COBEM 2023 27TH INTERNATIONAL CONGRESS OF MECHANICAL ENGINEERING

ABCM Brazilian Society of Mechanical Sciences and Engineering
27th ABCM International Congress of Mechanical Engineering
December 04-08, 2023, Florianópolis-SC, Brazil

COB-2023-0986
INTELLIGENT REGRESSION MODELING FOR PERFORMANCE PREDICTION OF A VAPOR COMPRESSION REFRIGERATION PROTOTYPE USING MACHINE LEARNING TECHNIQUES

Paulo Rafael Costa Silva
Evandro Pedro Alves de Mendonça
Sérgio da Silva Franco
Alvaro Antonio Villa Ochoa
Gustavo de Novaes Pires Leite
José Ângelo Peixoto da Costa
Kilvio Alessandro Ferraz
Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco, Campus Recife – IFPE,
Av. Prof. Luiz Freire, 500, Cidade Universitária, Recife/PE – CEP: 50740-545.
prcs@discente.ifpe.edu.br
evandromendonca@recife.ifpe.edu.br
sergiofranco@recife.ifpe.edu.br
ochooaalvaro@recife.ifpe.br
gustavonovaes@recife.ifpe.edu.br
angelocosta@recife.ifpe.edu.br
kilvioferraz@recife.ifpe.edu.br

Abstract. *Vapor compression refrigeration systems are essential for cooling and refrigeration in various settings, but they require maintenance and optimization for optimal performance. Analyzing data on their behavior can provide valuable insights for improving efficiency and lifespan. This study focuses on analyzing the energy performance of a refrigeration prototype using the R404A fluid and tube-in-tube evaporator. The prototype consists of two circuits: one with the R404A fluid refrigeration system and the other with ethylene glycol and thermal load provided by electrical resistances. The system incorporates an integrated supervisory system for data acquisition and control. The collected data includes temperatures, pressures, power, voltage, current, and consumed energy. The study aims to predict the system's performance using intelligent regression models based on linear regression, decision tree, random forest, and artificial neural network techniques. The dataset is divided into training, validation, and test sets, and performance parameters such as determination coefficient, root mean square error and mean absolute error are used to evaluate the models. The results show that the proposed approach accurately estimates the transient behavior of the refrigeration system and can be used to optimize its performance, making it a valuable tool for dynamic analysis of vapor compression equipment operating at full or partial loads.*

Keywords: *Vapor Compression Refrigeration, Machine Learning, Regression Modeling, Transient Behavior, Performance Prediction.*

1. INTRODUCTION

Refrigeration is a process that has been gaining increasing attention over the years, considering its essential role in thermal comfort as it allows for temperature and relative humidity control of the air (Novaes Pires Leite et al., 2019). Moreover, it is crucial for preserving various food and pharmaceutical products, among others, which typically require lower temperatures to maintain their properties or prevent spoilage (Salehy et al., 2020). Mechanical compression refrigeration is the commonly used method, encompassing both household refrigerators and industrial cooling systems (Xia et al., 2021). This process is based on thermodynamic principles and follows a thermodynamic cycle, involving four main components: compressor, condenser, expansion valve, and evaporator, with the presence of a refrigerant fluid that passes through them (Wu et al., 2021).

The need to enhance control and operation systems in refrigeration systems has been studied by several authors (Dong et al., 2022; Franco et al., 2022; Huang et al., 2023; Usman et al., 2017) with the aim of maximizing cooling capacity, reducing energy consumption, and achieving higher coefficient of performance (COP) (Franco et al., 2022, 2019). A promising methodology involves artificial intelligence techniques utilizing machine learning algorithms (Adelekan et al., 2022; Ho and Yu, 2021; Panahizadeh et al., 2021; Zhang et al., 2018), which are capable of identifying behavioral patterns in processes based on previously obtained data, thereby improving system control and adapting it to real situations with greater precision, ultimately enhancing system efficiency. This can be observed in the work conducted by (Sousa

Alcântara et al., 2023), where intelligent regression models were used to simulate the energy behavior of an absorption chiller using temperature and flow parameters, demonstrating a potential alternative for the analysis and study of such equipment.

Considering the growing demand for refrigeration system improvement, the present study aims to optimize a refrigeration prototype by implementing regression models using machine learning methods with real operation data from a refrigeration system. The objective is to predict system behavior based on operational parameters such as temperature, pressure, flow rate, electricity, among others.

2. THEORETICAL FRAMEWORK

2.1 Vapor compression refrigeration

A mechanical compression refrigeration system, as shown in Figure 1, is divided into four parts: compression, condensation, expansion, and evaporation, and operates based on the concepts of heat and work, using a refrigerant fluid (Ferraz, 2008). In the compression stage, the fluid is compressed and begins to circulate in the system. The fluid arrives at the condenser as superheated vapor and is cooled there until it condenses. It is then directed to the expansion valve, where it undergoes pressure reduction, causing a portion of the liquid to vaporize. The mixture then reaches the evaporator, which absorbs heat from the refrigerated environment, transforming the liquid into vapor. This vapor is then sent back to the compressor, restarting the cycle (Oliveira, 2021).

