Data-Driven Performance Mapping of a Transcritical CO_2 Booster System

Dimitrios Dalavouras[a], Michail Lougkos[a], Tryfon C. Roumpedakis[a,], Petros Dalavouras[a], Irene Koronaki[a], and Konstantinos Braimakis[b]*

[a] *Laboratory of Applied Thermodynamics, National Technical University of Athens (NTUA), Greece, troumpedak@mail.ntua.gr*

[b] *Laboratory of Refrigeration, Air-Conditioning and Solar Energy, National Technical University of Athens (NTUA), Greece, mpraim@central.ntua.gr*

Abstract:

Transcritical CO_2 booster cycles are a key enabling technology for low-GWP refrigeration and heat recovery, yet their performance is strongly influenced by operating conditions, control settings, and component interactions under off-design regimes. This study presents a data-driven modelling and experimental investigation of a CO_2 booster refrigeration cycle installed and instrumented at the National Technical University of Athens. A shallow feed-forward Artificial Neural Network (ANN) is developed to predict cycle performance based on measured laboratory data, providing a computationally efficient surrogate to support performance mapping and control-oriented studies. An incremental architecture assessment is performed to identify an optimal shallow network configuration, balancing predictive accuracy and generalization. The selected ANN achieves high predictive performance across the examined operating envelope, with R^2 values in the range of 0.95–0.99 for primary performance indicators, while maintaining stable behavior at part-load and during shifts in high-side control conditions. Results demonstrate that shallow feed-forward ANNs can accurately approximate the nonlinear response of transcritical CO_2 booster systems, enabling rapid performance prediction without the computational burden and parameter-identification effort of detailed physics-based models. The proposed approach supports accelerated experimentation, real-time performance estimation, and efficient generation of performance maps for design, optimization, and supervisory control applications.

Keywords:

Transcritical CO_2 refrigeration; Artificial neural networks; Surrogate modelling; Energy optimization

1. Introduction

As the cooling loads increase, a direct effect of global warming, refrigeration and air-conditioning sectors face a dual challenge: maximizing energy efficiency and at the same time minimizing their environmental footprint [1]. Driven by the phase-out of ozone-depleting substances and increasingly strict regulations on fluorinated refrigerants, mainly associated with the revised EU F-gas Regulation [2], the industry is actively pivoting to more sustainable alternatives, causing a renewed focus on natural refrigerants [3]. Carbon dioxide (R744) is one of the top choices amongst natural refrigerants due to its negligible ozone depletion potential (ODP) and minimal GWP. On the other hand, due to the low critical temperature of CO_2, its high-side heat rejection differs fundamentally from conventional subcritical systems, mainly forced to operate in supercritical conditions [4]. As a result, CO_2 systems' performance is highly sensitive to operating conditions, particularly when ambient temperatures rise.

Several studies discuss the strong dependence between transcritical CO_2 performance and the control of the high-side pressure. Ge and Tassou [5] showed that the efficiency of supermarket booster systems in transcritical operation is governed by several parameters under high ambient conditions, reflecting the complex nature of the problem. In the same direction, Peñarrocha-Alós et al.[6] proposed a real-time optimization and control strategy that continuously tracks maximum-efficiency conditions by regulating compressor speed and back-pressure-valve opening, directly underscoring the importance of discharge-side pressure control for competitive CO_2 operation. Apart from the control based investigations to tackle the efficiency penalties of the basic transcritical cycle, a number of studies focused on architectural improvements. Parallel compression is considered a promising route to higher COP, as discussed by Sarkar and Agrawal [7]. On the other hand, Elbel and Lawrence [8] reviewed advanced ejector configurations as a means of recovering expansion work, while Hafner et al. [9] investigated the multi-ejector concept for supermarket refrigeration. More recently, Tashtoush

et al. [10] investigated a hybrid combination of two-phase ejectors and parallel compression for supermarket applications, revealing an exergy efficiency increase of up to 40%.

While previous research thoroughly explored thermodynamic optimization and architectural modifications, the practical implementation of advanced control requires fast, accurate modelling of key operational variables. Data-driven and surrogate models are gaining interest [11,12], yet a significant gap remains in CO_2 refrigeration as there is a lack in robust models trained on real plant data that can accurately predict directly controllable variables, with current literature focusing on training surrogate models from other simulation results [13]. High-side pressure (HP) dictates COP and power consumption, while being subjected to direct supervisory control actions. This makes HP an ideal variable for control-oriented modelling. Based on the authors' knowledge, there is a disconnect between computationally heavy physics-based optimization studies and lightweight, deployable tools capable of capturing the nonlinear dynamics of transcritical CO_2 systems in real time.

This study addresses the aforementioned gap by proposing a measurement-driven surrogate modelling framework to predict gas-cooler/high-side pressure in a transcritical CO_2 booster installation. Using real operational data across a wide range of conditions, a feedforward artificial neural network (ANN) was developed and evaluated. The study systematically assesses how input-variable selection, dataset size, network architecture, training algorithms, and lagged pressure information impact the model's ability to capture system behaviour. Ultimately, this work provides both a highly accurate predictive model for a critical control variable and a framework for selecting modelling parameters for CO_2 systems. The findings of the study indicate that thermodynamically informed inputs and extensive datasets drastically improve accuracy. The most effective model, a five-hidden-layer architecture trained using Bayesian Regularization algorithm, demonstrates strong generalization across diverse operating regimes, paving the way for its use in real-time monitoring, optimization, and advanced control.

