

Article

A Symmetry-Based Hybrid Model of Computational Fluid Dynamics and Machine Learning for Cold Storage Temperature Management

Yang Liu ¹ , Lanting Guo ², Xiaoyu Hu ³ and Mengjie Zhou ^{4,*}

¹ Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609, USA; harryliu@ieee.org

² The Department of Food Science and Human Nutrition, University of Illinois Urbana-Champaign, Champaign, IL 61801, USA; lanting.guo@ieee.org

³ Department of Chemical Engineering and Materials Science, Stevens Institute of Technology, Hoboken, NJ 07030, USA; xiaoyh5@uci.edu

⁴ Department of Computer Science, University of Bristol, Bristol BS8 1QU, UK

* Correspondence: ix18497@bristol.ac.uk

Abstract: Cold chain temperature management is crucial for preserving product quality and safety across various industries. While Computational Fluid Dynamics (CFD) provides detailed insights into thermal analysis and fluid dynamics, its computational intensity limits practical applications. This study presents a novel hybrid approach combining CFD and machine learning to enhance both computational efficiency and prediction accuracy in cold storage temperature management. A validated 3D CFD model was developed to analyze temperature distribution and airflow patterns in a refrigerated container with multiple storage boxes. Taking advantage of the cold room's symmetrical design along its longitudinal axis significantly reduced computational requirements while maintaining model accuracy. Over 200 cases were simulated by varying key process parameters to generate training data for machine learning models. Random Forest (RF) and Neural Network (NN) models were developed and compared, with RF demonstrating consistently superior performance across all storage locations. Feature importance analysis revealed cold air temperature as the dominant control variable, while SHAP analysis identified optimal operational ranges for air velocity and heat transfer coefficients that balance product quality with energy efficiency. This research work also revealed distinct patterns in the influence of process parameters, with cold air and ambient temperatures showing hierarchical impacts on system performance. The hybrid methodology successfully addresses the computational limitations of traditional CFD approaches while maintaining high prediction accuracy, offering a practical solution for sustainable temperature management in cold storage applications. Finally, this research provides valuable insights for optimizing cold chain operations and demonstrates the potential of hybrid modeling approaches in thermal management systems.

Keywords: cold chain management; computational fluid dynamics; hybrid machine learning; symmetrical design; insulation material



Academic Editor: Toshio Tagawa

Received: 3 March 2025

Revised: 27 March 2025

Accepted: 30 March 2025

Published: 1 April 2025

Citation: Liu, Y.; Guo, L.; Hu, X.; Zhou, M. A Symmetry-Based Hybrid Model of Computational Fluid Dynamics and Machine Learning for Cold Storage Temperature Management. *Symmetry* **2025**, *17*, 539. <https://doi.org/10.3390/sym17040539>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cold chain temperature management is crucial for maintaining the quality, safety, and shelf life of temperature-sensitive products across various industries, including pharmaceuticals, food, and biochemicals. The critical role of cold chain management is highlighted

by [1], who found that improper temperature control during extended transportation results in about 40% value loss for fruits and vegetables.

Temperature management in cold storage rooms is crucial for maintaining product quality and reducing energy consumption. There are several critical weak points in cold chains: precooling operations, ground transportation operations, retail display, and consumer handling [2]. Studies show temperature variations can be as high as 10 °C during transport, with vulnerability during loading and unloading operations [3]. Notably, consumer handling represents the weakest link, with domestic refrigerator temperatures often exceeding recommended levels by 2–3 °C [4].

