INTERNATIONAL JOURNAL OF AUTOMOTIVE AND MECHANICAL ENGINEERING
ISSN: 2229-8649 e-ISSN: 2180-1606
VOLUME 23, ISSUE 1, 2026, 13405 - 13418
DOI: https://doi.org/10.15282/ijame.23.1.2026.19.1017

RESEARCH ARTICLE

Machine learning-based prediction of the coefficient of performance for low global warming potential refrigerants in a vapor compression system

Nguyen Duy Tue[1], **Vo Van An**[2]*

[1]Faculty of Mechanical, Electrical and Computer Engineering, Van Lang School of Technology, Van Lang University, Ho Chi Minh City 70000, Vietnam.

[2]Institute of Engineering Technology, Thu Dau Mot University, Ho Chi Minh City 70000, Vietnam

Abstract – This study addresses the energy significance of the coefficient of performance (COP) in vapor compression systems and the practical need to forecast COP quickly and reliably. Because COP directly reflects the amount of cooling delivered per unit of input power, accurate prediction supports energy savings, refrigerant selection, and early-stage design decisions, especially for low global warming potential (GWP) refrigerants. Authors develop data-driven models to estimate COP without full thermodynamic calculations. A synthetic dataset of 2,000 samples is generated in the Engineering Equation Solver (EES) for four refrigerants (R1234yf, R134a, R290, R600a) by using five inputs: refrigerant type, evaporation temperature, condensing temperature, subcooling, and superheat. Five supervised learning algorithms are trained and compared: linear regression, polynomial regression, random forests, decision trees, and support vector machines. The study evaluates model performance using the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) based on an 80/20 train/test split. Results show Polynomial Regression (degree 3) delivers the highest accuracy ($R^2 \approx 0.9999$; RMSE $\approx$ 0.0071; MAE $\approx$ 0.0053), with Random Forest as the next strongest baseline. The findings suggest that lightweight, well-tuned regressors can provide fast, precise COP predictions, reduce analysis time, and guide system design and parameter optimization. The approach offers an accessible tool for engineers seeking efficient, low-carbon refrigeration solutions.

Article History

Received	:	30 April 2025
Revised	:	05 February 2025
Accepted	:	24 February 2026
Published	:	15 March 2026

Keywords
Low-GWP refrigerant
Machine learning
Thermodynamic calculation
Refrigerant system
Vapor compression system

1. Introduction

In recent years, refrigeration technologies have been widely adopted across sectors, from domestic appliances to large-scale industrial systems. The coefficient of performance (COP) is a key indicator of a cooling system's energy efficiency. It represents the ratio of the thermal energy extracted in the evaporator (Qo in kW) to the input power required (W in kW). A higher coefficient of performance indicates better energy utilization, which is vital for reducing consumption. Given that each refrigerant has distinctive thermophysical characteristics, estimating COP during system development becomes a fundamental step. Accurate COP estimation is essential, as refrigerant behavior differs under operational conditions. Over the past few years, artificial intelligence, particularly machine learning, has gained traction as an influential approach in the analysis of heating, ventilation, and air conditioning (HVAC) and cooling systems. These techniques enable performance forecasting from input data without relying on traditional thermodynamic modeling. Several contemporary studies have investigated the application of data-driven algorithms to enhance COP prediction, achieving greater speed and precision. This study conducts a literature review to address three related questions: (i) how researchers have applied machine learning (ML) techniques to predict the coefficient of performance and enhance the energy performance of vapor compression and related refrigeration or heat pump systems; (ii) what operating ranges of superheat and subcooling previous thermodynamic studies have recommended; and (iii) what main advantages and limitations characterize commonly used machine learning methods. The search was carried out in Scopus, Web of Science, and Google Scholar using combinations of keywords such as "coefficient of performance", "COP", "vapor compression", "refrigeration system", "heat pump", "HVAC", "machine learning", "neural network", "subcooling degree", and "superheat". This review included peer-reviewed journal articles and full conference papers published in English between 2011 and 2026. The study prioritized papers that clearly reported input variables, target outputs, and performance metrics, including the coefficient of determination (R^2), root mean square error, and mean absolute error. It also emphasized studies that critically analyzed the advantages and limitations of regression models, tree-based methods, support vector machines, and neural networks.

Mshragi et al. [1] proposed a fast machine learning framework for real-time optimization of heating, ventilation, and air conditioning systems in building management systems. Their method combines a Random Forest surrogate model implemented on a field programmable gate array with an adaptive genetic algorithm running on the processing unit. The system achieved more than 1.67 million predictions per second and reduced electricity use by more than 50 percent while maintaining occupant comfort within standard thermal comfort limits. Researchers Zhou and Zhu [2] proposed an intelligent virtual twin framework combining a convolutional neural network (CNN), a multilayer perceptron, and a long short-term memory (LSTM), with real-time optimization via the particle swarm optimization (PSO) algorithm. The study selected key influencing factors according to their correlation with the coefficient of performance. CNNs and MLPs capture nonlinear features, while LSTMs handle time series. The proposed model achieved better prediction performance: MAE, MAPE, and RMSE were 7.01%, 1.71%, and 0.08% lower than CNN; 0.92%, 0.24%, and 0.01% lower than LSTM;

*CORRESPONDING AUTHOR | Vo Van An | ✉ anvv@tdmu.edu.vn

and 2.41%, 0.58%, and 0.02% lower than MLP. The results demonstrate that the proposed model achieves higher COP prediction accuracy compared to the individual baseline methods. Pektezel et al. [3] compared the performance of R290 and R600a refrigerants in a vapor compression system and applied machine learning to predict system parameters. Although R600a reduced compressor power by 33.44%, R290 improved COP by 23.77%, cooling capacity by 82.55%, and second law efficiency by 20.99%. This study employs multilayer perceptron (MLP), support vector machine (SVM), and decision tree (DT) algorithms to predict system performance. Among them, SVM achieved the highest accuracy and the lowest MAE in COP prediction (0.0317), compared to MLP (0.0324) and DT (0.0989). Results confirm R290 performs better thermodynamically, and SVM is the most effective prediction model. Wang et al. [4] evaluated four ML models: support vector machine, artificial neural network (ANN), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM) for predicting HVAC system behavior in a building energy management context. The results show that XGBoost achieved the highest prediction accuracy, with $R^2 = 0.978$ for power and 0.983 for temperature. However, LightGBM demonstrated the fastest training and prediction times, making it more suitable for real-time applications.

Deng et al. [5] proposed a novel framework for predicting the chiller Coefficient of Performance by employing a dynamic graph convolutional network enhanced through association rule mining. This approach uses a historical association graph bank, fused with real-time sensor correlations, to effectively capture the complex, evolving interactions within HVAC systems. Empirical validation using data from an actual building demonstrated that this dynamic GCN architecture achieves superior predictive accuracy compared to current state-of-the-art benchmarks. Nguyen et al. [6] investigated machine learning for fault detection in refrigeration systems using difluoromethane (R32). The authors developed a synthetic dataset with 1,998 samples in Python. They tested four models: Naïve Bayes, generalized linear model, decision tree, and random forest to detect refrigerant leakage and air filter clogging. All models achieved high accuracy above 96 percent, while the decision tree and random forest showed strong generalization with accuracies of 97.9 percent and 97.4 percent, respectively. The study shows that machine learning can support early fault detection, reduce energy losses, and improve system service life. Cho et al. [7] explored the integration of machine learning to estimate COP in ground-source heat pump applications. Their study considered three algorithms: ANN, Tree-based ensemble learning, and SVM, optimized via Bayesian strategies for enhanced accuracy. Using TRNSYS simulation data, they modeled a residential building based on temperature and fluid flow parameters. The ANN method delivered the strongest predictive performance, achieving the highest R^2 and lowest RMSE, followed by the Tree Ensemble model, and lastly SVM. Notably, the SVM's R^2 failed to meet ASHRAE guidelines, whereas the ANN model exhibited the narrowest maximum deviation (±0.025), confirming its greater precision in GSHP system control.