2. System description and methodology

2.1. The NEVA system

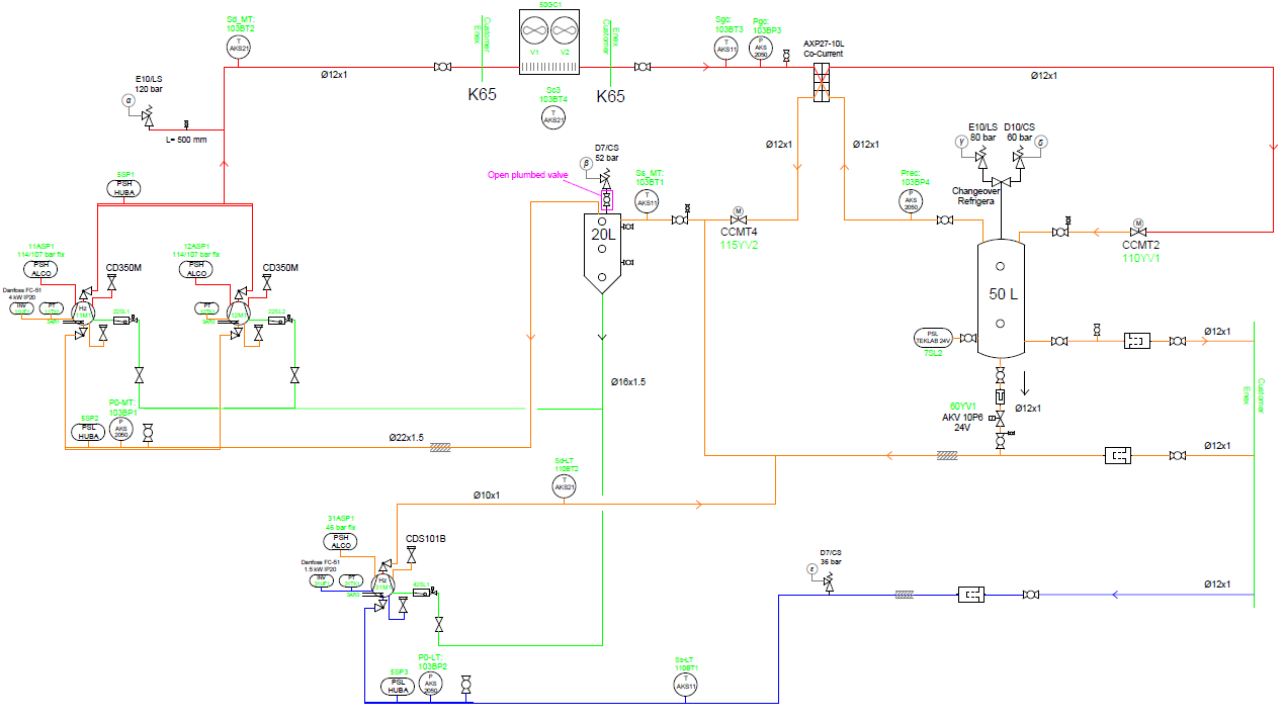
The experimental system considered in this study is a transcritical CO_2 refrigeration unit installed in a supermarket in Athens, Greece, serving a large number of medium-temperature and low-temperature display cases and cold rooms. The unit designed by ENEX, is a NEVA industrial-grade CO_2 booster system intended for commercial food-retail applications, using R744 as refrigerant and simultaneously covering medium-temperature (MT) and low-temperature (LT) cooling loads. The NEVA model series is available in a wide capacity range, depending on the number of compressors employed, with nominal capacities ranging between 70–800 kW for the MT level and 4–200 kW for the LT level. In the installation under investigation, a 3+3 configuration is used, i.e., three compressors on the MT rack and three compressors on the LT rack, with one inverter-driven compressor in each rack. An oil-management system is integrated, including a dedicated oil reservoir and automatic oil-level regulators. Figure 1(a) provides a physical view of the NEVA unit from a product catalogue, while Figure 1(b) summarizes the overall refrigerant circuit and the main functional subsystems, which are more thoroughly discussed below.

On the HP side, the discharge flow from the three MT compressors, after separating the oil content, is directed to the gas cooler. Downstream of the gas cooler, an internal heat exchanger (IHX) provides slight additional cooling of the high-pressure CO_2 stream. The pressure reduction from the HP side to the MT level is achieved through two parallel expansion valves, one electronic expansion valve controlled and one thermostatic expansion valve, improving both controllability and operational reliability. At MT level, a flash tank/receiver is used for phase separation of the two-phase stream generated after expansion. A fraction of the stream is diverted to an auxiliary mechanical subcooling loop, and the cooled branch is then mixed with the flash-gas stream to reduce vapor quality and specific enthalpy, thereby reducing the energetic penalty associated with flash gas recirculation in CO_2 systems [14]. A second IHX downstream of the mixing point cools further the HP fluid before the next expansion stage, limiting the flash-gas generation. At MT level, the liquid stream leaving the bottom of the flash tank supplies both the MT and LT evaporators through the corresponding expansion devices. Before entering the common liquid supply line, this stream passes through a third IHX, where it exchanges heat with the LT return stream. This arrangement provides a dual benefit: slight subcooling of the liquid supplied to the evaporators and controlled superheating of the LT suction gas, thus protecting the LT compressors from liquid carryover. The MT return stream, after absorbing the refrigeration load, mixes with the flash-gas stream that has been expanded down to MT pressure. The combined stream then enters an accumulator placed upstream of the MT compressor suction line, where any entrained liquid is removed. Finally, on the LT side, the discharge of the low-pressure compressors is routed to the MT suction-return line, so that the MT compressors ultimately handle the return flow from the MT evaporators, the processed flash gas, and the LT compressor discharge.

Measurements were recorded for 32 operating variables from 31 May 2025 to 1 September 2025, with a sampling interval of 1 min, resulting in 132,740 samples per variable. The recorded variables include directly

measured thermodynamic quantities, such as temperatures and pressures at key locations of the cycle, as well as controller-based operational variables describing the current control state of the plant. For the MT side, the available signals include ambient dry-bulb temperature (T_{amb}), MT rack running capacity (RC_{MT}), inverter compressor speed (MT_{Speed}), MT suction pressure ($P_{o,MT}$), saturation temperature corresponding to MT suction pressure ($T_{o,MT}$), gas-cooler/high-side pressure (P_{gc}), high-pressure valve opening degree ($V_{hp,OD}$), MT suction temperature ($T_{s,MT}$), MT discharge temperature ($T_{d,MT}$), gas-cooler outlet temperature (S_{gc}), flash-tank temperature (T_{rec}), and flash-tank pressure (P_{rec}). Similar variables are available for the LT side. Among these variables, the MT high-side pressure P_{gc} is a key control variable in transcritical CO_2 operation and directly affects the system operating point, compressor work and thus the overall energy performance. At the same time, variables such as T_{amb}, RC_{MT}, $P_{o,MT}$, $T_{o,MT}$, P_{rec}, and S_{gc} provide physically meaningful information on heat-rejection conditions, refrigeration load, suction state, intermediate-pressure behavior, and gas-cooler performance, making them highly relevant candidates for data-driven prediction of the high-pressure level. At this point, it is worth noting that high values of $P_{o,MT}$ above 30 bar are typically associated with standby or transient, non-representative operating conditions and for this reason, such points were treated separately during the data-cleaning stage of the modeling workflow.