As noticed from the above presented critical weak points, proper temperature management could greatly mitigate the risk of exceeding the required temperature range for transported goods. Proper temperature management largely depends on reliable methodologies to accurately measure and predict cold storage room temperature. CFD has proven valuable for optimizing cooling equipment placement. CFD is a method that uses computer simulations to predict and analyze fluid flow and heat transfer by solving mathematical equations across a discretized domain, enabling engineers to understand complex fluid behavior in various systems [5]. In their work, Ref. [6] validated the accuracy of 3D CFD modeling for predicting temperature distributions in refrigerated containers through experimental studies. Meanwhile, Ref. [7] used CFD models to study airflow in cold storage apple crates, finding that wider spacing between stacks improves cooling efficiency by 25.2% compared to conventional arrangements, while maintaining more uniform temperatures throughout. A 3D CFD model for predicting temperature distribution in cold storage rooms using ultrasonic humidifiers was validated by Ref. [6], achieving accuracy with limited error for air velocity and temperature measurements in experiment validation. Several research works have highlighted the importance of proper airflow design in maintaining uniform temperature distribution and advanced phase change material to improve energy efficiency [8,9]. In Ref. [10], the authors utilized CFD simulations to analyze the impact of door openings on temperature fluctuations in cold storage rooms. Their research revealed that traditional CFD models, while accurate, required significant computational resources and time for complex scenarios.

CFD modeling of cold storage rooms is constrained by high computational costs, especially for large facilities and complex operations like door openings. Model validation challenges arise from limited experimental data, while accuracy depends heavily on boundary condition assumptions and heat load estimations [11]. The models struggle with rapid transient events and complex turbulent flows, and the detailed geometric requirements for products and equipment make mesh generation demanding, often requiring simplifications that can affect prediction accuracy of temperature and airflow patterns.

Recent years have seen an increase in the application of machine learning techniques to cold chain management. The authors of [12] developed a machine learning model combining optimal control and k-means clustering to manage cold storage temperatures, identifying optimal storage conditions between -0.87 °C and 3.31 °C for preserving perishable foods. Their model significantly reduced computational time compared to traditional CFD approaches. By using a machine learning model, [13] predicted cold chain temperatures in real-time, although accuracy depends on the training data quality. Experimental data provide 20–40% better results than synthetic data, which can lead to overfitting. The study recommends optimizing sensor placement and improving thermal models rather than increasing data volume. The authors of [14] developed a machine learning pipeline to forecast energy usage, temperature, and humidity in food cold storage facilities one week in advance. The research reveals that Random Forest models achieved the best electricity consumption predictions when using real-world data from two facilities, and [15] developed a

deep learning model combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture with transfer learning to predict egg freshness during cold storage, using nondestructive weight loss measurements from various temperature conditions. The model achieved superior accuracy compared to traditional methods like Random Forest, enabling effective real-time quality monitoring in cold chain environments. However, as noted by [2], challenges remain in standardizing these approaches across different storage facility configurations and product types.

While the advancement of machine learning greatly improves temperature management for cold storage, machine learning in cold storage temperature studies is limited by the need for extensive training data and struggles with generalization across different facility layouts. Model accuracy depends heavily on data quality, while real-time implementation requires robust sensor networks and processing capabilities. The “black box” nature of ML models makes physical validation challenging, and they may fail to handle unexpected scenarios not present in training data.

2. Methodology

In the present study, it is intended to create an integration of CFD and machine learning to create a powerful tool that enhances both computational efficiency and accuracy in the prediction of temperature for cold storage rooms. This hybrid approach reduces computation time by using machine learning models trained on previous CFD simulations, enabling rapid prediction and optimization. While machine learning models require less computational power than full CFD simulations, they can improve prediction accuracy to handle the complex thermal and fluid dynamics systems of cold storage more efficiently.

In the present study, a full-scale 3D CFD model is firstly developed by referring to the recent research of [16], in which the authors systematically studied temperature distribution inside a refrigerated container. The baseline results are compared and validated with the experimental measurements. Afterwards, more than 200 cases with varying process conditions were run to be used by developing machine learning models. The machine learning models predict the temperature distribution of the cold storage boxes in the container to provide comprehensive guidance for temperature management by considering different process conditions in food and vaccine cold supply chain management.