Senthilkumar [8] analyzed experimental data from an R600a vapor compression refrigeration system enhanced with CuO/SiO_2 nano lubricant (0-0.6 g/L) and varying refrigerant charge (50-70 g). Three ML models were developed, such as LR, ANN, and SVM, using a 70/15/15 train/validation/test split. Based on RMSE, MAE, MSE, and R^2, the quadratic SVM achieved the strongest predictive performance ($R^2 = 0.970196$, MAE = 0.055486) relative to the ANN and linear regression models. Yu et al. [9] optimized energy management in commercial building air conditioning systems using AI-based regression, identifying ridge regression as the most suitable method for modeling chiller system coefficient of performance and key influencing variables. The researchers evaluated regularised regression models: ridge, lasso, and elastic net on time series data from five chillers with two capacities. Ridge regression achieved the best predictive accuracy when fine-tuned with optimal hyperparameters. Important features affecting SCOP include part load ratio, number of active chillers and pumps, and water temperatures. The model revealed July as the peak month for energy improvement. This work offers a practical SCOP prediction model that supports more sustainable building operation strategies. Ghanbarpour et al. [10] used an ANN to evaluate the potential of R449A as a lower-GWP alternative to R404A in supermarket refrigeration systems. This study trained the ANN using real-world field data collected from systems operating at both low and medium temperature levels. Results showed that R449A improved the coefficient of performance by approximately 10% (low temperature) and 5% (medium temperature) compared to R404A, with similar cooling capacities. The ANN model accurately predicted system energy performance, demonstrating its reliability for retrofit analysis under variable real-world conditions. Additionally, the study confirmed a reduction in carbon emissions when using R449A. Overall, previous studies have applied a wide range of machine-learning techniques to COP prediction and related performance analysis. K-means clustering is mainly used as an unsupervised pre-processing tool to group operating conditions, rather than to predict COP directly. Kernel-based methods, such as support vector machines and their least squares variants, can model nonlinear relationships with relatively small datasets; however, their performance is sensitive to kernel selection and hyperparameter tuning, and the resulting models are less transparent to engineers. Artificial neural networks and deep learning approaches achieve high accuracy and enable advanced control strategies. Still, they usually require large, high-quality datasets, careful regularization to avoid overfitting, and behave as black box models from a physical point of view.

The present work develops and compares simple machine learning regression models for fast COP prediction in a single-stage vapor compression system used for food storage. The study models the coefficient of performance as a function of five key input variables: refrigerant type, evaporation temperature, condensing temperature (t_c), subcooling, and superheat. To construct the thermodynamic database, 2000 operating conditions were generated using the Engineering Equation Solver (EES). The dataset comprises 500 operating points for each of four selected refrigerants: R1234yf, R134a, R290, and R600a. This study selects these refrigerants because they comply with current environmental regulations and support the ongoing transition toward low global warming potential (low GWP) alternatives. Among them, R134a exhibits a relatively high global warming potential (GWP ≈ 1430). International environmental agreements and regional F-gas regulations are gradually phasing out the use of R134a. Consequently, in this study, R134a is primarily employed as a benchmark refrigerant to facilitate comparative performance evaluation with emerging low-GWP alternatives. R1234yf, by contrast, has a near-zero GWP and zero ODP and is promoted as a lower-impact replacement for R134a in many applications. R290 and R600a are natural hydrocarbons that exhibit very low global warming potential (GWP ≈ 3) and zero ozone depletion potential (ODP), making them environmentally attractive refrigerants. However, their high flammability class (A3) requires strict charge limits and appropriate safety measures during system design and operation. Consequently, engineers commonly use these refrigerants in small hermetic vapor-compression systems, particularly in domestic and light commercial refrigeration applications. Using this dataset, this study applies five supervised learning methods: linear regression, polynomial regression, random forests, decision trees, and support vector machines. It compares their COP prediction performance to identify simple models that provide engineers with a fast, practical alternative to manual thermodynamic calculations during preliminary design and parametric analysis.

2. Materials and Methods

This research aims to accurately evaluate the coefficient of performance of refrigeration systems using four common refrigerants: R1234yf, R134a, R290, and R600a. Food preservation systems and domestic refrigerators widely use these refrigerants. To accomplish this objective, the study applies several machine learning models to predict COP. Figures 1 and 2 illustrate the overall research workflow. To develop a mathematical model for estimating COP, it is essential to understand the basic principles of the refrigeration system. In addition, researchers must carefully consider the system state points, as they directly affect the accuracy of the COP calculation. Figure 2 illustrates the operating principle of the refrigeration system. The vaporized refrigerant leaves the evaporator and enters the compressor. There must be no liquid in the refrigerant, so it must be completely dry. To evaluate that, the superheat value, which is the difference between the temperature at the evaporator's outlet and the saturation temperature corresponding to the evaporator's pressure (also called suction pressure), is used. The compressor then compresses this superheated vapor (State 1) into a gas at high pressure and high temperature (State 2), which then comes to the condenser. Inside the condenser, the hot, superheated refrigerant gas is cooled and turned into a liquid while the pressure remains constant. To make the refrigeration system work more efficiently (higher COP) and prevent gas bubbles from forming in the liquid line, the liquid refrigerant needs to be cooled further, a process called subcooling. Subcooling represents the difference between the condensing temperature (determined from the condensing pressure) and the refrigerant temperature at the condenser outlet. Next, the cooled liquid refrigerant passes through a filter that removes tiny dust particles, keeping the system clean (State 3). After that, it goes into the expansion valve. At this point, the high-pressure condenser discharge drops to the evaporator's lower pressure, and the temperature decreases to match the evaporator's conditions (State 4). Inside the evaporator, the cold refrigerant absorbs heat from the surrounding area. As it absorbs this heat, it changes from a liquid to a saturated vapor, then becomes superheated vapor (State 1). The compressor then pulls the vapor back into the compressor chamber. The entire evaporation process occurs while the pressure in the evaporator remains constant. This cycle keeps repeating itself in the refrigeration system.

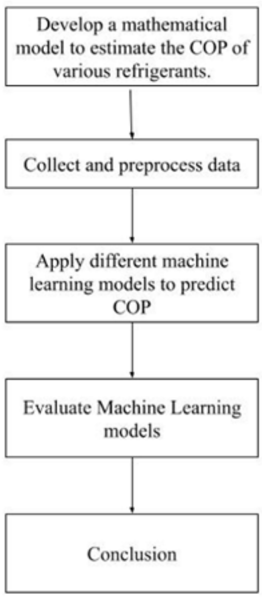


Figure 1. Framework for the study of COP prediction for different refrigerants

According to Babu et al. [11], maintaining a subcooling temperature between 3°C and 5°C is recommended for efficient system operation. This temperature range helps ensure the refrigerant remains fully liquid before entering the expansion valve, which is important for stable, consistent performance. Ensuring the refrigerant is entirely liquid at this stage also improves the system's overall efficiency, as measured by the Coefficient of Performance. Altuntas et al. [12] conducted experiments on a refrigeration system using different refrigerants, including R-22, R-404A, R-507, and the alternative refrigerant R-407C. The authors showed that applying 5 K of subcooling in the condenser can reduce energy use by 4% to 8% and improve exergy efficiency by 3% to 25%. Regarding the superheat value, Sunu et al. [13] suggest a range of 4.6 to 7.6 K. This helps ensure the compressor does not take in wet vapor, which could damage the system. In this study, we examine a refrigeration system designed for food storage using four refrigerants: R1234yf, R134a, R290, and R600a. We generate the refrigerant data using EES software under the following operating conditions:

i) Evaporation temperature: from -29°C to -5°C (suitable for food storage and home refrigerators)
ii) Condensing temperature: from 41°C to 52°C (commonly used in household cooling systems)
iii) Superheat range: 4 to 15 K
iv) Subcooling range: 1 to 7 K

