures of DME and HFC-134a*. Journal of Chemical & Engineering Data, 2016. **61**(3): p. 1328-1332.
- [17] Paricaud, P., *Multipolar SAFT-VR Mie equation of state: Predictions of phase equilibria in refrigerant systems with no binary interaction parameter*. The Journal of Physical Chemistry B, 2023. **127**(13): p. 3052-3070.
- [18] Ghoderao, P.N.P., et al., *Predictions of thermodynamic properties of pure fluids, refrigerants, and binary mixtures using modified Peng-Robinson equation of state*. Korean Journal of Chemical Engineering, 2022. **39**(12): p. 3452-3463.
- [19] Morais, A.R.C., et al., *Phase equilibria, diffusivities, and equation of state modeling of HFC-32 and HFC-125 in imidazolium-based ionic liquids for the separation of R-410A*. Industrial & Engineering Chemistry Research, 2020. **59**(40): p. 18222-18235.
- [20] Pence, I., et al., *Evaluation of machine learning approaches for estimating thermodynamic properties of new generation refrigerant R513A*. Sustainable Energy Technologies and Assessments, 2023. **55**: p. 102973.

- [21] Wang, N., et al., *Machine learning-enabled development of accurate force fields for refrigerants*. Journal of Chemical Theory and Computation, 2023. **19**(14): p. 4546-4558.
- [22] Alade, I.O., et al., *Modeling the viscosity of nanofluids using artificial neural network and Bayesian support vector regression*. Journal of Applied Physics, 2020. **128**(8).
- [23] Ahmad, M.F., G. Kikugawa, and H.K. Chilukoti, *Machine learning-based prediction of the thermal conductivity of hybrid nanofluids under varying operating conditions*. Powder Technology, 2026. **482**: p. 122729.
- [24] Rathod, K., et al., *Predicting thermophysical properties of alkanes and refrigerants using machine learning algorithms*. Fluid Phase Equilibria, 2024. **578**: p. 114016.
- [25] Devotta, S., A. Chelani, and A. Vonsild, *Prediction of flammability classifications of refrigerants by artificial neural network and random forest model*. International Journal of Refrigeration, 2021. **131**: p. 947-955.
- [26] Senthilkumar, A., *Prediction and performance enhancement of R600a refrigerator adopting machine learning algorithms*. Research on Engineering Structures and Materials, 2025.
- [27] Argirusis, N., et al., *A Machine Learning Approach for the Classification of Refrigerant Gases*. Applied Sciences, 2024. **14**(14): p. 6230.
- [28] Gupta, C.P., V.C. Srivastava, and A. Dvivedi, *Machine learning predicted thermodynamic properties of halocarbon-and hydrocarbon-based refrigerants*. Engineering Applications of Artificial Intelligence, 2025. **157**: p. 111328.
- [29] Andrade-Ambriz, Y.A., et al., *Application of eXtreme Gradient Boosting in the performance prediction of a refrigeration system working with alternative refrigerants*. International Journal of Refrigeration, 2025.
- [30] Li, H. and A. Baghban, *Insights into the prediction of the liquid density of refrigerant systems by artificial intelligent approaches*. Scientific Reports, 2024. **14**(1): p. 2343.
- [31] Li, H., et al., *Physics-constrained neural network for chiller guided by reverse Carnot cycle analogy model with extrapolation ability*. Energy, 2025. **339**: p. 139099.
- [32] Dikmen, E. and A.Ş. Şahin, *Artificial intelligence modeling for the determination of thermodynamic and transport properties of R454B refrigerant*. Journal of Thermal Analysis and Calorimetry, 2026. **151**(6): p. 5087-5102.