2.1. Baseline Validation

Figure 1 shows the schematic view of the cold container used in the baseline study of the present CFD model. The container is 11.65 m long, 2.28 m in width, and 2.56 m in height. Cold air is supplied from a refrigeration unit and distributed throughout the whole container and returned from the grille on the side. There are a total of 10 arrays of boxes that contain perishable frozen food or vaccines that are placed along the length direction, which are numbered from 1 to 10 from the cold air inlet location to the door (whose dimensions are 0.5 m × 0.5 m × 0.54 m); see details in Figure 1a. Thus, by referring to Figure 1b, one can read that the computational domain spans 0–11.65 m in the x direction for domain length, 0–2.28 m in the y direction for domain width, and 0–2.56 m in the z direction for domain height. The cold temperature management of the container is largely dependent on how well the container is insulated, and Figure 1b shows the insulation material layers on the outside of the container. Both the top cover and side wall are insulated with polyurethane foam (PUR), whose thicknesses are 13.6 cm and 6.6 cm, respectively.

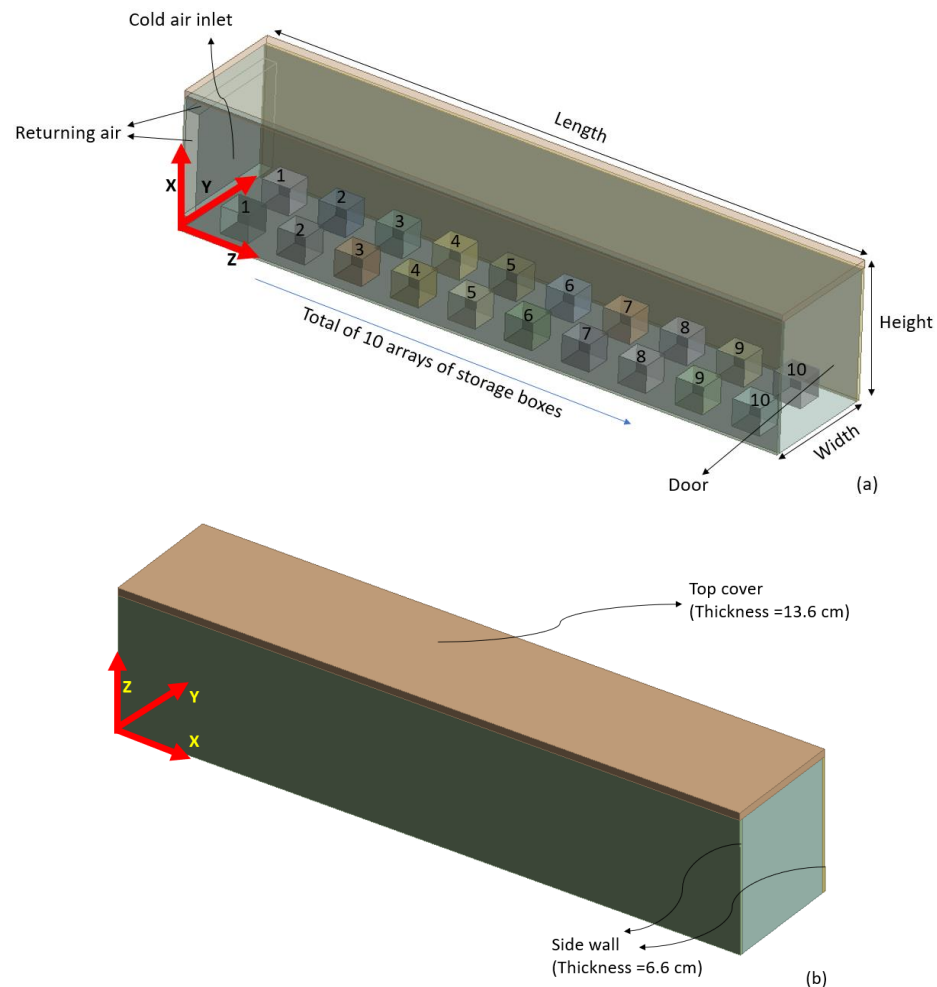


Figure 1. Schematic view of symmetrically designed cold storage room for CFD model. (a) Transparent view for visualizing the placement of frozen food or vaccine boxes and (b) insulation on the outside of the container, which highlights the thickness of top cover and side wall.