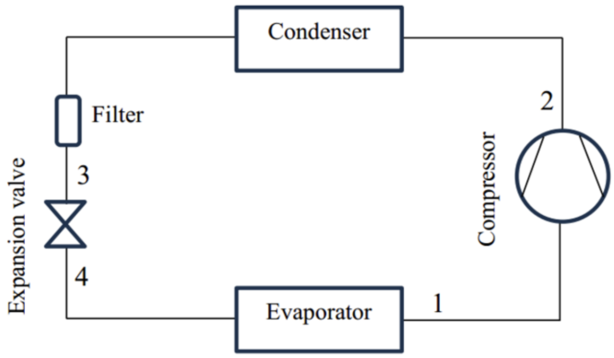


Figure 2. Refrigerant principle diagram

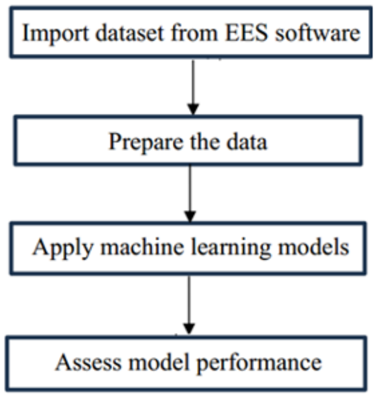


Figure 3. Detailed machine learning pipeline for the EES-generated dataset used in COP prediction

The COP is used to evaluate the efficiency of a refrigerant system. A higher COP value indicates more efficient system operation and lower energy consumption for the same cooling capacity. Equation (1) calculates the COP value.

$$COP = (h_1 - h_4)/(h_2 - h_1) \tag{1}$$

where h represents the refrigerant's enthalpy at each specific state.

This study used EES software to obtain the refrigerant enthalpy values at each thermodynamic state and subsequently calculated the Coefficient of Performance. The analysis considered four refrigerants, including R1234yf, R134a, R290, and R600a. The calculation process employed five input parameters: evaporation temperature, condensing temperature, refrigerant type, superheat, and subcooling. These inputs were randomly varied to produce different COP values. A total of 500 COP values were generated for each refrigerant, yielding a dataset of 2000 data points. The dataset was analyzed using the Google Colab platform, and the overall analysis procedure is presented in Figure 3. Figure 3 shows the machine learning workflow used in this study. First, 2,000 raw data points generated by the EES software were imported into a Python program using the Google Colab platform. The next important step was data preparation, which is essential because data quality directly affects ML model performance. The study carefully checked the dataset to ensure that no values were missing or inconsistent. If any data were missing, they were either imputed using statistical methods, such as mean or median imputation, or removed when deemed unreliable. The data were then processed to ensure consistent units and a structure suitable for analysis. The refrigerants such as R1234yf, R134a, R290, and R600a were encoded as 0, 1, 2, and 3, respectively. After data preparation, the study split the dataset into two subsets: a training set and a test set. This study used 80% of the data for training and the remaining 20% for testing. This approach, known as the 80/20 split, helps evaluate the model's performance on new, unseen data. Training on a larger portion allows the model to learn better, and testing on a smaller portion checks its ability to generalize. This step is important to ensure the model is not only accurate but also works well with similar data in the future. Finally, five ML models, such as linear regression, polynomial regression, random forest, decision tree, and support vector machine, were applied to predict COP. Polynomial Regression with different degrees was tested and compared with other models to identify the most effective one. These five models are commonly used in HVAC and refrigerant systems because they can handle complex data relationships.

Linear and polynomial regression are fundamental algorithms for exploring the relationship between dependent and independent variables. While LR is a basic statistical method, it has limitations for modeling complex, nonlinear relationships compared to more advanced ML models [14]. In contrast, PR extends LR by including polynomial terms, thereby capturing more complex trends in the data. The proposed model is better suited to cases where the relationship between variables is not strictly linear. As a result, PR has been applied in studies of refrigerant systems to model their performance more accurately under various operating conditions [15]. Decision trees and random forests are machine learning algorithms known for their robustness and interpretability. Since they perform well in nonlinear and high-dimensional environments, they are particularly useful for predicting refrigerant system performance and estimating energy consumption, especially when working with noisy or incomplete datasets. The DT algorithm builds a tree-structured model in which each branch represents a decision rule based on input features, and each leaf node represents an outcome. This clear structure makes it easy to understand and interpret. However, single decision trees may overfit, reducing prediction accuracy. To overcome this limitation, RF combines predictions from multiple decision trees, thereby improving overall accuracy and reducing the risk of overfitting. This ensemble method enhances stability and generalization, making it more reliable for practical applications. Both algorithms are widely used in refrigerant system studies for tasks such as fault detection and regression modeling because they deliver strong predictive performance and effectively handle complex data relationships [16-17].

Researchers apply SVM, a supervised learning algorithm, to both classification and regression tasks. It operates by identifying an optimal hyperplane that maximizes the margin between data points of different classes. In some studies on HVAC and refrigerant systems, this model has been combined with Random Forest to improve fault-detection and diagnosis accuracy, even with limited data [18]. In this study, we implemented the RF, DT, and SVM models using the scikit-learn library as RandomForestRegressor(random_state = 42), DecisionTreeRegressor(random_state = 42), and SVR(kernel="rbf"), respectively, while keeping all other hyperparameters at their default values to establish a standard baseline. To evaluate model performance, the coefficient of determination, Root mean square error, and mean absolute error were used, with their formulas given in Eqs. (1-3). The coefficient of determination measures how well the model's predictions approximate the actual data values. It indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y_i})^2} \tag{2}$$

where y_i is the actual observed value, $\hat{y}$ is the predicted value from the model, $\bar{y}$ is the mean of the actual values, and n is the number of observations

RMSE is the square root of the average squared difference between predicted and actual values. The RMSE metric evaluates model performance; lower values indicate higher prediction accuracy and are expressed in the same units as the predicted variable.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \tag{3}$$

The MAE metric quantifies the mean absolute difference between predicted and actual outcomes. A smaller MAE value indicates higher prediction accuracy and is expressed in the same units as the target variable.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| \tag{4}$$

3. Results and Discussion

3.1 Data visualization

The study uses a pair plot to visualize the overall data distribution and the relationships among numerical variables, as shown in Figure 4. Figure 4 reveals potential relationships, patterns, and outliers among key variables, including evaporation temperature, condensing temperature, subcooling, superheat, and COP. In addition, the study calculates the skewness of the input variables to assess the shape of their distributions. The skewness values for te, tc, subcooling, superheat, and COP are -0.036, -0.113, -0.05, 0.0269, and 0.37, respectively. Since all values fall within the range of -0.5 to 0.5, the data distribution is approximately symmetrical. Therefore, the dataset is well-suited for ML models, which often perform better with balanced and normally distributed input features. The lower half of the pair plot shows how each numerical variable relates to the others. Moreover, when combined with a heatmap (Figure 5), these relationships become even clearer. Figure 5 shows a correlation heatmap of Pearson correlation coefficients between the variables. The analysis helps clarify the relationships among the variables. The strongest positive correlation (0.89) is observed between COP and evaporation temperature, as shown by the upward trend in Figure 4. The analysis indicates that any change in the evaporation temperature (t_e) has a strong impact on the coefficient of performance. Specifically, as evaporation temperature increases, the COP rises; as it decreases, the COP drops. On the other hand, condensing temperature has a negative correlation with COP (-0.40), aßs shown by the downward trend in the pair plot. Higher condensing temperatures lead to lower system efficiency and increased energy consumption. In contrast, subcooling and superheat exhibit very weak correlations with COP (0.13 and 0.03, respectively), indicating that these parameters have limited effects on system efficiency under the examined conditions. In this dataset, subcooling and superheat show only a weak correlation with

COP compared with the evaporation and condensing temperatures. The limited impact of these variables arises from the fact that both were varied only within narrow "useful" ranges that ensure safe system operation. Previous studies indicate that about 5 K of subcooling can reduce energy use by roughly 4-8% [12], which is a modest COP improvement but still smaller than the effect of changing the main temperature levels. As a result, within the operating range considered here, subcooling and superheat have a secondary influence on COP, while evaporation and condensing temperatures remain the dominant factors.