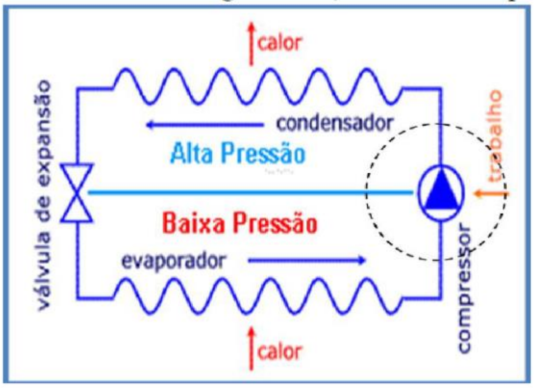

Figure 1. Compression refrigeration cycle.

2.2 Machine Learning

In 1959, Arthur Samuel was one of the researchers who paved the way for machine learning. He conducted his work based on the game of checkers, using hundreds of recorded games to refine his algorithm, which eventually became capable of winning a game against a person (Gabel, 2019). As a subfield of artificial intelligence, machine learning enables computers to learn and improve their performance through a dataset (Calanca et al., 2023). A machine learning algorithm can be classified based on the task it performs, including regression, classification, clustering, data reduction, or anomaly detection (Kang et al., 2020). It also has four main subdivisions:

- Supervised learning: Computers are trained using pre-labeled input data and their corresponding outputs. This approach is typically used when there is prior knowledge of the desired results. The more data available, the better the final outcome.
- Unsupervised learning: Unlike supervised learning, unsupervised learning operates with unlabeled data. It is used to discover patterns or unknown relationships within the data (Oliveira, 2021).
- Semi-supervised learning: This approach utilizes both labeled and unlabeled data for training. It is employed when there is limited labeled data available.
- Reinforcement learning: In reinforcement learning, an agent is trained to make a sequence of decisions and receives rewards or penalties in return. This encourages the agent to take actions that maximize the rewards (Calanca et al., 2023).

3. METHODOLOGY

The work was developed through two aspects. The first aspect involved experimental investigation to gather data from a vapor compression refrigeration system. The second aspect focused on the implementation of machine learning methods.

The Steps followed in this work are shown in Figure 2.

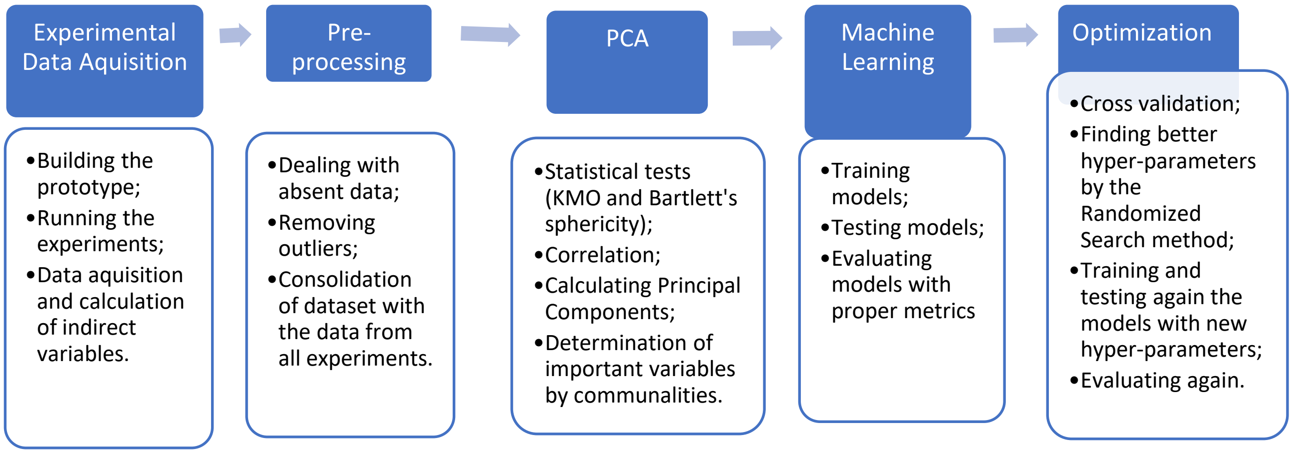

Figure 2. Methodology.

3.1 Experimental apparatus

The experimental analysis was conducted using a vapor compression refrigeration prototype, as shown in Figure 3, which utilizes a dual-tube evaporator and the R404A refrigerant. The system includes a secondary operating circuit with a 50 % ethylene glycol solution as the cooling fluid. The prototype was fully instrumented with sensors to measure various operational parameters such as temperatures, pressures, current, voltage, power factor, energy consumption, flow rate, among others. It was monitored using an automated supervisory system, the Sitrad software (Full Gauge Controls). The circuit incorporates an electrical resistance as a thermal load source for the refrigeration system, with a power rating of 625 W.

Figure 3. Vapor compression refrigeration prototype.