The cold room exhibits symmetry along its longitudinal axis, which allows the CFD model to leverage this symmetrical nature in our modeling approach. By taking advantage of this symmetry, we can significantly reduce computational requirements while maintaining model accuracy. This symmetrical domain assumption is justified by the uniform arrangement of storage boxes and the balanced distribution of cold air from the refrigeration unit.

The developed CFD model performs a steady-state simulation. Turbulent flow is generated by the supplying cold air; the unit is modeled by a two-equation Reynold Average Navior–Stokes (RANS) model and Realizable k - ϵ model, as they provide better prediction of flow recirculation in confined spaces [17]. By referring to the experiment condition in [16], the boundary condition used in the baseline CFD model has “velocity inlet” at the cold air inlet, which is m/s at $-37.3\text{ }^{\circ}\text{C}$ and provides a total cold air volumetric flowrate of $0.38\text{ m}^3/\text{s}$. The thermal conductivity of the PUR insulation material is $0.0245\text{ W}/(\text{m}\cdot\text{K})$, both for the top cover and side walls, as shown in Figure 1b. As described above, the CFD model used “symmetry” boundary conditions so that the computational domain with half of the total width, shown in Figure 1b, is modeled.

The authors of Ref. [16] used thermocouples to measure multiple locations, including 6 data points for temperature distribution in the container and 1 data point for the returning air grille. In the above baseline study, a total of $\sim 68,3900$ tetrahedral meshes were used in the CFD model, which is based upon an independent mesh study to ensure this solution

(e.g., temperature at the thermocouple measurement locations does not change with a further increase of mesh number).

The comparison between the experiment and the present model is shown in Figure 2. The numerical model systematically slightly overpredicts temperatures by 2–3 °C compared to experimental measurements across all data points. However, the model accurately captures the overall temperature distribution from the experiment, showing good agreement between the present CFD model and the experimental measurements.

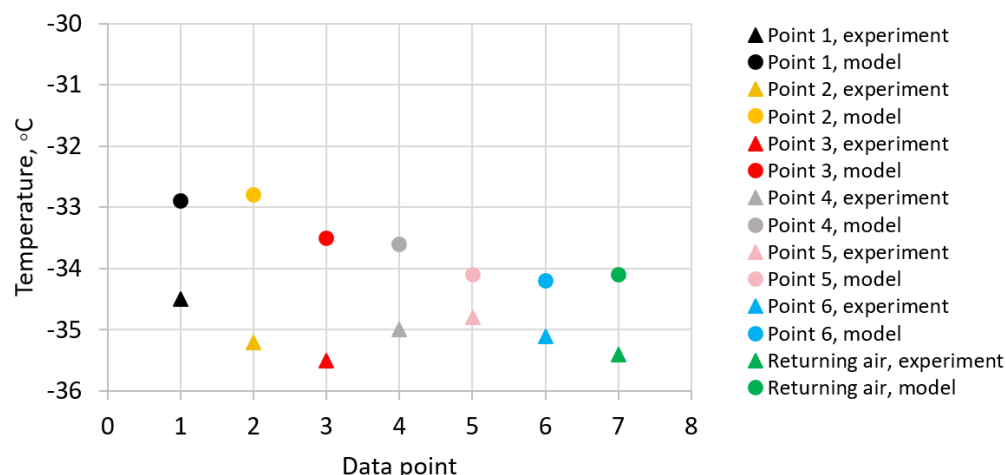


Figure 2. Comparison of results of temperature data points from [16]’s experiment and the present model.

2.2. CFD Model Results Analysis

From the above presented validated CFD model, further analysis is performed on the CFD model to reveal comprehensive temperature and velocity vector distribution within the cold container. In Figure 3, the temperature distribution visualization reveals a significant thermal gradient along the container’s length, with temperatures ranging from -37 °C near the cold air inlet (left side) to approximately -32 °C at the door (right side). The ambient temperature is 20 °C and seems to be well insulated by the high-performance PUR insulation material, thus the thermal stratification shows a clear transition from cold (blue) to warm (red) regions, with the cold zone effectively maintaining low temperatures around the first six boxes but showing reduced cooling efficiency for the boxes closer to the door. The thermal boundary layer development along the container walls and around the boxes is evident, with warmer temperatures accumulating in the upper right region of the container, suggesting potential optimization opportunities for airflow distribution to achieve more uniform cooling throughout the container length.