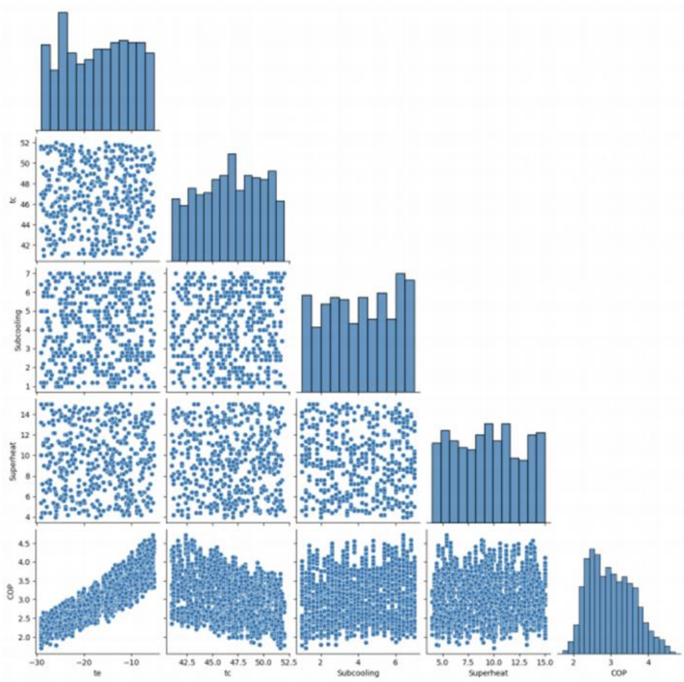


Figure 4. Scatter plot matrix of input features and target variable

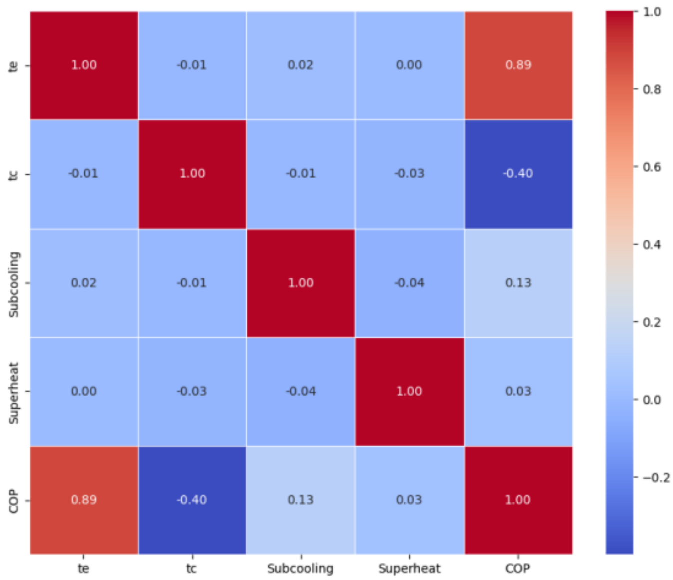


Figure 5. Correlation matrix of operating parameters and COP

3.2 Model Performance Evaluation

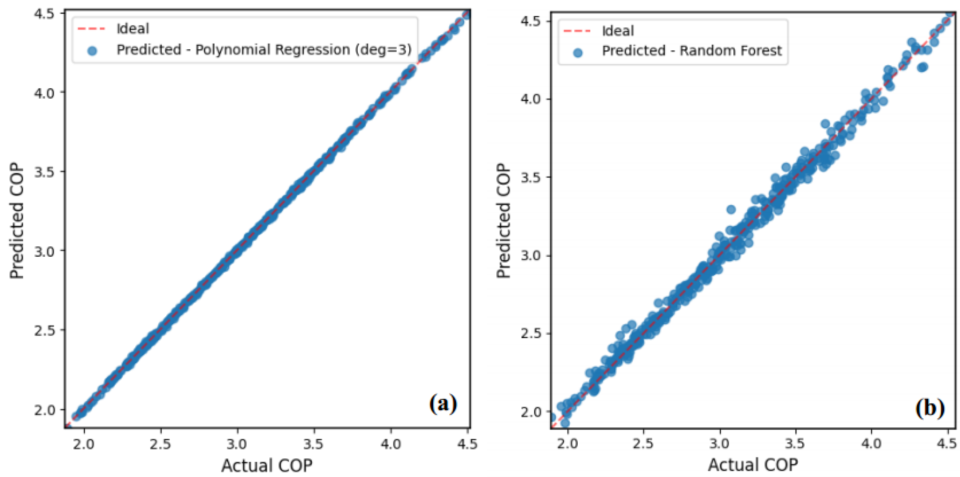
The performance of five machine learning models, including linear regression, polynomial regression, random forest, decision tree, and support vector machine, was evaluated based on three metrics: R^2, RMSE, and MAE. The study employs these metrics to assess the models' accuracy in predicting COP values. Each model was trained and tested on the same dataset to ensure fair comparison. The results provide insight into which model offers the best balance between prediction accuracy and error minimization, as shown in Table 1. Table 1 compares the predictive accuracy of five ML algorithms using three standard performance indicators: R^2, RMSE, and MAE. Polynomial regression was applied with degrees ranging from 1 to 3, where the first-degree case is mathematically equivalent to linear regression. Among the evaluated models, PR of degree 3 exhibited the most favorable performance, achieving an R^2 of 0.999857 and the lowest error metrics (RMSE = 0.007084, MAE = 0.005254). These values suggest a near-perfect fit to the dataset, making it the most accurate model in this context. Following closely are RF and PR with degree 2. Although both models yielded high predictive accuracy, RF slightly outperformed the second-degree PR across all three metrics, particularly in MAE, where it achieved 0.036628 compared to 0.043716. SVM and DT models demonstrated moderate accuracy, with comparable R^2 scores around 0.988 and marginally higher error values. Researchers may select these models depending on the application, particularly when model interpretability or training speed is prioritized over maximum accuracy. In contrast, LR, corresponding to the lowest polynomial degree, showed the weakest performance, reflecting its limited capacity to model the complex, nonlinear relationships in the data. This comparative analysis highlights the advantages of using higher-order polynomial regression and ensemble-based methods, such as Random Forests, for nonlinear patterns. However, researchers should consider the potential for overfitting when using very high polynomial degrees, even though degree 3 performs best in this case.

Table 1. Assessment measures of learning algorithms

Model	Degree	R^2	RMSE	MAE
LR	1	0.976597	0.090657	0.072418
PR	2	0.992592	0.051007	0.043716
PR	3	0.999857	0.007084	0.005254
RF	N/A	0.993429	0.048037	0.036628
DT	N/A	0.987123	0.067246	0.051437
SVM	N/A	0.988751	0.062853	0.053906

3.3 Model Accuracy Visualization

In this study, the dataset was split into two parts using an 80/20 ratio. Specifically, the study uses 80% of the dataset to train the machine learning models and reserves the remaining 20% for testing. The testing phase did not reuse the training data to ensure an unbiased evaluation. Based on the obtained performance results, the study selected the two best-performing models to predict the coefficient of performance. A scatter plot comparing the actual and predicted COP values for these models is shown in Figure 6 to assess their predictive performance visually.