(a)



(b)

Figure. 1. *Overview of the transcritical CO2 system NEVA and (b) schematic of the main components*

2.2. ANN approach

The modeling approach followed in this study is based on supervised regression using ANNs trained with on-field measurements from the transcritical CO_2 refrigeration system described in the previous section. The objective of the ANN is the prediction of the gas-cooler/high-side pressure, P_{gc}, using plant measurements and operating-state variables. For this purpose, the recorded signals were first organized in MATLAB into a unified time-series dataset, used afterwards for model development, training, and evaluation. The overall approach treats the ANN as a surrogate model of the system behavior, attempting to approximate the nonlinear relationship between measured operating conditions and the target control variable. Therefore, a dedicated single-output model was developed for P_{gc}, since this variable is both reliably measured and directly relevant to system control. The methodological focus was placed on constructing regression models that map thermodynamically meaningful input variables to the instantaneous HP pressure. The modeling workflow was structured progressively, beginning with a limited input space and gradually expanding both the number of input variables and the complexity of the ANN architecture. In parallel, the effects of sample size, data ordering, training algorithm, and lagged information were systematically examined, as is discussed in section 3. The selected ANN family was a multi-layer feedforward neural network (FNN). Feedforward networks were chosen because they provide a flexible nonlinear regression framework and are well suited for surrogate modeling of complex input-output relationships [15]. A representative FNN structure is shown in **Figure 2**.

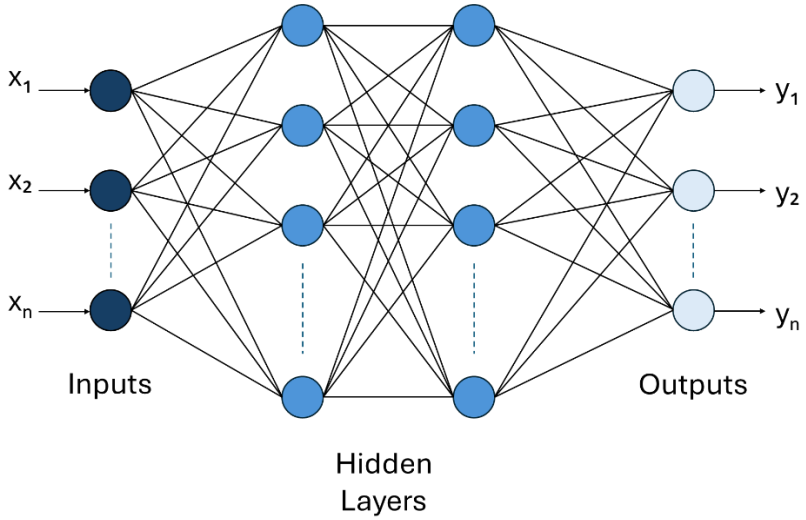


Figure. 2. *Architecture of a feedforward neural network with two hidden layers.*

The ANN models were configured as regression networks with one output neuron corresponding to the target variable. In the initial stage of development, architectures with one to three hidden layers were examined, using different combinations of neurons per layer. At a second stage, more complex architectures with additional hidden layers were also considered as part of the architecture-sensitivity study. For the hidden layers, the nonlinear sigmoid transfer function tansig was used in order to capture nonlinear relationships among the operating variables [16]. The initial training strategy employed the Levenberg-Marquardt (LM) algorithm, selected because of its fast convergence and good performance in small- to medium-sized nonlinear least-squares problems [17]. During the subsequent methodological exploration, additional training algorithms were also considered, including resilient backpropagation (RP) and Bayesian Regularization (BR), to assess the influence of the optimizer selection on model robustness and computational cost.

Model training was formulated as a nonlinear least-squares regression problem, while predictive performance was assessed using the mean squared error (MSE) and the coefficient of determination, R^2. MSE was used as the optimization objective during training, whereas $R^2$was used as a complementary indicator of the agreement between measured and predicted values.

The MSE is defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad (1)$$

where $y_i$is the measured value of the target variable, $\hat{y}_i$is the ANN prediction, and N is the number of samples. In addition to MSE, model fit was assessed through the coefficient of determination, R^2, given by

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \qquad (2)$$

where $\bar{y}$ is the mean of the measured target values.

With respect to the ANN training, the LM algorithm updates the parameter vector w according to:

$$\mathrm{w}_{k+1} = \mathrm{w}_k - (\mathrm{J}^T\mathrm{J}+\mu\mathrm{I})^{-1}\mathrm{J}^T\mathrm{e} \qquad (3)$$

where J is the Jacobian matrix of network errors with respect to the parameters, e is the error vector, $\mu$is the damping factor, and I is the identity matrix. In this form, LM combines the fast local convergence of Gauss-Newton methods with the robustness of gradient-descent-like updates, making it well suited to nonlinear least-squares training problems [17].

In BR algorithm, a modified objective function is implemented which balances data fitting and model complexity:

$$F = \beta E_D + \alpha E_W \qquad (4)$$

where the data-error term is equal to:

$$E_D = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad (5)$$

and the weight-penalty term is:

$$E_W = \sum_{j=1}^{n_w} w_j^2 \qquad (6)$$

with n_w to be the total number of network parameters. In contrast to standard error minimization, BR penalizes excessively large weights and therefore helps improve generalization and reduce overfitting [18].