3.2 Data collection

The collected data regarding temperatures, pressures, and electrical parameters in the refrigeration prototype were obtained during tests lasting between two and three hours. Data collection was automated, occurring every five seconds, resulting in 72008 values for each variable. All data were collected within a closed room at a constant average temperature of 25 °C.

To determine the enthalpies of the states in the refrigeration cycle, as well as the variables related to ethylene glycol, the Engineering Equation Solver (EES) software was utilized. The Python platform was used to calculate the heat in the evaporator (Q_{ev}), Eq. (1), the electrical coefficient of performance (COP), Eq. (2), the thermal coefficient of performance Eq. (3), and the Energy Efficiency Ratio (EER) Eq. (4).

$$Q_{ev} = \dot{m}_1 \times \Delta h \tag{1}$$

$$COP_{ele} = Q_{ev}/Pow_{active} \tag{2}$$

$$COP_{therm} = Q_{ev}/[\dot{m}_2 \times (h_{discharge} - h_{suction\ evap})] \tag{3}$$

$$EER = (Q_{ev} \times h \times 3{,}412)/E_{active} \tag{4}$$

where $\dot{m}_1$, Δh, ΔT, $\dot{m}_2$, $h_{discharge}$, $h_{suction\ evap}$ and h are the mass flow rate of the solution, the enthalpy variation in the condensation process, the temperature variation in the evaporator, the mass flow rate of the R404A fluid, the discharge enthalpy, the suction enthalpy of the evaporator, and the time in hours, respectively. The evaporator heat was obtained through the mass flow rate of ethylene glycol, specific heat and the temperature difference, considering that the heat of the solution is equal to the heat absorbed by R404A, which is calculated from Eq. (1).

3.3 Data analysis

This step is focused on examining the variables of interest by creating a correlation matrix and identifying numerical variables that show the highest correlation with the target variables, namely the COP and EER of the system.

For the next step, it was necessary to remove outliers in order to calculate the Kaiser-Meyer-Olkin (KMO) statistic, which is a criterion for assessing whether a factor analysis model is well-suited to the data by testing its overall consistency. In addition to this criterion, the Bartlett's test of sphericity was also performed. This test involves comparing the correlation matrix "ρ" with an identity matrix "I" of the same dimension. If the differences between the corresponding values off the main diagonal of each matrix are not statistically different from zero at a certain level of significance, it can be considered that factor extraction will not be appropriate.

3.4 Principal Components Analysis (PCA)

The factor analysis procedure involves first obtaining eigenvalues and eigenvectors from the correlation matrix of the data. The eigenvector is divided by the square root of its corresponding eigenvalue, generating factor scores for each variable. These factor scores are then multiplied by the standardized variables, and the results are summed for each of the factors, generating the Principal Components (PCs). For the implementation of Principal Component Analysis (PCA), it was necessary to standardize the variables using the ZScore method Eq. (5).

Based on the analysis, the variance of the factors was examined, with the number of factors determined by the Kaiser criterion. Using computational tools, it was observed that the first eight factors explained 92% of the variance.

$$Z = (X - \mu) / \sigma \tag{5}$$

where Z, X, μ e σ are the ZScore value, the individual value to be normalized, the mean of the data series and the standard deviation, respectively.

3.5 Regression Methods

The regression methods used were multiple linear regression (LR), decision tree regression (DT), random forest regression (RF), and artificial neural networks (NN), implemented using the Python libraries Scikit-Learn (Pedregosa et al., 2011) and Keras, which is part of the TensorFlow.

In the process of identifying the most important variables for the dataframe, the communalities procedure was used to select essential variables for regression prediction models. These variables range from 0 to 1, where 0 indicates that none of the variance is explained by the principal components, and 1 indicates that all the variance is explained. Based on the analysis of correlations and communalities, the step of removing variables that are not important for the model was initiated. For communalities, the criterion used was that if the value was below 0.8, the corresponding variable would be removed, with the exception of EER, as it is an output variable. Regarding correlations, variables with a value greater than 0.9 with more than six other variables and those with values less than 0.1 with the output variables were removed. Finally, with the selection of important variables, the regression models were constructed.

Four machine learning models were built: multiple linear regression, decision tree regression, random forest regression, and neural network, aiming to compare their performance and predict the behavior of the COP and EER of the system.

The dataset was divided into two other sets: training and test sets, with 75 % and 25 % of the original dataset, respectively.

3.5.1 Multiple linear regression (LR)

The data will be separated into independent variables, which are chosen through PCA, and dependent variables: COP and EER. The model was created using the LinearRegression() object from the sklearn (Sci-kit Learn) library and subsequently trained using the input and output variables.

Unlike simple linear regression, this model uses more than one independent variable to predict the output value.

3.5.2 Decision Tree for regression

The decision tree model is developed by splitting the data based on different criteria to minimize prediction errors. It was created using the DecisionTreeRegressor() object, from sklearn, and trained with the input and output data. Afterwards, predictions can be made from the test data.

3.5.3 Random forest para regressão

Just like the decision tree, the random forest is built following the same principles. However, what sets these two models apart is that, while the decision tree is an individual model, the random forest is a collection of decision trees that work together to provide a more accurate result. This is achieved through the technique of ensemble, which combines multiple simple data structures to form a more complex one that can generalize better and yield improved results. The model was created using the RandomForestRegressor() object from the sklearn library.