(a) Polynomial regression (deg = 3) (b) Random forest

Figure 6. Scatter plot for polynomial regression (deg = 3) and for random forest

Figure 6 shows the results of predicting COP using two models: polynomial regression (degree 3) and a random forest. The horizontal axis shows the actual COP values, and the vertical axis shows the model-predicted values. The red dashed line is the ideal line where the prediction matches the actual result. In the first graph, the polynomial regression predictions are very close to the ideal line. The points are almost perfectly aligned, meaning the model predicted the values very accurately. There is very little error, and the model captures the relationship between the variables well. The results indicate that a third-degree polynomial regression model is highly effective for modeling this dataset. Figure 6(b) shows that the RF model performs well; however, the points are more spread out than in the first graph. Some values are slightly higher or lower than the ideal line, especially at the edges. This result indicates that the RF model produces slightly higher prediction errors and lower precision compared with the PR model. However, caution is needed to avoid overfitting. If the polynomial degree is too high or the model is applied to different datasets, the model may fit the training data too closely and consequently perform poorly on the test data. At the same time, random forests typically handle complexity better, reducing overfitting through ensemble learning and thereby enhancing generalization to unseen data [19-20]. Furthermore, the actual and predicted COP values are compared and illustrated in Figure 7. Figure 7 shows the comparison between actual and predicted COP values for two models: (a) polynomial regression and (b) random forest. Each vertical line connects the true value and the predicted one for each test sample. In Figure 7(a), the lines are generally short and consistent, showing that the Polynomial Regression model gives accurate and stable predictions. In contrast, Figure 7(b) shows greater variation in line lengths, indicating that the Random Forest model has larger, less consistent errors. Overall, Figure 7 suggests that Polynomial Regression (degree = 3) performs better in this case, while Random Forest shows slightly less accuracy but still captures the general trend. Even though Polynomial Regression yields strong results here, it is still important to monitor for overfitting, especially when using higher-degree models.

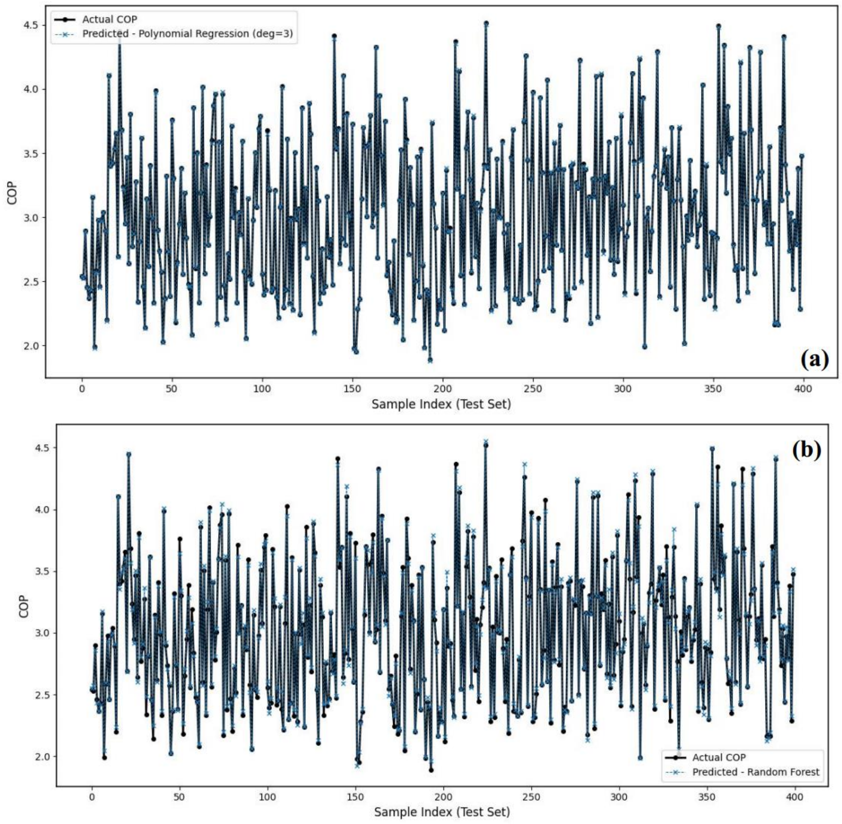


Figure 7. Comparison between actual and predicted COP values on the test set using (a) Polynomial Regression (degree = 3) and (b) Random Forest

3.4 Analysis of Model Parameters and Residuals

This study analyzes feature importance and coefficient values for the two models (random forest and polynomial regression) to understand how each input variable contributes to COP prediction. Figure 8 shows the direction and strength of the relationship between each variable and COP. In the random forest model, importance scores measure how much each feature contributes to decision-making across the trees. A higher importance score means that the feature has more influence on the model's output. In contrast, for polynomial regression, the model provides coefficient values that represent the strength and direction of the relationship between each input variable and the target variable. By examining these values, we can see how strongly each factor affects COP and whether its impact is positive or negative. Regarding condensing temperature, lower condensing temperatures increase COP. Among the input variables, evaporation temperature has a strong impact on COP. As the evaporation temperature increases, the COP also rises; when the temperature drops, the COP decreases. Both models confirm the importance of this factor, reflected in its high importance score (Random forest) and coefficient value (Polynomial regression). According to Al-Rashed et al. [21], the increase in COP is due to a higher evaporator temperature, which lowers the compressor pressure ratio. Reducing compressor workload improves the system's overall efficiency. Similarly, condensing temperature plays a key role in the stem's performance. A lower condensing temperature typically results in a higher COP, making the system more energy efficient. Ismael et al. [22] found that using geothermal cooling improved COP by 41%, while evaporative cooling achieved up to 65% enhancement by reducing condenser temperature. In the case of Polynomial Regression, subcooling shows a positive correlation with COP, having a coefficient of around 0.12. The results suggest that increased subcooling slightly improves system performance. However, in the random forest model, subcooling is assigned a very low importance score, implying it has little effect on prediction. This difference shows how different algorithms evaluate the same feature differently, depending on their internal structures and variable interactions. Subcooling not only reduces the work input required by the compressor but also boosts the overall COP by exceeding 10% in practical refrigeration cycles [23].

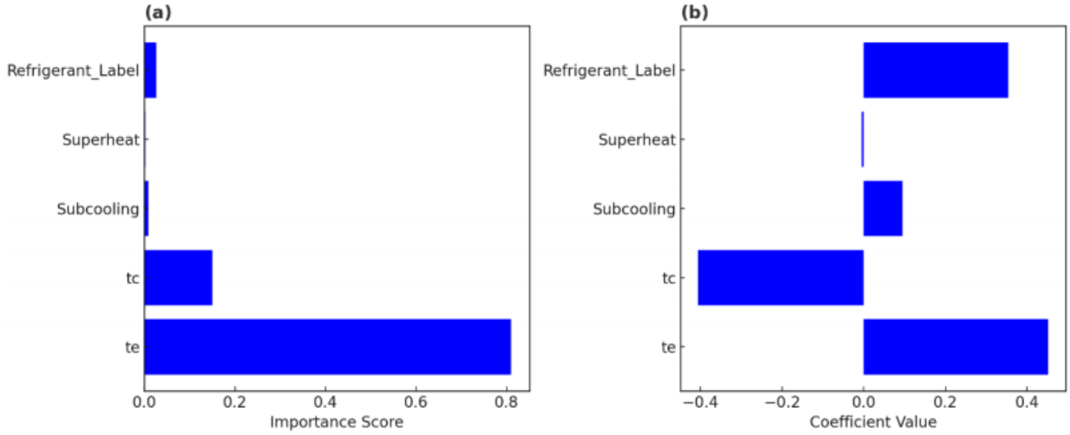


Figure 8. Contribution of input features to COP prediction for two models: (a) Feature importance scores for the Random Forest model, (b) Coefficient values from the Polynomial Regression model (degree = 3).