The methodological development followed a staged procedure, as depicted in Figure 3. The first stage used a reduced set of inputs in order to establish a baseline regression structure for P_{gc}. The second stage addressed the effect of data arrangement by applying random shuffling to the available samples before partitioning into training, validation, and testing subsets. This step was introduced to reduce the mismatch that can arise when chronologically continuous blocks of data correspond to different operating regimes, thereby improving the representativeness of all subsets [19]. The third stage focused on input-variable selection; starting from a compact set of operating descriptors, additional variables were progressively introduced on the basis of their thermodynamic relevance and their physical connection to the high-side pressure. The explored sets of inputs included combinations of ambient temperature T_{amb}, medium-temperature capacity ratio RC_{MT}, suction pressure $P_{o,MT}$, suction saturation temperature $T_{o,MT}$, receiver pressure P_{rec}, and gas-cooler outlet temperature S_{gc}. In this way, the methodology aimed not only to construct an accurate prediction model, but also to clarify which measured variables contribute most effectively to the representation of the system state. A fourth methodological stage included the investigation of the dataset size. Different sample counts were considered during model development, beginning with smaller subsets and extending to larger datasets drawn from the full monitoring period. This was done to examine the effect of data availability on model training and generalization under varying operating conditions. In addition, a lagged version of the target variable, $P_{gc}(t-1)$, was examined as an extra input to incorporate information about the dynamic inertia of the refrigeration system. This choice reflects the fact that thermodynamic variables in refrigeration cycles are temporally correlated, so the current pressure state is partly influenced by the immediately preceding operating state. In that way, the study captures short-term temporal dependence through explicit lagged inputs within the feedforward framework.

Overall, the adopted methodology combines a physically informed choice of inputs with a structured ANN-development procedure. This structure was designed to produce a control-relevant surrogate model for high-side pressure prediction while keeping the modeling sufficiently transparent and computationally practical for engineering applications.

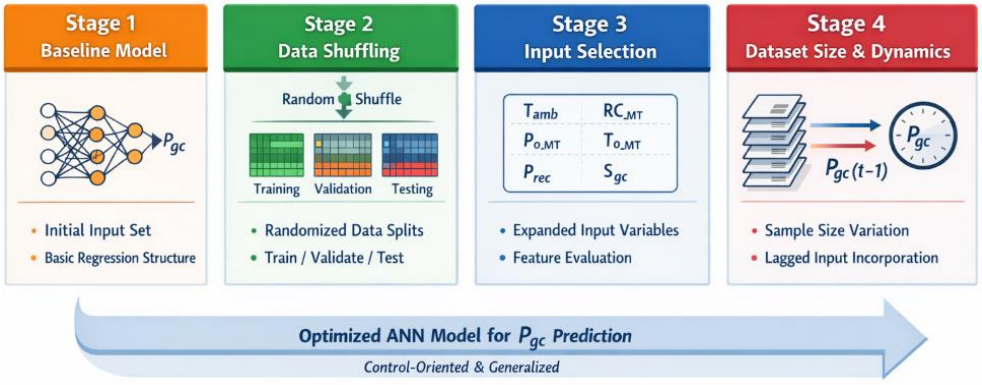


Figure. 3. *Schematic of the staged procedure for the development of the models*

3. Results and discussion

3.1. Baseline model and data preparation influence

The results of the baseline model show that the prediction accuracy of the ANN depends strongly on both sample size and the dataset quality. Using a reduced two-input set, namely T_{amb} and RC_{MT}, the baseline model exhibited a clear improvement when the number of samples increased from 5,000 to 10,000, with the R^2increasing from 0.6433 to 0.9323. However, when the same chronological-order sampling strategy was extended to larger sample rates, performance deteriorated substantially, with R^2 decreasing to 0.8376 at 20,000 samples, 0.5643 at 40,000, 0.3921 at 60,000 and 0.0164 at 80,000. This behavior indicates that, although a small increase in sample count initially improves regression quality, a chronologically expanded dataset also introduces broader operating variability and more transient regimes, making the mapping from the selected inputs to P_{gc}increasingly difficult for a simple ANN. The regression plots, listed in **Figure 4**, confirm this trend. At 5,000 samples, the predicted values show broad scatter and a noticeable deviation of the fitted line from the ideal $x = y$ line, whereas at 10,000 samples the point-cloud becomes appreciably tighter and the regression line approaches the ideal trend.

A more robust behavior emerged when random shuffling was applied before splitting the dataset into training, validation, and testing subsets, comprising 70%, 15%, and 15% of the samples, respectively. Under this strategy, the same two-input ANN retained significantly higher predictive quality across the full range of dataset sizes. The shuffled results yielded R^2 values of 0.8117 at 5,000 samples, 0.9499 at 10,000, 0.9303 at 20,000, 0.8520 at 40,000, 0.7589 at 60,000, 0.7232 at 80,000, and 0.6122 at 100,000 samples. Compared with the non-shuffled case, shuffling clearly reduced the train-test mismatch caused by contiguous operating periods belonging to different regimes, thereby making all subsets more representative of the full operating envelope. At the same time, the results also show that shuffling alone is not sufficient to eliminate the drop in performance at larger sample counts. Once the dataset spans a broader range of system conditions, the pair T_{amb} - RC_{MT} becomes insufficient to describe the full thermodynamic state governing P_{gc}.