Figure 4 shows that CFD simulation results reveal a complex airflow pattern within the refrigerated container housing 10 uniformly spaced storage boxes. The velocity magnitude distribution, ranging from 0 to 0.94 m/s, demonstrates a predominantly longitudinal flow pattern with significant secondary recirculation zones. This recirculation zone is not observed in [16], which might be because [16] modeled a transient flow regime, while the present CFD model used a steady-state simulation. However, air recirculation commonly occurs as the high-velocity flow entrains the surrounding fluid through viscous forces. The entrainment creates pressure gradients within the confined space, forming counter-rotating vortices near the jet entry and secondary patterns downstream, which is observed in similar studies by previous researchers [18]. From the cross-sectional view in Figure 4a, the primary air stream exhibits higher velocities along the upper region of the container, while the presence of storage boxes creates distinct wake regions and vertical mixing patterns. The top view in Figure 4b illustrates the formation of symmetrical vortex structures around

each storage box, indicating potential areas of reduced heat transfer efficiency. The return flow configuration, where cold air enters and exits from the left side, generates a gradient in velocity magnitude along the container length, with notably stronger flow characteristics near the inlet region. From Figure 4, the flow architecture explains the thermal stratification and varying cooling effectiveness across the storage volume that presented in Figure 4, particularly in the downstream regions where velocity magnitudes decrease substantially due to flow resistance and energy dissipation through the box array.

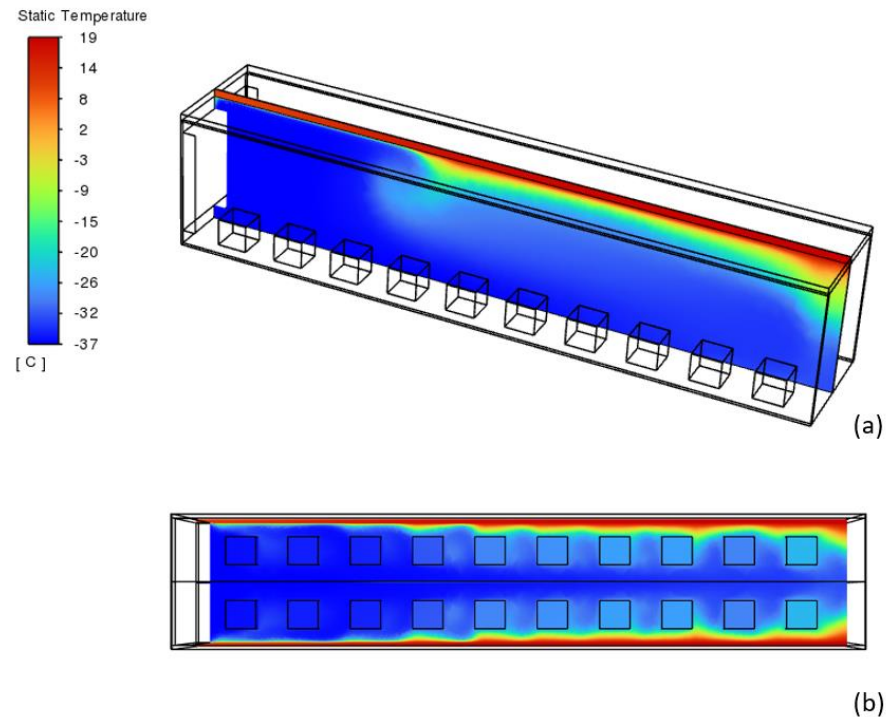


Figure 3. Temperature distribution of the symmetrically designed cold room with array-arranged storage boxes. (a) Cross-sectional and (b) top view.