The impact of the Refrigerant Label on COP varies between the two models. In the random forest model, this feature has a very low importance score, meaning it contributes little to the prediction. In contrast, the polynomial regression model yields a high coefficient for refrigerant label, indicating a strong effect on COP. The observed variation in system performance is reasonable, as each refrigerant has unique thermodynamic properties that influence the system's operating characteristics. Overall, this difference shows how each algorithm interprets the same feature differently. Polynomial regression captures the overall influence of refrigerant types, whereas random forests may focus more on numerical features. Eq. (5) presents the polynomial regression formula used to calculate the COP for each refrigerant.

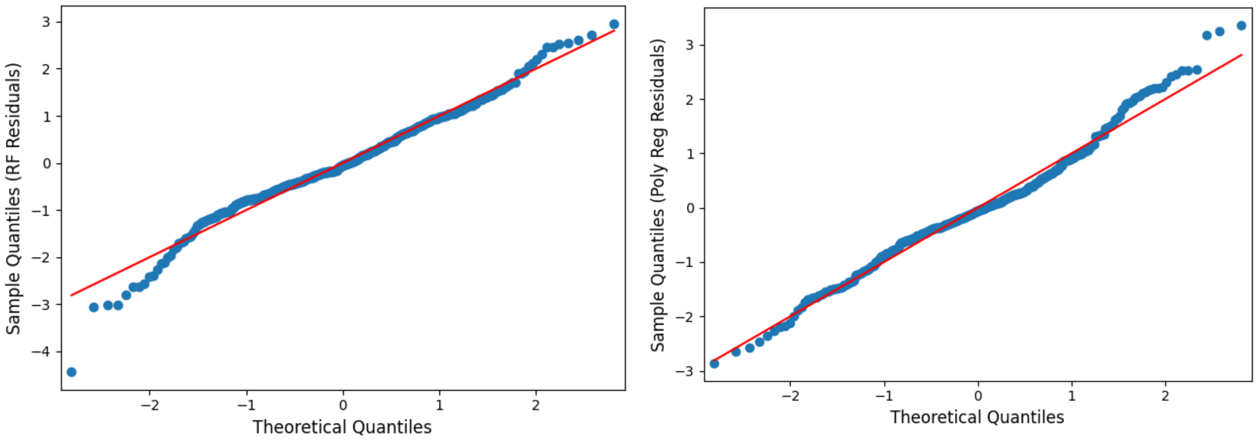
$$
\begin{aligned}
& COP = 0.4524 * t_e - 0.4057 * t_c + 0.9475\times10^{-1} * Subcooling - 0.4049\times10^{-2} * Superheat + 0.3543 * \\
& Refrigerant\_Label + 0.4975\times10^{-2} * t_e^2 - 0.9456\times10^{-2} * t_e * t_c + 0.1814\times10^{-2} * t_e * Subcooling \\
& + 0.9041\times10^{-4} * t_e * Superheat - 0.2068\times10^{-2} * t_e * Refrigerant\_Label + 0.3878\times10^{-2} * t_c^2 \\
& - 0.8381\times10^{-3} * t_c * Subcooling + 0.4899\times10^{-3} * t_c * Superheat + 0.3252\times10^{-2} * t_c * \\
& Refrigerant\_Label - 0.4628\times10^{-3} * Subcooling^2 - 0.3822\times10^{-3} * Subcooling * Superheat \\
& - 0.8426\times10^{-2} * Subcooling * Refrigerant\_Label + 0.2655\times10^{-3} * Superheat^2 - 0.6562\times10^{-2} * \\
& Superheat * Refrigerant\_Label - 0.3045 * Refrigerant\_Label^2 + 0.1988\times10^{-4} * t_e^3 - 0.5871\times10^{-4} * \\
& t_e^2 * t_c + 0.1040\times10^{-4} * t_e^2 * Subcooling + 0.4970\times10^{-6} * t_e^2 * Superheat + 0.7171\times10^{-5} * t_e^2 * \\
& Refrigerant\_Label + 0.5407\times10^{-4} * t_e * t_c^2 - 0.1239\times10^{-4} * t_e * t_c * Subcooling + 0.4694\times10^{-6} * \\
& t_e * t_c * Superheat + 0.8950\times10^{-5} * t_e * t_c * Refrigerant\_Label - 0.5172\times10^{-5} * t_e * \\
& Subcooling^2 - 0.4904\times10^{-5} * t_e * Subcooling * Superheat - 0.5236\times10^{-4} * t_e * \\
& Subcooling * Refrigerant\_Label + 0.6019\times10^{-6} * t_e * Superheat^2 - 0.4909\times10^{-5} * \\
& t_e * Superheat * Refrigerant\_Label + 0.9320\times10^{-3} * t_e * Refrigerant\_Label^2 - 0.1343\times10^{-4} * \\
& t_c^3 + 0.2882\times10^{-5} * t_c^2 * Subcooling - 0.3512\times10^{-5} * t_c^2 * Superheat + 0.4751\times10^{-5} * t_c^2 * \\
& Refrigerant\_Label + 0.2932\times10^{-5} * t_c^2 * Subcooling^2 + 0.8739\times10^{-6} * t_c * Subcooling * \\
& Superheat + 0.1537\times10^{-5} * t_c * Subcooling * Refrigerant\_Label - 0.4534\times10^{-5} * t_c * \\
& Superheat^2 - 0.3781\times10^{-5} * t_c * Superheat * Refrigerant\_Label - 0.8616\times10^{-3} * t_c * \\
& Refrigerant\_Label^2 + 0.1416\times10^{-4} * Subcooling^3 - 0.2870\times10^{-5} * Subcooling^2 * \\
& Superheat + 0.1212\times10^{-4} * Subcooling^2 * Refrigerant\_Label + 0.3065\times10^{-5} * Subcooling * \\
& Superheat^2 + 0.2400\times10^{-4} * Subcooling * Superheat * Refrigerant\_Label + 0.1543\times10^{-2} * \\
& Subcooling * Refrigerant\_Label^2 - 0.1697\times10^{-5} * Superheat^3 + 0.1987\times10^{-5} * Superheat^2 * \\
& Refrigerant\_Label + 0.1921\times10^{-2} * Superheat * Refrigerant\_Label^2 + 0.7168\times10^{-1} * \\
& Refrigerant\_Label^3 + 16.0234
\end{aligned}
\tag{5}
$$

where t_e is evaporation temperature (^oC), and t_c is condensing temperature (^oC)

The refrigerant label indicates the type of refrigerant used in the model. The values are encoded numerically as follows: R1234yf = 0, R134a = 1, R290 = 2, and R600a = 3. Overall, the degree 3 polynomial regression model combines high predictive accuracy with a simple closed-form equation, making it easy to implement in spreadsheets or control algorithms. Random forests also offer strong performance and robustness to noise, but they lack the transparency and explicit functional form of the polynomial model. These features make the polynomial regression model a particularly attractive option for practical COP estimation in preliminary design.

3.5 Residual Analysis Plots

To further assess the quality of the fitted models, residual analyses were conducted for the Random forest and degree-3 polynomial regression. For this purpose, Q-Q plots (Figure 8) are used to compare the sample quantiles of the residuals with the theoretical quantiles of a normal distribution. If the points lie close to the reference line, the residuals are approximately normally distributed, and the models can be considered adequate within the studied operating range. Figure 9 presents Q-Q plots for the residuals of the Random Forest and degree-3 polynomial regression models. In both plots, the points closely follow the red reference line in the central region. The residual distribution closely approximates a normal distribution and exhibits minimal skewness. At the two ends of the plots, some points move away from the line. For the Random Forest model, a few residuals in the lower tail are more negative than expected. In contrast, for the polynomial model, several residuals in the upper tail are larger than the normal reference. These deviations suggest slightly heavier tails, but they affect only a small portion of the data. Overall, the Q-Q plots show that the residuals of both models are approximately normal, with minor departures in the extremes, and there is no clear sign of severe misspecification of the COP prediction models. Between the two best-performing models, the degree-3 polynomial regression achieves much lower RMSE and MAE than the Random Forest model, and its residuals are approximately normal. The results indicate that the polynomial model provides the highest accuracy on the present dataset.