3.5.4 Rede neural artificial

To create the neural network model, the data will first need to be normalized. Then, the architecture will be defined, followed by compilation, and finally, the model will be trained. The architecture design will consider a regression problem, and the compilation stage will also be based on the same problem. The loss function and optimizer will be mean squared error and Adam, respectively. The defined architecture consists of 10 nodes in the input layer, 10 nodes in the hidden layer, and 2 nodes in the output layer, corresponding to the two output variables (COP and EER). The neural network model was created, trained, and tested using objects from the Keras library in TensorFlow.

3.6 Metrics for evaluation

The models' performances were evaluated according to the following metrics: coefficient of determination (R^2), mean absolute error (MAE) and root mean square error (RMSE).

- Coefficient of determination (R^2), which ranges from -inf to 1. The higher the value of the coefficient, the more explanatory the model is in relation to the predicted data. A value of 0 means that the model could predict only the mean value of the test set. Negative values indicate that the model didn't adjust properly to the input data;
- Mean Absolute Error (MAE), which measures the average difference between the actual and predicted values. It varies from 0 to +inf, and values closer to 0 are ideal as they indicate a smaller error;
- Root Mean Square Error (RMSE), which calculates the average difference between the predicted and actual values, similar to the MAE metric. However, instead of using the absolute value of the difference between y and $\hat{y}$, this metric squares the difference and then takes the square root. Its result can be interpreted in the same way as MAE, but its value is less sensitive to the variance, in comparison to the MAE.

4. RESULTS

This section presents the results of the intelligent regression models built using real operational data from the refrigeration prototype to predict the energy behavior of the system based on COP and EER.

4.1 Data collection and analysis

Data collection was carried out at the Federal Institute of Pernambuco – *Campus* Recife facilities using the refrigeration prototype, with the collection of 24 variables. Subsequently, with the help of the Engineering Equation Solver and Python, values for an additional 35 variables and 4 constants were obtained.

After processing all the data, some variables and all constants were removed in order to retain only the most important data for model creation. The elimination of these data was based on previously mentioned criteria regarding correlation analysis and communalities.

With the reduced variables, a final analysis (Figure 4) was conducted to confirm that the correlation between all of the variables wasn't too high (greater than 0.9), to avoid redundancy, and the correlation between input and output variables is greater than 0.1, to avoid the noise of non-important variables.

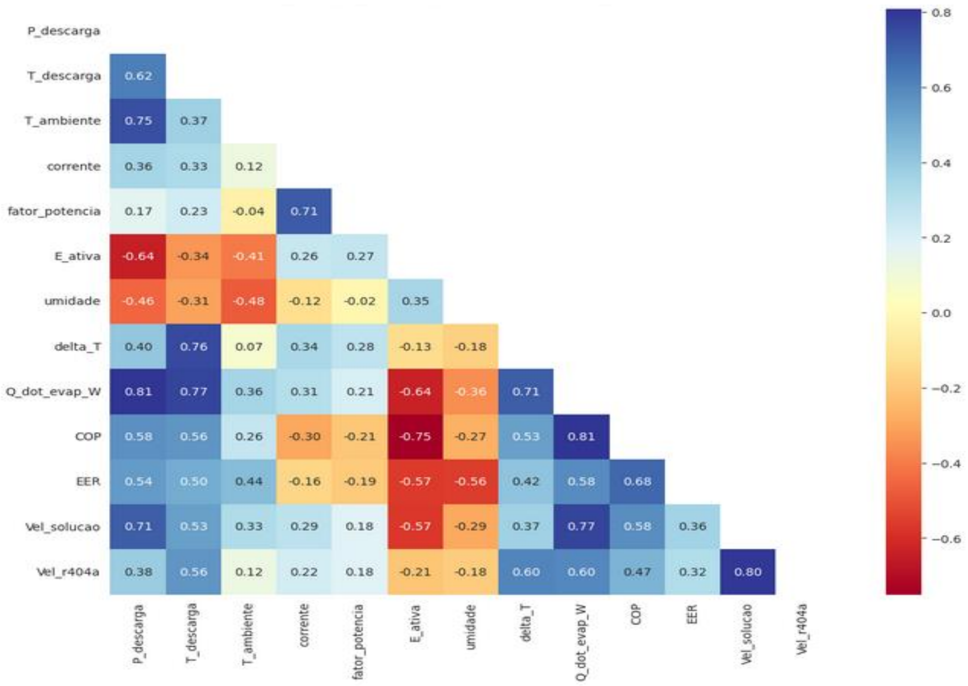

Figure 4. Correlation among the variables used as input for the machine learning models.

4.2 Performance of the Linear Regression model

As seen in Figure 5, the model showed good results for predicting the coefficient of performance, except for some outliers (negative values). However, when it comes to predicting the EER, it can be deduced from the graph in Figure 6 that the error is much higher compared to the COP graph. This suggests that the model is inappropriate for predicting energy efficiency.