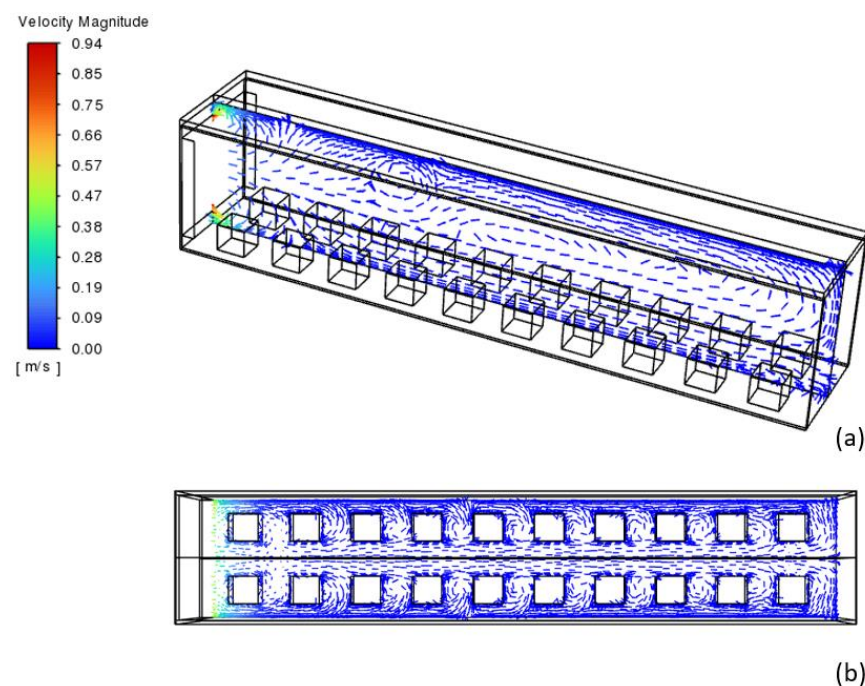


Figure 4. Velocity field distribution of the symmetrically designed cold room with array-arranged storage boxes. (a) Cross-sectional and (b) top view.

2.3. Generating Datasets for Machine Learning

As discussed above, while CFD simulation provides detailed spatiotemporal insights into the complex flow patterns within the cold container, the computational intensity and time requirements of these simulations limit their wide applications in operational improvement. By leveraging the high-fidelity CFD data as a training foundation for machine learning models, one can develop rapid prediction capabilities for critical parameters such as local velocity distributions, temperature gradients, and cooling effectiveness. Such machine learning models, once properly trained and validated against CFD results, could have the potential to enable even more effective thermal management. Furthermore, the integration of machine approaches with CFD-generated datasets opens new avenues for identifying nonobvious correlations between process parameters and performance metrics, potentially revealing optimal process control configurations that might not be immediately apparent through traditional parametric CFD studies alone. This hybrid CFD–machine learning methodology represents a promising direction for advancing the efficiency and intelligence of refrigerated container systems while maintaining the physical accuracy inherent in CFD simulations.

To develop a robust machine learning framework for predicting temperature distributions in refrigerated containers, a comprehensive training dataset was generated by systematically varying six key process variables within their operational ranges: air velocity, ambient temperature, cold air temperature, external wall heat transfer coefficient, food thermal conductivity, and PUR insulation thermal conductivity. The output parameters consist of the temperature for each of the ten storage boxes arranged sequentially within the container (see Figure 1). This parametric approach enables the capture of complex thermal interactions and nonlinear relationships between input variables and resulting temperature distributions, providing a rich dataset for training advanced machine learning algorithms. To minimize bias in generating cases, a Gaussian random generator is used to generate the 205 CFD cases when considering the above input variables. A Gaussian random generator is used in this study to produce case input conditions in the present machine learning study, as it helps avoid bias by naturally representing real-world phenomena, which ensures balanced data representation, and provides useful mathematical properties for weight initialization and data augmentation [19]. By producing values that follow a Gaussian random distribution, one creates realistic noise patterns that enhance model robustness while maintaining mathematical tractability. The statistical distribution of input parameters, characterized by their respective means and standard deviations, ensures comprehensive coverage of realistic operating conditions while maintaining physical consistency in the generated data, which are detailed in Table 1. The CFD model cases were run by varying the boundary conditions and material properties in the model setup in the parametric study. For example, different air velocities, ambient temperatures, and external wall heat transfer coefficients are input in the model setup for boundary conditions. The thermal conductivity values for both food and PUR insulation as material properties were varied. In total, 205 cases were run in parallel on a 48-core workstation.