(a) (b)

Figure 9. Q-Q plots of model residuals: (a) Random Forest; (b) degree-3 polynomial regression

3.6 Comparison with Other Studies

Table 2 compares several studies that applied ML to predict the performance of refrigerant systems. It shows the types of models used and examines their R^2, RMSE, and MAE, underlining the importance of choosing the right model for this research. Overall, the studies listed in Table 2 show that most machine learning applications for COP prediction in vapor compression systems focus on a single refrigerant, a specific system configuration, or a particular advanced algorithm. As a result, the reported models are often difficult to compare directly. They are not always easy to implement in everyday engineering practice, especially when they rely on complex black box techniques or small experimental datasets. In contrast, the present study uses a unified dataset for four low-GWP refrigerants operating in the same vapor-compression refrigeration system. It benchmarks several simple, widely available regression models. This approach provides a more consistent basis for comparing refrigerant options and offers practical tools for engineers to estimate COP quickly and perform preliminary design. Compared with advanced models reported in Table 2, such as ANN, MTL-XGBoost, or gradient-boosted trees, the proposed degree-3 polynomial regression and Random Forest models offer a favorable balance between accuracy, simplicity, and transparency. In particular, the closed-form polynomial model can be implemented directly in spreadsheets or controller code without retraining, which is useful for engineering practice.

Table 2. Comparative analysis of ML models for the refrigerant system

Reference	Used Method	Research content	Result
Zhou and Zhu [2]	Developed a digital twin model combining CNN, MLP, and LSTM for comprehensive performance prediction.	- Real-time optimization achieved using the PSO algorithm. - Key influencing factors were selected based on correlation analysis with COP.	MAE, MAPE, and RMSE were 7.01%, 1.71%, and 0.08% lower than CNN; 0.92%, 0.24%, and 0.01% lower than LSTM; and 2.41%, 0.58%, and 0.02% lower than MLP.

		CNNs and MLPs extract nonlinear spatial features.	
Pektezel et al. [3]	MLP, SVM, and DT.	To predict COP in R290 and R600a refrigeration systems	SVM achieved the highest accuracy and the lowest MAE in COP prediction (0.0317), compared to MLP (0.0324) and DT (0.0989).
Wang et al. [4]	SVM, ANN, XGBoost, and LightGBM	Predict HVAC system behavior in a building energy management context	-XGBoost achieved the highest prediction accuracy with R^2 = 0.978 (power) and R^2 = 0.983 (temperature). -LightGBM is the fastest training and prediction time.
Nguyen et al. [6]	Naïve Bayes, Generalized linear model, DT, and RF	To detect refrigerant leakage and air filter clogging.	-All models reached an accuracy above 96%. -Decision tree and random forest performed well on generalization (97.9% and 97.4%).
Ma et al. [24]	MLP, SVM, and RF	Predict the COP of solar-assisted heat pumps	Random forest achieved the best prediction, with MAE = 2.35% and RMSE = 3.84% on the test set. MAE of 2.42%, RMSE of 4.01% on the train set. Training time was 6.57 seconds.
Ye et al. [25]	MTL-XGBoost (Optuna-tuned)	Performance prediction of R290 scroll compressors under multiple operating conditions	R^2>0.98, MSE/MAE were substantially lower than baselines.
Akyol et al. [26]	CPSO-TSA	To predict the energy consumption of the R600a refrigerant system	Accuracy of 84.4%, MAE of 0.1556, and RMSE of 0.3949, Kappa statistic of 0.6490.
Ashour et al. [27]	ANN, RF, XGB, AB, and KNN	A hybrid refrigeration system was developed by integrating thermoelectric cooling modules with a vapor compression system to achieve lower freezer temperatures.	ANN is the most accurate model (R^2 > 0.98) and has the lowest Mean Absolute Error (MAE) in forecasting all system variables.
Senthilkumar [28]	XGBoost, LR, and RF	COP prediction of R600a	XGBoost achieves the highest accuracy with (MSE) of 0.00465.
Current study	LR, PR, RF, DT, and SVM	Predict the COP of types of refrigerants	-Polynomial Regression of degree 3 achieves an R^2 of 0.999857 and the lowest error metrics (RMSE = 0.007084, MAE = 0.005254). -Random Forest reached R^2 of 0.993429, RMSE of 0.048037, and MAE of 0.036628.

4. Conclusions

This study applied machine learning techniques to predict the coefficient of performance of several refrigerants in a single-stage vapor-compression refrigeration system. COP was modeled as a function of five inputs: refrigerant type, evaporation temperature, condensing temperature, subcooling, and superheat. A synthetic dataset with 2,000 operating points was generated in EES for four low-GWP refrigerants (R1234yf, R134a, R290, and R600a) and then imported into Python for model development. Five regression algorithms were trained and tested: linear regression, polynomial regression, random forests, decision trees, and support vector machines. The results show that COP can be estimated with high accuracy using relatively simple models. Among the tested methods, degree 3 polynomial regression and random forests achieved the best performance. The cubic polynomial model achieved an R^2 of 0.999857, RMSE of 0.007084, and

MAE of 0.005254 on the test set, while the random forest also achieved high accuracy and robustness. The machine learning models proposed here offer refrigeration engineers a fast alternative to traditional thermodynamic calculations. Once trained, these models allow COP to be predicted directly from a small set of operating parameters, thereby reducing the time required for performance estimation and enabling quicker system design and optimization. The explicit polynomial expression is particularly useful, as it can be coded in spreadsheets or basic software without any need for retraining. Combined with information on safety and environmental constraints, the models can assist in comparing refrigerant options and selecting suitable operating conditions for food storage applications and similar vapor compression systems. This work is limited to simulated steady-state data for a single-stage cycle and to a restricted set of machine learning methods. Future research should validate the models with experimental or field measurements, extend the analysis to other system configurations and refrigerants, and investigate additional algorithms, such as gradient boosted trees or deep learning approaches. Integrating the COP models into multi-objective optimization frameworks that also account for cost and environmental impact would further enhance their practical value. Even though the degree-3 polynomial model achieves an R^2 of 0.999857 on the test set, this very high value may be influenced by the smooth, ideal nature of the simulated data. For this reason, the model should be applied with care outside the studied operating range, and future studies should assess its performance using experimental or real-world system data.

Acknowledgements
The authors thank Van Lang School of Technology, Van Lang University, Vietnam, for providing the laboratory facilities that enabled this work.

Funding
This study was not supported by any grants from funding bodies in the public, private, or not-for-profit sectors.

Declaration of Competing Interest
The author declares no conflicts of interest.

CRediT Authorship Contribution Statement
Tue Nguyen Duy: Methodology, Experimental Investigation, Draft Preparation, and Supervision
Vo Van An: Methodology, Reviewing, and Editing

Availability of Data and Materials
The data supporting this study's findings are available on request from the corresponding author.

Ethics Declarations
This study did not involve human participants or animals. Ethical approval was therefore not required.

Generative Artificial Intelligence Declarations
The authors stated that generative AI was not used to generate content, ideas, or theories. We have just utilized AI to enhance readability and refine the language. This was used with extreme human control and oversight. The authors take full responsibility for reviewing and approving the content.