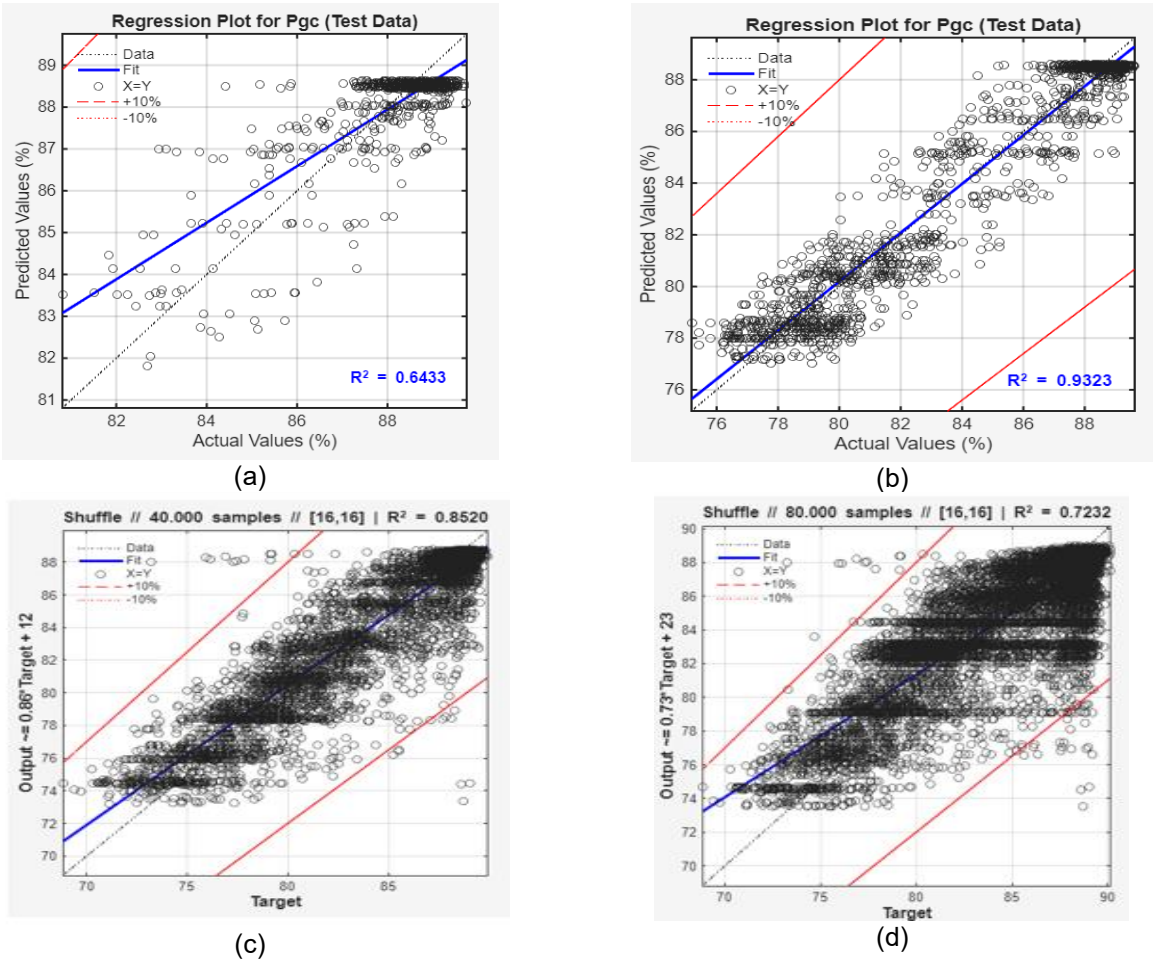


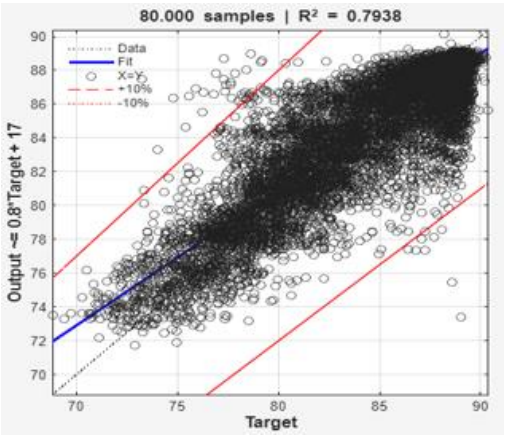
Figure. 4. *Regression plots illustrating the baseline two-input $\{T_{amb}, RC_{MT}\}$ ANN behavior and the effect of shuffling on P_{gc} prediction: (a) 5,000 samples; (b) 10,000 samples; (c) shuffled 40,000 samples; (d) shuffled 80,000 samples*

The article title appears centred at the top of the first page. To format the title authors should use the “Title” style from the formatting menu. Only the first word and proper nouns for the title should be capitalized. The use of acronyms and abbreviations in the title should be avoided, unless they are widely understood, or they are accompanied by the expanded expression.

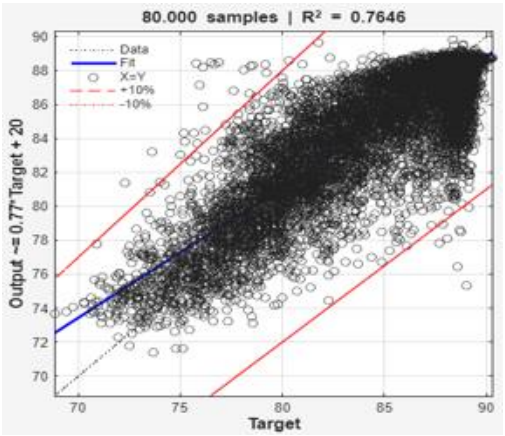
3.2. Effect of input selection enrichment

After establishing the shuffled two-input model as the reference case, the next step was to examine whether the addition of thermodynamically meaningful variables could yield higher accuracy for larger and more heterogeneous datasets. The first comparison considered the incremental addition of single variables to the baseline $\{T_{amb}, RC_{MT}\}$ set. At 40,000 samples, the addition of $P_{o,MT}$, $T_{o,MT}$, P_{rec}, and S_{gc} produced R^2values of 0.8600, 0.8609, 0.8587, and 0.8767, respectively. At 80,000 samples, the corresponding values were 0.7562, 0.7532, 0.7498, and 0.7938. Among the single-variable additions, the gas-cooler outlet temperature S_{gc}provided the largest improvement at both dataset sizes, suggesting that high-side heat-rejection information contributes strongly to the representation of the pressure state in transcritical operation.