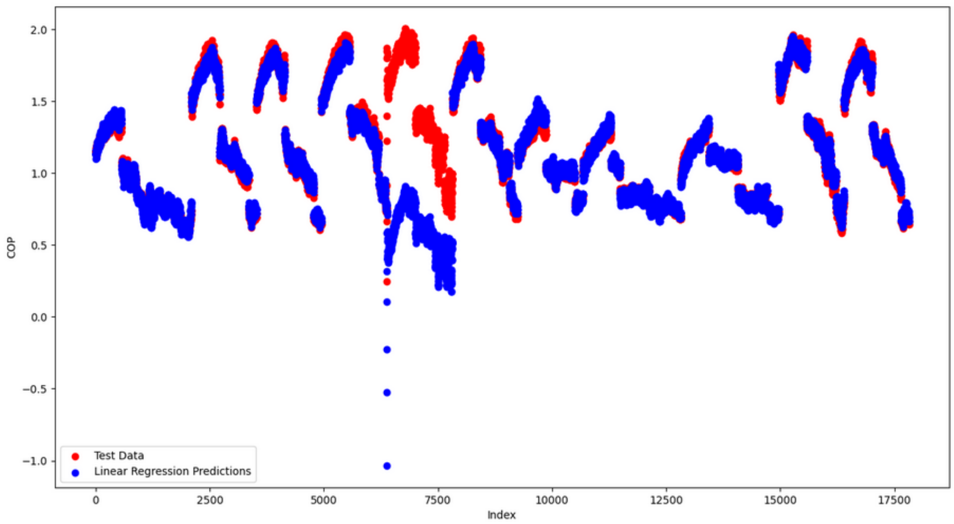

Figure 5. Comparison graph between test data and predictions of multiple linear regression for COP.

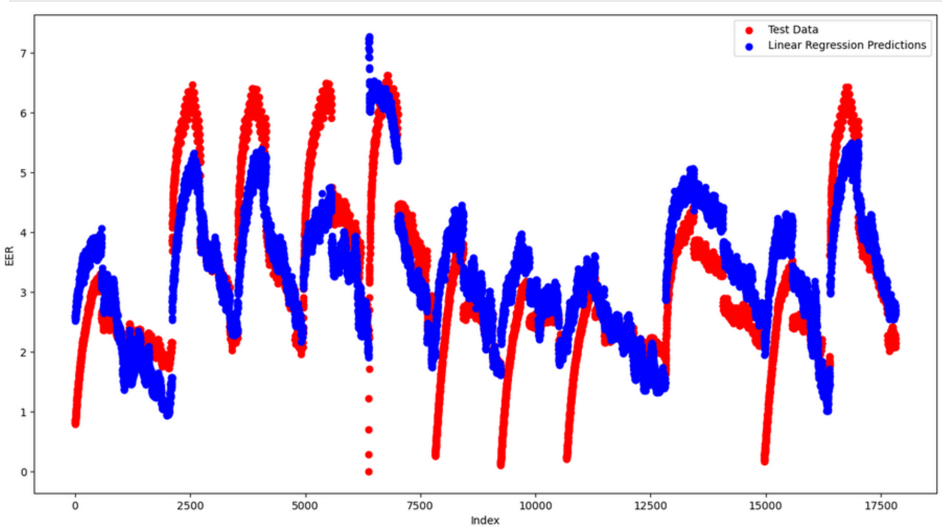

Figure 6. Comparison graph between test data and predictions of multiple linear regression for EER.

4.3 Performance of the Decision Tree model

Similar to the multiple linear regression, the decision tree model showed good results for predicting COP (Figure 7). However, unlike linear regression, it did not have any outliers with negative values, making it more suitable for use. Regarding EER (Figure 8), it also achieved good values, but still had a higher margin of error than desired.

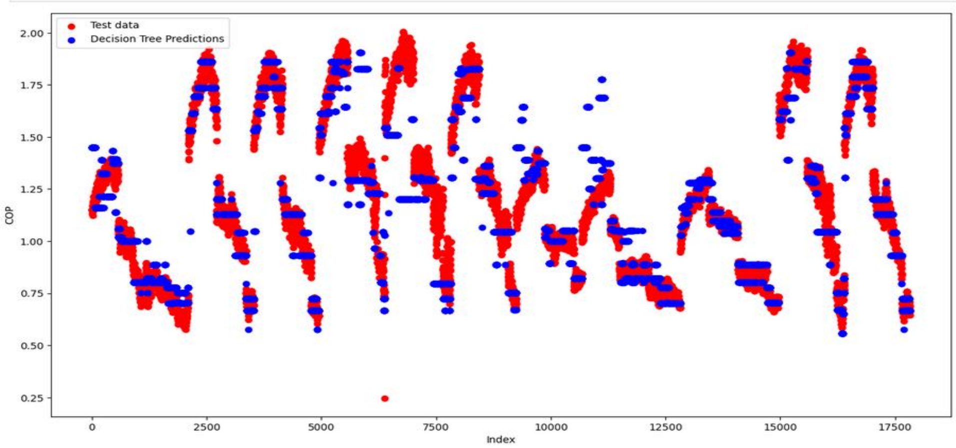

Figure 7. Comparison graph between test data and predictions of the decision tree for COP.