It is important to ensure that the input variables were within reasonable ranges so that the modeled CFD cases represent meaningful physical conditions, especially for material properties. The external wall heat transfer coefficient ranges from low values (e.g., $0.019 \text{ W}/(\text{m}^2 \cdot \text{K})$), representing minimal natural convection, to high values (e.g., $86.4 \text{ W}/(\text{m}^2 \cdot \text{K})$), representing windy weather conditions.

Food thermal conductivity spans from $25.5 \text{ W}/(\text{m} \cdot \text{K})$ to approximately $100 \text{ W}/(\text{m} \cdot \text{K})$, as referenced in [20]. Cold air temperature mostly ranges from very strict frozen conditions, such as storing vaccines at -39.7°C [16], to general frozen food conditions at approximately 5°C [20].

Table 1. Value range limits for input variables for generating CFD model cases for developing machine learning models.

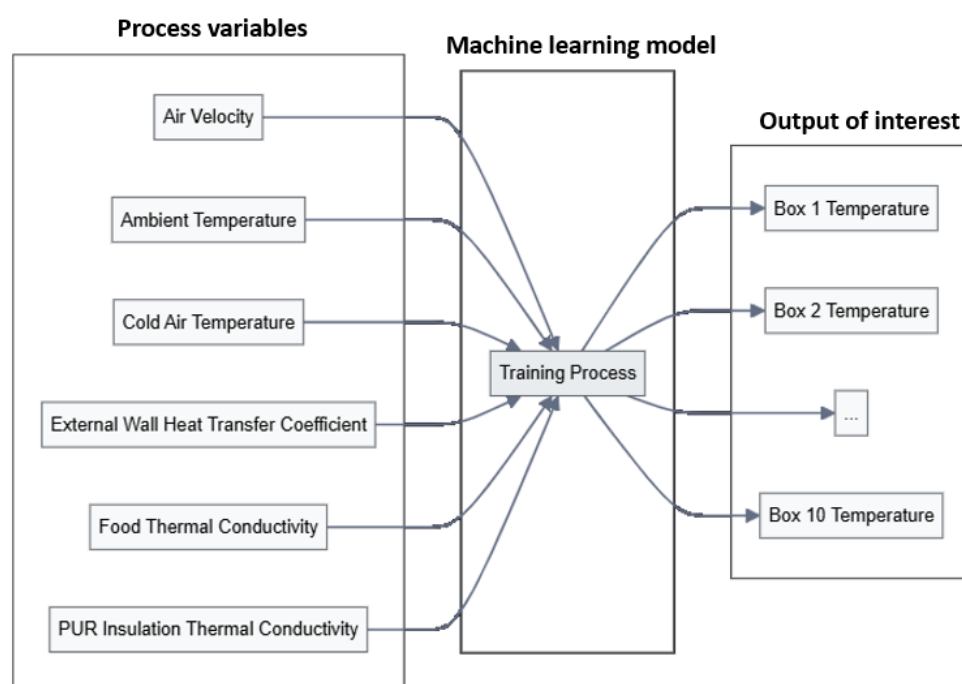
Input Variables	Min	Max	Mean	Std
Air velocity (m/s)	0.021	0.79	0.267	0.18
Ambient temperature (°C)	10.02	40.40	25.202	8.35
Cold air temperature (°C)	−39.702	−5.08	−23.201	10.36
External wall heat transfer coefficient (W/(m ² ·K))	0.019	86.4	15.789	7.84
Food thermal conductivity (W/(m·K))	25.528	99.85	53.676	27.01
PUR insulation thermal conductivity (W/(m·K))	0.003	0.19	0.025	0.02

PUR insulation material made of fiberglass initially has a thermal conductivity of approximately 0.003 W/(m·K) [21]. However, PUR insulation material degrades when exposed to environmental conditions (e.g., UV radiation from sunlight) and is assumed to degrade to a value only marginally better than particleboard (approximately 0.19 W/(m·K)) in the present study in the worst case [22].