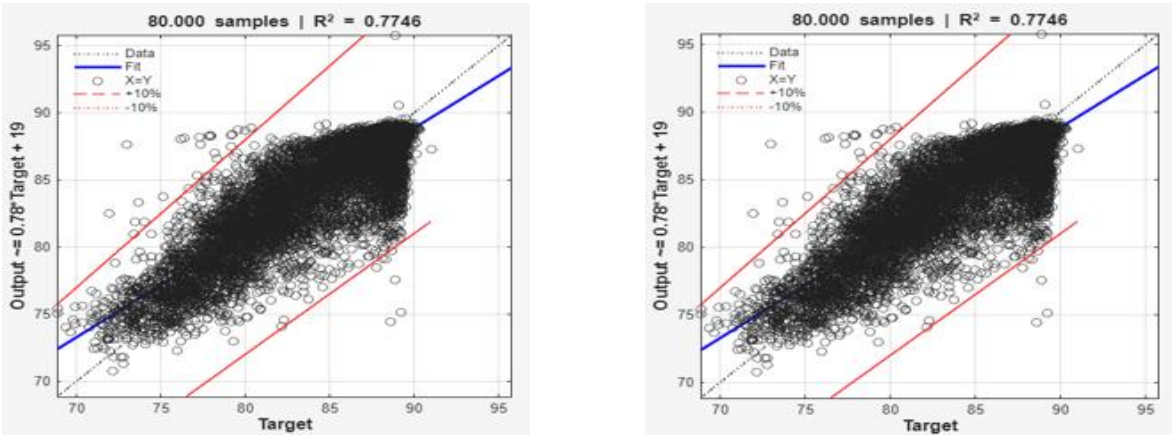
These comparisons show that the model benefits from variables that better reflect the physical mechanisms controlling the high-side pressure. The modest spread among the $P_{o,MT}$, $T_{o,MT}$, and P_{rec} cases indicates that suction-side and receiver-side variables all contain useful information about the operating state, but the consistently stronger performance of S_{gc} points to the central role of gas-cooler conditions in determining P_{gc}, an observation consistent with the thermodynamic behavior of the transcritical CO_2 cycle, in which the high-pressure level is tightly coupled to ambient-driven heat rejection and the state of the refrigerant at the gas-cooler outlet. The next step of the study was the cumulative input enrichment. When the two-input baseline was expanded to $\{T_{amb}, RC_{MT}, P_{o,MT}, T_{o,MT}\}$, the regression quality increased to $R^2 = 0.8619$ at 40,000 samples and $R^2 = 0.7646$ at 80,000. Adding receiver pressure to form $\{T_{amb}, RC_{MT}, P_{o,MT}, T_{o,MT}, P_{rec}\}$ yielded a small but consistent improvement, raising R^2 to 0.8631 and 0.7746 for 40,000 and 80,000 samples, respectively. The richest six-input set, $\{T_{amb}, RC_{MT}, P_{o,MT}, T_{o,MT}, P_{rec}, S_{gc}\}$, delivered the best performance in this stage of the investigation, with $R^2 = 0.8786$ at 40,000 samples and $R^2 = 0.8043$ at 80,000 samples. A graphic overview of these findings for the 80,000 samples is shown in **Figure 5**. These results indicate that progressively enriching the input space makes the ANN substantially more robust to the increased heterogeneity of the larger datasets. As shown from this analysis, once the network was supplied with variables carrying information on suction conditions, intermediate-pressure behavior, and gas-cooler outlet state, the mapping to P_{gc} became significantly more accurate. This result supports the use of physically informed feature selection in control-oriented surrogate modeling of refrigeration systems, where predictive accuracy depends not only on algorithmic sophistication but also on choosing variables that span the relevant thermodynamic degrees of freedom.



(a)



(b)

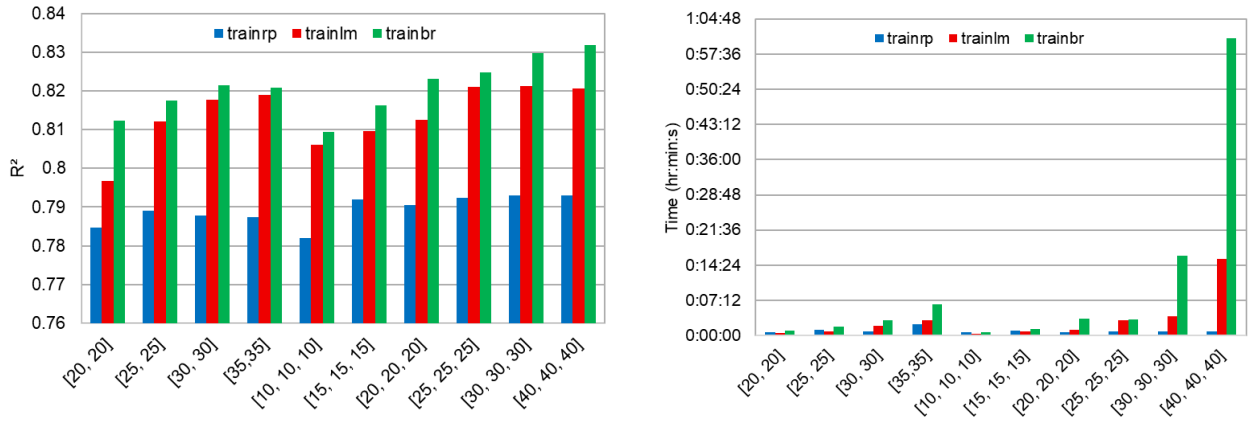


(c) (d)

Figure. 5. *Effect of input-variable enrichment on ANN regression quality for P_{gc}prediction for 80,000 samples: (a) $\{T_{amb}, RC_{MT}, S_{gc}\}$; (b) $\{T_{amb}, RC_{MT}, P_{o,MT}, T_{o,MT}\}$; (c) $\{T_{amb}, RC_{MT}, P_{o,MT}, T_{o,MT}, P_{rec}\}$; (d) $\{T_{amb}, RC_{MT}, P_{o,MT}, T_{o,MT}, P_{rec}, S_{gc}\}$*

3.3. Training algorithm, network architecture and final model selection

After the six-input configuration was identified as the most accurate set of inputs, the focus shifted to the combined effect of training algorithm and network architecture. For the 80,000 sample six-input case, the three training algorithms were cross-evaluated, displaying clearly different performance levels. The RP algorithm produced the lowest results overall, with R^2values ranging from 0.7820 to 0.7931 and MSE values around 4.38 to 4.14 across the tested architectures. The LM algorithm performed better, reaching $R^2 = 0.8212$ and MSE $= 3.593$ for a [30,30,30] architecture, while BR delivered the best results among the three training algorithms, rising to $R^2 = 0.8318$ and MSE $= 3.379$ for a [40,40,40] architecture. At the same time, the results reveal a clear diminishing-returns trend with increasing ANN complexity. Under BR algorithm, the progression from [20,20,20] to [25,25,25], [30,30,30] and [40,40,40] improved R^2 from 0.8231 to 0.8248, 0.8298, and 0.8318, respectively. However, the gain from [30,30,30] to [40,40,40] was only 0.0020 in terms of R^2, while the training time increased from 00:16:24 to 01:00:45. In other words, an improvement of only about 0.24% in R^2required approximately a fourfold increase in training time. The same tradeoff is visible in the comparison between LM and BR, as shown in **Figure 6**, the latter provides better generalization, but at a substantially higher computational cost, especially as network depth and width increase.