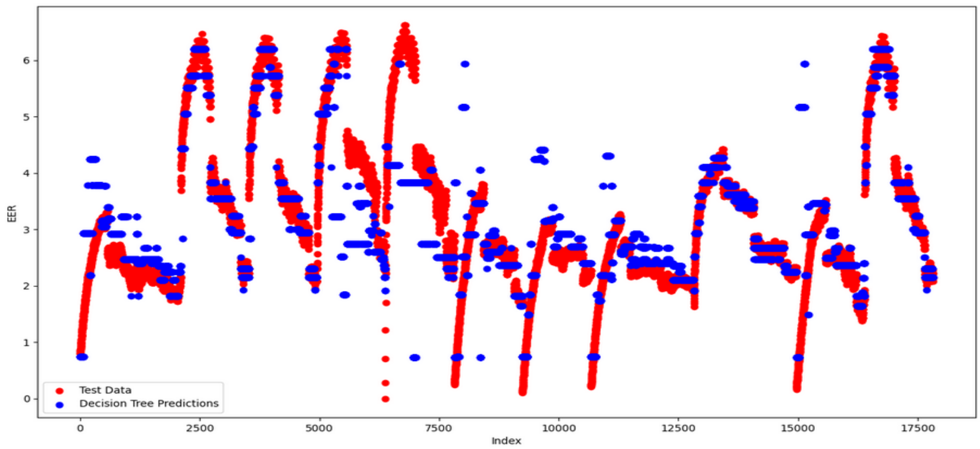

Figure 8. Comparison graph between test data and predictions of the decision tree for EER.

4.4 Performance of the Random Forest model

The random forest model achieved the best results, as it was able to predict both COP and EER with the lowest error among the four models. The values obtained for COP (Figure 9) were better compared to EER (Figure 10). When analyzing the performance metrics, the results for the coefficient of performance were close to the ideal values.

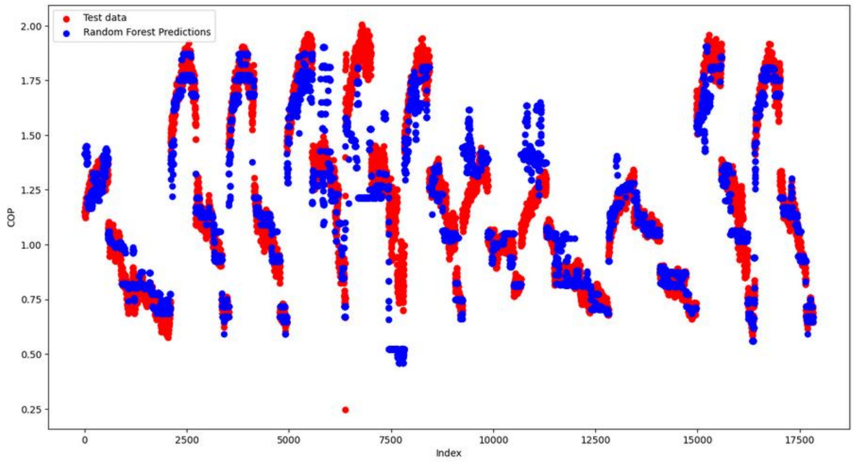

Figure 9. Comparison graph between test data and predictions of the random forest for COP.

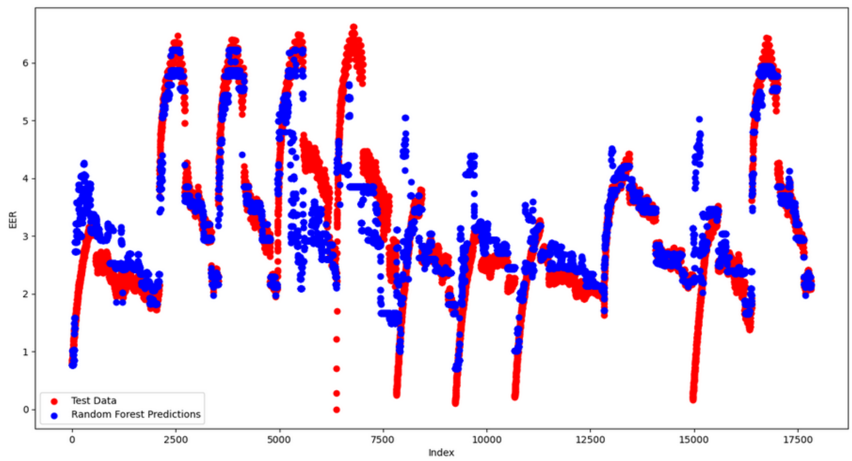

Figure 10. Comparison graph between test data and predictions of the random forest for EER.

4.5 Performance of the Neural Network model

The neural network model achieved results somewhat similar to multiple linear regression, meaning it is capable of predicting COP more accurately than EER. This can be supported by analyzing the prediction graphs for COP and EER (Figures 11 and 12). In the first graph, there is a close resemblance between the test data and the predicted data, while in the second graph, there is a great offset and scale difference between the predicted data and the true data.