References
[1] M. Mshragi, I. Petri, "FastML-GA: FPGA-accelerated machine learning for real-time energy HVAC optimization in buildings," *Neural Computing and Applications*, vol. 38, no. 25, pp. 1–18, 2026.
[2] D. Zhou, Z. Zhu, "A digital twin model for cop prediction in refrigeration system using combined machine learning method, "*In IEEE Electrical Power and Energy Conference,* Victoria, BC, Canada, pp. 412-417, 2022
[3] O. Pektezel, H.I. Acar, "Experimental comparison of R290 and R600a and prediction of performance with machine learning algorithms, " *Science and Technology for the Built Environment*, vol. 29, pp. 1-17, 2023
[4] H. Wang, D. Mai, Q. Li, Z. Ding, "Evaluating machine learning models for HVAC demand response: the impact of prediction accuracy on model predictive control performance," *Buildings,* vol. 14, no. 7, pp. 1-17, 2024
[5] Q. Deng, Z. Chen, W. Zhu, Z. Li, Y. Yuan, W. Gui, "A performance prediction method for on-site chillers based on dynamic graph convolutional network enhanced by association rules," *Building Simulation*, vol. 17, pp. 1213–1229, 2024.
[6] T.D. Nguyen, H.M Bui "Machine learning for fault detection in refrigeration systems using difluoromethane refrigerant," *Sustainable Chemistry One World*, vol. 5, art. no. 100085, 2025.
[7] H.U. Cho, Y. Nam, E.J. Choi, Y.J. Choi, H. Kim, S. Bae, J.W. Moon, "Comparative analysis of the optimized ANN, SVM, and tree ensemble models using Bayesian optimization for predicting GSHP COP, " *Journal of Building Engineering,* Vol. 44, p. 103411, 2021
[8] A. Senthilkumar, "Prediction and performance enhancement of R600a refrigerator adopting machine learning algorithms," *Research on Engineering Structures & Materials*, vol. 11, pp. 3207-3223, 2025.
[9] F.W. Yu, W.T. Ho, C.F.J. Wong, "Improved energy management of chiller system with AI-based regression," *Applied Soft Computing*, vol. 150, p. 111091, 2024

[10] M. Ghanbarpour, A. Mota-Babiloni, P. Makhnatch, B.E. Badran, J. Rogstam, R. Khodabandeh, “ANN modeling to analyze the R404A replacement with the low GWP alternative R449A in an indirect supermarket refrigeration system, ” *Applied Sciences*, vol. 11, no. 23, 2021

[11] R. Babu, R. Srikanth, “Design and analysis of different material to increase the performance of AC condenser,” *International Journal of Research*, vol. 3, no. 11, pp. 518–547, 2016.

[12] O. Altuntas, M.Z. Sogut, E. Yalcin, T.H. Karakoc, “Assessment of thermodynamic and environmental performances in subcooling process for different refrigerants, ”*International Journal of Exergy*, vol. 24, no. 2–4, pp. 216–234, 2017.

[13] P.W. Sunu, D.A. Simon, I.M. Rasta, Y.B. Lukiyanto, I.D.M.Susila, I.K. Suarsana, “Effect of various superheat condition on working condition of air-cooled chiller, ”*International Conference on Applied Science and Technology,* 2018.

[14] R. Dharmkar, J. Yeboah, I. Nti, “Smart temperature management in buildings using predictive analysis by machine learning algorithms,” *International Conference on Computational Science and Computational Intelligence*, pp. 284-290, 2023

[15] J. Marchante-Avellaneda, J.M. Corberan, E. Navarro-Peris, S.S. Shrestha, “A critical analysis of the AHRI polynomials for scroll compressor characterization,” *Applied Thermal Engineering*, vol. 219, p. 119432, 2023

[16] K. Nong, H. Zhang, Z. Liu, “Comparative study of different machine learning models for heat transfer performance prediction of evaporators in modular refrigerated display cabinets, ” *Energies*, vol. 17, no. 23, 2024

[17] A. Dhakar, B. Singh, P. Gupta, “Fault diagnosis of air compressor set-up using decision tree-based J48 classification algorithm,” *Journal of Engineering Research*, vol. 12, pp. 1011-1025, 2024.

[18] W. Tun, J.K. Wong, S. Ling, “Hybrid random forest and support vector machine modeling for HVAC fault detection and diagnosis, ” *Sensors,* vol. 21, no. 24, p. 8163, 2021

[19] A. Ivanov, S.M. Ailuro, “TMPNN: High-order polynomial regression based on Taylor map factorization, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 11, pp. 12726-12734, 2024

[20] E. Halabaku, E. Bytyçi, E., “Overfitting in machine learning: a comparative analysis of decision trees and random forests, ” *Intelligent Automation and Soft Computing*, vol. 39, no. 6, pp. 987–1006, 2024.

[21] A. Al-Rashed, “Effect of evaporator temperature on vapor compression refrigeration system,” *Alexandria Engineering Journal*, vol. 50, pp. 283-290, 2011

[22] M.A. Ismael, S.G. Yahya, M. Azhar, I.M. Mahbubul, O. Ilori, “Improving the efficiency of a conventional vapor-compression refrigeration system used for geothermal and evaporative cooling techniques: case study in Iraq,” *DJES*, vol. 16, no. 1, pp. 15–26, 2023

[23] P. Singha, C. Das, M.S. Dasgupta, M.K. Sahu, “New approach to improve COP and heat recovery in transcritical CO_2 refrigeration system for milk processing application,” *Scientific Reports*, vol. 15, art. no. 5502, 2025.

[24] M. Ma, O. Pektezel, V. Ballerini, P. Valdiserri, E.R. di Schio, “Performance predictions of solar-assisted heat pumps: methodological approach and comparison between various artificial intelligence methods,” *Energies*, vol. 17, no. 22, p. 5607. 2024.

[25] X. Ye, J. Lin, Y. Liu, C. Zhuo, S. Xiong, J. Zhao, et al., “Performance prediction of R290 refrigerant scroll compressors based on multitask Lasso-XGBoost algorithm,” *Applied Thermal Engineering*, Vol. 285, pp. 129285, 2026.

[26] S. Akyol, M. Das, B. Alatas, “Modeling the energy consumption of R600a gas in a refrigeration system with new explainable artificial intelligence methods based on hybrid optimization,” *Biomimetics,* vol. 8, pp. 1-20, 2023.

[27] A.M. Ashour, A.M. Al Jubori, A. Mota-Babiloni, H.M. Ali, H.M. Jaffar, A. Bouabidi, “Performance assessment of a hybrid vapor compression-thermoelectric refrigeration system using machine learning models,” *Energy Conversion and Management*, vol. 348, art. no. 120627, 2026.

[28] A. Senthilkumar, A., “Enhancing COP in R600a refrigerators by dispersing CuO/SiO_2 Nanolubricants: A machine learning prediction,” *Jordan Journal of Mechanical and Industrial Engineering*, vol. 19, no. 3, pp. 547–556, 2025.

Abbreviation

AI – Artificial Intelligence
ANN – Artificial Neural Network
CNN – Convolutional Neural Network
COP – Coefficient of Performance
DL – Deep Learning
DT – Decision Trees
GPR – Gaussian Process Regression
GWP – Global Warming Potential
HVAC – Heating, Ventilation, and Air Conditioning
KNN – K-Nearest Neighbours
LR – Linear Regression
LSTM – Long Short-Term Memory
MAE – Mean Absolute Error
MAPE – Mean Absolute Percentage Error
ML – Machine Learning

MLP – Multilayer Perceptron
ODP – Ozone Depletion Potential
PR – Polynomial Regression
PSO – Particle Swarm Optimisation
R^2 – Coefficient of Determination
RBF – Radial Basis Function
RF – Random Forest
RMSE – Root Mean Squared Error
SVM – Support Vector Machine
SVR – Support Vector Regression