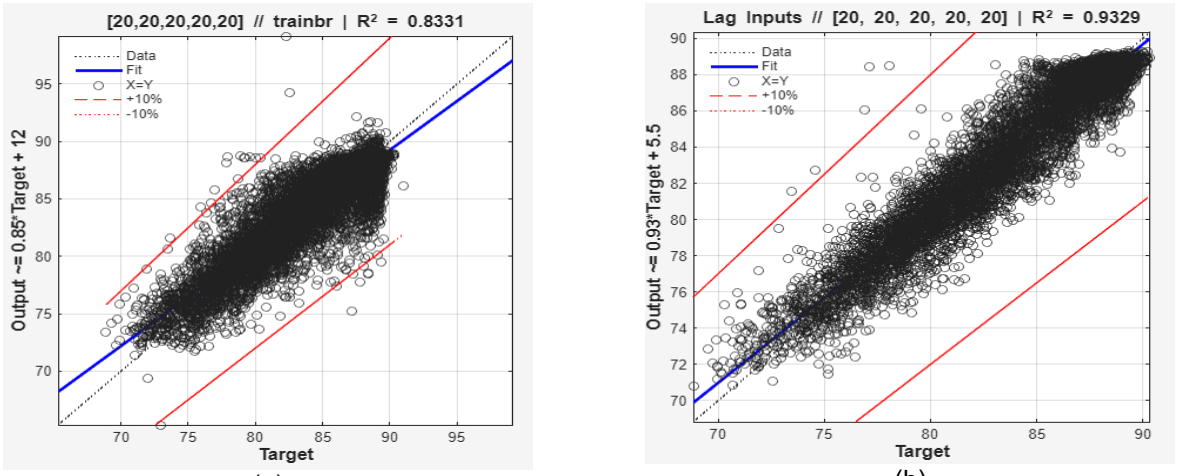
(a) (b)

Figure. 6. *Comparative effect of training algorithm and ANN architecture for the six-input, 80,000-sample model: (a) R^2as a function of architecture for RP (trainrp), LM (trainlm), and BR (trainbr) algorithms; (b) corresponding training time.*

The extension to even deeper architectures confirmed that performance can still improve beyond the three-layer networks, but only up to a point. Among the more complex models, LM algorithm yielded $R^2 = 0.8117$, 0.8230, and 0.8218 for [10,10,10,10], [20,20,20,20], and [30,30,30,30], respectively. The corresponding BR networks performed better, with $R^2 = 0.8237$ for [20,20,20,20], 0.8281 for [25,25,25,25], 0.8315 for [30,30,30,30], and 0.8331 for the five-layer [20,20,20,20,20] architecture. Increasing the depth further to six hidden layers, [20,20,20,20,20,20], did not improve performance; instead, R^2 decreased slightly to 0.8304, while the training time rose sharply to 02:17:00. This confirms that the optimum for the present problem, and the given data sample, lies at a balanced architecture that captures sufficient nonlinearity without incurring

unnecessary computational burden. As a result, the five-hidden-layer BR network with [20,20,20,20,20] architecture was selected as the final non-lagged model, as shown in **Figure 7(a)**. Its choice reflects the priority given to predictive accuracy and generalization on the independent test set, while accepting a higher but one-time training cost.

A final methodological extension examined the addition of lagged pressure information, $P_{gc}(t-1)$, as an extra ANN input. This step was motivated by the temporal correlation and inertia of thermodynamic variables in refrigeration systems [20]. The lag-augmented model yielded a visibly tighter regression cloud and a better fit to the ideal prediction line than the corresponding non-lagged final ANN, showing that short-term dynamic information can further improve predictive accuracy when incorporated explicitly into the feedforward framework, as shown in **Figure 7(b)**. The final outcome is eventually a structured demonstration that accurate P_{gc} prediction in transcritical CO_2 refrigeration requires the combined consideration of data preparation, physically meaningful inputs, algorithm choice, model complexity, and, where appropriate, short-term temporal information.



(a) (b)

Figure. 7. Regression plots of the final selected ANN models for P_{gc}prediction: (a) non-lagged final model, trainbr, $[20, 20, 20, 20, 20]$, with inputs $\{T_{amb}, RC_{MT}, P_{o,MT}, T_{o,MT}, P_{rec}, S_{gc}\}$; (b) lag-augmented version including $P_{gc}(t-1)$.

4. Conclusions

Summarizing, this study examined the use of feedforward artificial neural networks as surrogate models for predicting the gas-cooler/high-side pressure, P_{gc}, in a transcritical CO_2 refrigeration system using real operating data from a supermarket installation. The analysis focused on the effects of data preparation, input selection, dataset size, network architecture, training algorithm, and lagged information on predictive accuracy and model generalization. The main conclusions of the study are listed below:

- Feedforward ANNs were shown to provide an effective framework for modeling the nonlinear relationship between measured operating variables and P_{gc}.
- The baseline results demonstrated that increasing the number of samples does not necessarily improve prediction quality when the input space is insufficiently informative. Random shuffling prior to dataset partitioning significantly improved the representativeness of the training, validation, and test subsets, leading to more stable and meaningful model behavior across different sample sizes.
- Input-variable selection proved to be one of the dominant factors affecting predictive performance. Progressively enriching the input set with thermodynamically relevant variables substantially improved the prediction of P_{gc}, confirming the importance of physically informed feature selection in control-oriented surrogate modeling.
- Among the examined training strategies, Bayesian Regularization provided the best overall predictive performance, although at higher computational cost than Levenberg-Marquardt and resilient backpropagation. At the same time, increasing network complexity led to diminishing returns, indicating that model selection must balance accuracy gains against training effort.
- The best non-lagged model was obtained with a five-hidden-layer ANN trained using Bayesian Regularization, while the inclusion of the lagged variable $P_{gc}(t-1)$further improved predictive quality by capturing short-term dynamic effects of the refrigeration system with a final R^2 of 93.29% for a five layer 20 neurons/layer ANN.