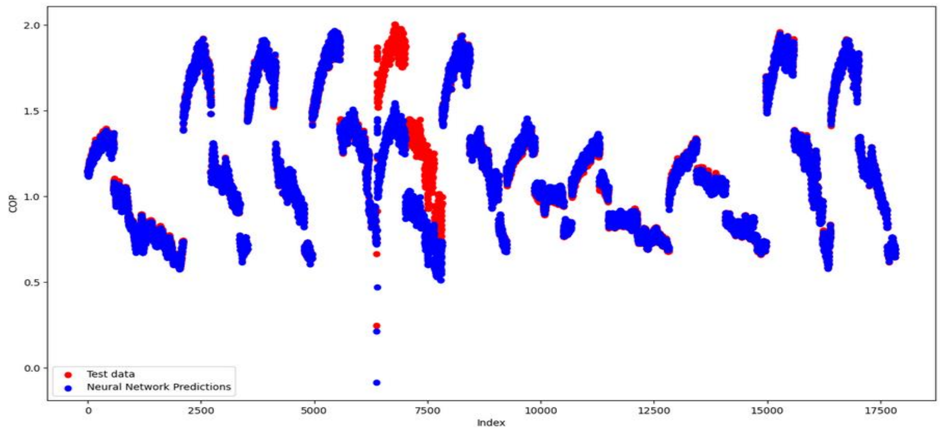

Figure 11. Comparison graph between test data and predictions of the neural network for COP.

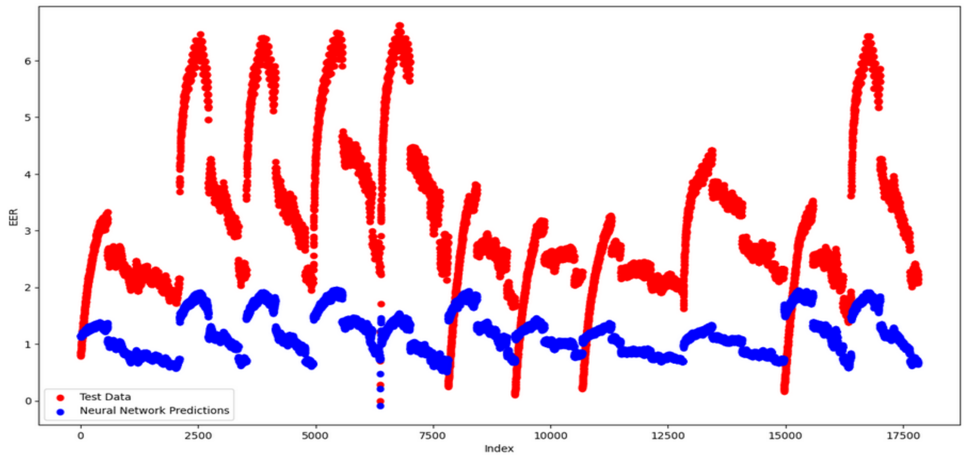

Figure 12. Comparison graph between test data and predictions of the neural network for EER.

4.6 Discussion of the results

After the graphical analysis of each model, the performance of all models was evaluated using three different methods: RMSE, MAE, and R^2. This evaluation was done both overall and for each output variable. Through this evaluation, it was concluded that the models are not very effective in predicting EER. However, all of them are able to predict COP values with small errors, as observed in Table 1.

Table 1. Values related to the overall and specific performance of each model

Model	RMSE	MAE	R^2	RMSE (COP)	MAE (COP)	R^2 (COP)	RMSE (EER)	MAE (EER)	R^2 (EER)
LR	1.833	0.672	-0.914	0.155	0.060	0.825	2.587	1.283	-2.652
DT	0.581	0.251	0.725	0.164	0.077	0.804	0.805	0.426	0.647
RF	0.518	0.211	0.826	0.090	0.044	0.941	0.727	0.378	0.712
NN	1.686	1.076	-0.600	0.122	0.041	0.892	2.381	2.111	-2.092

5. CONCLUSION

The results achieved through this work were satisfactory, considering the performance evaluation method used. Two out of the four models showed values close to the ideal, both overall and for each of the output variables. These models were the decision tree and the random forest, which were able to make predictions with the lowest error rate for both EER and COP, with the random forest achieving the best result. For COP, the errors obtained for RMSE, MAE, and R^2 were 0.089, 0.043, and 0.941, respectively. For EER, the values obtained were 0.727, 0.377, and 0.711, respectively. The other models performed well only in predicting the coefficient of performance, where they obtained better values than the decision tree. However, for predicting the other dependent variable, they generated significantly higher errors.

With the random forest model, it is possible to make efficiency predictions for the refrigeration bench, and with further studies, it would be possible to apply this model to various refrigeration systems or even create more effective models to further reduce the prediction error.

6. ACKNOWLEDGEMENTS

All the authors thank the Fullgauge Company and the IFPE for its support through this project's development. The authors also thank the CNPq for financial support to the research project - Universal 402323/2016-5. The fourth author thanks the CNPq for the Productivity Grant 3303417/2022-6.

7. REFERENCES

Adelekan, D.S., Ohunakin, O.S., Paul, B.S., 2022. Artificial intelligence models for refrigeration , air conditioning and heat pump systems. Energy Reports 8, 8451–8466. https://doi.org/10.1016/j.egyr.2022.06.062

Dong, X., Ma, Y., Wang, Y., Chen, Q., Liu, Z., Jia, X., 2022. An improved power flow calculation method based on linear regression for multi-area networks with information barriers. Int. J. Electr. Power Energy Syst. 142, 108385. https://doi.org/10.1016/j.ijepes.2022.108385

Franco, S.S., Henríquez, J.R.., Ferraz, K.A.., Da Costa, J.A.P.., Michalewicz, J.S.., Oliveira, R.E.., De Lima, T.L.A.., Ochoa, Á.A.V., 2019. Análise experimental do desempenho de um protótipo de refrigeração por compressão mecânica para conservação de alimentos através do uso de dispositivos de expansão Termostática e eletrônica, in: XVI CONBRAVA - CONGRESSO BRASILEIRO DE REFRIGERAÇÃO, AR-CONDICIONADO, VENTILAÇÃO, AQUECIMENTO E TRATAMENTO DO AR São Paulo Expo - 10 a 13 de Setembro de 2019. São Paulo.

Franco, S.S., Henríquez, J.R., Ochoa, A.A.V., da Costa, J.A.P., Ferraz, K.A., 2022. Thermal analysis and development of PID control for electronic expansion device of vapor compression refrigeration systems. Appl. Therm. Eng. 206, 118130. https://doi.org/10.1016/j.applthermaleng.2022.118130

Ho, W.T., Yu, F.W., 2021. Chiller system optimization using k nearest neighbour regression. J. Clean. Prod. 303. https://doi.org/10.1016/j.jclepro.2021.127050

Huang, C., Li, Z., Ye, Z., Wang, R., 2023. Performance analysis of two-stage CO2 transcritical refrigeration cycles with absorption subcooling and mechanical recooling, International Journal of Refrigeration. Elsevier B.V. https://doi.org/10.1016/j.ijrefrig.2023.05.023

Novaes Pires Leite, G. de, Weschenfelder, F., Araújo, A.M., Villa Ochoa, Á.A., Franca Prestrelo Neto, N. da, Kraj, A., 2019. An economic analysis of the integration between air-conditioning and solar photovoltaic systems. Energy Convers. Manag. 185, 836–849. https://doi.org/10.1016/j.enconman.2019.02.037

Panahizadeh, F., Hamzehei, M., Farzaneh-Gord, M., Villa, A.A.O., 2021. Evaluation of machine learning-based applications in forecasting the performance of single effect absorption chiller network. Therm. Sci. Eng. Prog. 26, 101087. https://doi.org/10.1016/j.tsep.2021.101087

Pedregosa, F., Varoquaux, G.., Gramfort, A.., Michel, V.., Thiriron, B.., 2011. Scikit-learn: Machine Learning in Python. J. ofMachine Learn. Res. 12, 2825–2830. https://doi.org/10.4018/978-1-5225-9902-9.ch008

Salehy, Y., Hoang, H.M., Cluzel, F., Leroy, Y., Delahaye, A., Fournaison, L., Yannou, B., 2020. Energy performances assessment for sustainable design recommendations: Case study of a supermarket's refrigeration system. Procedia CIRP 90, 328–333. https://doi.org/10.1016/j.procir.2020.01.102

Sousa Alcântara, S.C., Villa Ochoa, A.A., da Costa, J.Â.P., de Menezes, F.D., de Novaes Pires Leite, G., Arruda Michima, P.S., da Silva Marques, A., 2023. Development of a method for predicting the transient behavior of an absorption chiller using artificial intelligence methods. Appl. Therm. Eng. 231. https://doi.org/10.1016/j.applthermaleng.2023.120978

Usman, M., Imran, M., Yang, Y., Lee, D.H., Park, B.S., 2017. Thermo-economic comparison of air-cooled and cooling tower based Organic Rankine Cycle (ORC) with R245fa and R1233zde as candidate working fluids for different geographical climate conditions. Energy 123, 353–366. https://doi.org/10.1016/j.energy.2017.01.134

Wu, D., Hu, B., Wang, R.Z., 2021. Vapor compression heat pumps with pure Low-GWP refrigerants. Renew. Sustain. Energy Rev. 138. https://doi.org/10.1016/j.rser.2020.110571

Xia, Y., Ding, Q., Jing, N., Jiang, A., Zhang, X., Deng, S., 2021. A review on the operational instability of vapor compression system. Int. J. Refrig. 122, 97–109. https://doi.org/10.1016/j.ijrefrig.2020.10.024

Zhang, W., Liu, F., Fan, R., 2018. Improved thermal comfort modeling for smart buildings: A data analytics study. Int. J. Electr. Power Energy Syst. 103, 634–643. https://doi.org/10.1016/j.ijepes.2018.06.026

8. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.