- Overall, the results support the use of ANN-based surrogate models as practical tools for prediction of control-critical variables in transcritical CO_2 refrigeration systems, with clear potential for application in monitoring, supervisory control, and energy optimization.

Nomenclature

E_D	data-error term in Bayesian Regularization, -
E_W	weight-penalty term in Bayesian Regularization, -
$\mathbf{e}$	error vector, -
F	regularized objective function, -
I	identity matrix, -
J	Jacobian matrix of network errors, -
MSE	mean squared error, -
N	number of samples, -
n_w	total number of trainable network parameters, -
P	pressure, bar
R^2	coefficient of determination, -
RC	rack capacity ratio, -
S_{gc}	gas-cooler outlet temperature, °C
T	Temperature, °C
$V_{hp,OD}$	high-pressure valve opening degree, %
y_i	measured value of the target variable, bar
$\hat{y}_i$	predicted value of the target variable, bar
$\bar{y}$	mean measured value of the target variable, bar

Greek symbols

α	regularization parameter, -
β	regularization parameter, -
μ	damping factor in the Levenberg-Marquardt algorithm,-

Subscripts and superscripts

amb	ambient
d	discharge
gc	gas cooler
HP	high pressure
LT	low temperature
MT	medium temperature
o	saturation state corresponding to suction pressure
OD	opening degree
rec	receiver / flash tank
s	suction

References

[1] Tassou SA, De-Lille G, Ge YT. Food transport refrigeration – Approaches to reduce energy consumption and environmental impacts of road transport. Applied Thermal Engineering 2009;29:1467–77. https://doi.org/10.1016/j.applthermaleng.2008.06.027.

[2] Regulation (EU) 2024/573 of the European Parliament and of the Council of 7 February 2024 on fluorinated greenhouse gases, amending Directive (EU) 2019/1937 and repealing Regulation (EU) No 517/2014 (Text with EEA relevance). 2024.

[3] Mota-Babiloni A, Makhnatch P. Predictions of European refrigerants place on the market following F-gas regulation restrictions. International Journal of Refrigeration 2021;127:101–10. https://doi.org/10.1016/j.ijrefrig.2021.03.005.

[4] Cavallini A, Zilio C. Carbon dioxide as a natural refrigerant | International Journal of Low-Carbon Technologies | Oxford Academic. International Journal of Low-Carbon Technologies 2007;2:225–49.
[5] Ge YT, Tassou SA. Thermodynamic analysis of transcritical CO2 booster refrigeration systems in supermarket. Energy Conversion and Management 2011;52:1868–75. https://doi.org/10.1016/j.enconman.2010.11.015.
[6] Peñarrocha I, Llopis R, Tárrega L, et al. A new approach to optimize the energy efficiency of CO2 transcritical refrigeration plants. Applied Thermal Engineering 2014;67:137–46. https://doi.org/10.1016/j.applthermaleng.2014.03.004.
[7] Sarkar J, Agrawal N. Performance optimization of transcritical CO2 cycle with parallel compression economization. International Journal of Thermal Sciences 2010;49:838–43. https://doi.org/10.1016/j.ijthermalsci.2009.12.001.
[8] Elbel S, Lawrence N. Review of recent developments in advanced ejector technology. International Journal of Refrigeration 2016;62:1–18. https://doi.org/10.1016/j.ijrefrig.2015.10.031.
[9] Hafner A, Försterling S, Banasiak K. Multi-ejector concept for R-744 supermarket refrigeration. International Journal of Refrigeration 2014;43:1–13. https://doi.org/10.1016/j.ijrefrig.2013.10.015.
[10] Tashtoush B, Sahli H, Elakhdar M, et al. A new CO2 refrigeration system with two-phase ejector and parallel compression for supermarkets. Heliyon 2024;10:e27519. https://doi.org/10.1016/j.heliyon.2024.e27519.
[11] Al-Hammadi AA, Franzoi RE, Ibrahim OE, et al. Surrogate modeling for mixed refrigerant streams in the refrigeration cycle of an LNG plant. Computer Aided Chemical Engineering 2022;49:1795–800. https://doi.org/10.1016/B978-0-323-85159-6.50299-2.
[12] Ahmed R, Mahadzir S, Mota-Babiloni A, et al. 4E analysis of a two-stage refrigeration system through surrogate models based on response surface methods and hybrid grey wolf optimizer. PLOS ONE 2023;18:e0272160. https://doi.org/10.1371/journal.pone.0272160.
[13] Opalic SM, Goodwin M, Jiao L, et al. ANN modelling of CO2 refrigerant cooling system COP in a smart warehouse. Journal of Cleaner Production 2020;260:120887. https://doi.org/10.1016/j.jclepro.2020.120887.
[14] Wang J, Belusko M, Semsarilar H, et al. An optimisation study on a real-world transcritical CO2 heat pump system with a flash gas bypass. Energy Conversion and Management 2022;251:114995. https://doi.org/10.1016/j.enconman.2021.114995.
[15] Bebis G, Georgiopoulos M. Feed-forward neural networks. IEEE Potentials 1994;13:27–31. https://doi.org/10.1109/45.329294.
[16] Reyes-Téllez ED, Parrales A, Ramírez-Ramos GE, et al. Analysis of transfer functions and normalizations in an ANN model that predicts the transport of energy in a parabolic trough solar collector. Desalination and Water Treatment 2020;200:23–41. https://doi.org/10.5004/dwt.2020.26063.
[17] Moré JJ. The Levenberg-Marquardt algorithm: Implementation and theory | Springer Nature Link. Lecture Notes in Mathematics, vol. 630, 2006.
[18] Santos CFGD, Papa JP. Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks. ACM Comput Surv 2022;54:213:1-213:25. https://doi.org/10.1145/3510413.
[19] Bishop CM, Nasrabadi NM. Pattern recognition and machine learning. vol. 4. Springer; 2006.
[20] Bush J, Aute V, Radermacher R. Transient simulation of carbon dioxide booster refrigeration system with mechanical subcooler in demand response operation. Science and Technology for the Built Environment 2018;24:687–99.