The same CFD model meshes, which were used in the baseline study presented in Figures 2–4, were used for the above machine learning training data generation process. This is because the CFD cases only changed the process and material conditions instead of the structural geometry.

2.4. Machine Learning Model

The workflow diagram presented in Figure 5 illustrates the process of the machine learning model development to predict temperatures in multiple boxes (Box 1 through Box 10) based on six key process variables. The reason why the temperature of each box is output is because proper thermal management in a cold supply chain requires the maximum preservation of frozen food or vaccines, and thus the temperature uniformity of all boxes needs to be monitored and studied. These input variables are fed into a training process within the machine learning model, which then generates temperature predictions for different boxes as outputs of interest for cold container temperature management.

**Figure 5.** Development of machine learning model to predict the temperature of each box within the cold container.

In the present study, two types of machine learning models, Random Forest (RF) and Neural Network (NN), are used. Below shows a brief description of the two models, and more details of the mathematical derivation of these two models can be found in [23] and [24], respectively.

2.5. RF Model

A RF model is an ensemble learning method that combines multiple decision trees, each trained on different bootstrap samples of data. The trees independently learn data patterns by optimizing split parameters, and their collective predictions reduce the error propagation that could occur in a single tree, making the model more robust against noise and variations in the dataset [25].

2.6. NN Model

A NN model is a class of machine learning algorithms designed to replicate the information processing functions of the human brain. The objective of the NN is thermal efficiency. Hidden layers are placed in between the objective and input layers for prediction. In the present model, the activation function for hidden layers uses the Rectified Linear Unit Function (ReLU). The hyperparameters (e.g., L2 penalty parameter) are used as defaults within the scikit-learn library, with seven input parameters as the input layer and thermal efficiency as the objective [13].

NN models have developed rapidly in recent years. A NN can be mathematically considered as a nonlinear regression model $f(x) = \varphi(w, x)$, where φ is a nonlinear model function, and w is the vector that contains the parameters in which x is known as the input. The basic units of a NN, known as perceptrons, can be computed as $f(x) = \varphi(w^T, x)$, where the nonlinear function φ is called the activation function. The training of a perceptron model is conducted through the updating of weights as follows [26]:

$$w_i^{t+1} = w_i^t + \eta \left(y^m - \varphi \left(x^T w^t \right) \right) x_i^m,$$

where η is the learning rate.

3. Discussion

A RF model uses 200 decision trees with a max depth of 15 and bootstrap sampling, while the NN employs a deep architecture of four hidden layers ($256 \rightarrow 128 \rightarrow 64 \rightarrow 32$ neurons) with ReLU activation, an Adam optimizer, and an adaptive learning rate. When developing the above machine learning model, 20% of the total generated CFD datasets are used for testing. The accuracy of the developed machine learning models is examined. Figure 6 shows the comparative analysis of RF and NN models across ten storage temperature cases. It reveals that RF consistently outperforms the NN across the temperature of ten boxes, particularly evident for storage boxes 1–3, where RF achieves slightly lower RMSE values and higher R^2 values compared to the NN's. This pattern continues across Storages 4–10, where RF maintains better performance. The performance gap is most pronounced in Storage 6, where RF (RMSE = 2.02 °C, R^2 = 0.948) minorly outperforms the NN (RMSE = 2.49 °C, R^2 = 0.922), demonstrating a 19% advantage in prediction accuracy. These findings indicate that RF models should be the preferred choice for storage temperature prediction under steady-state conditions, offering both better accuracy and more consistent performance across varying storage conditions, while also typically requiring less computational overhead than NN models.

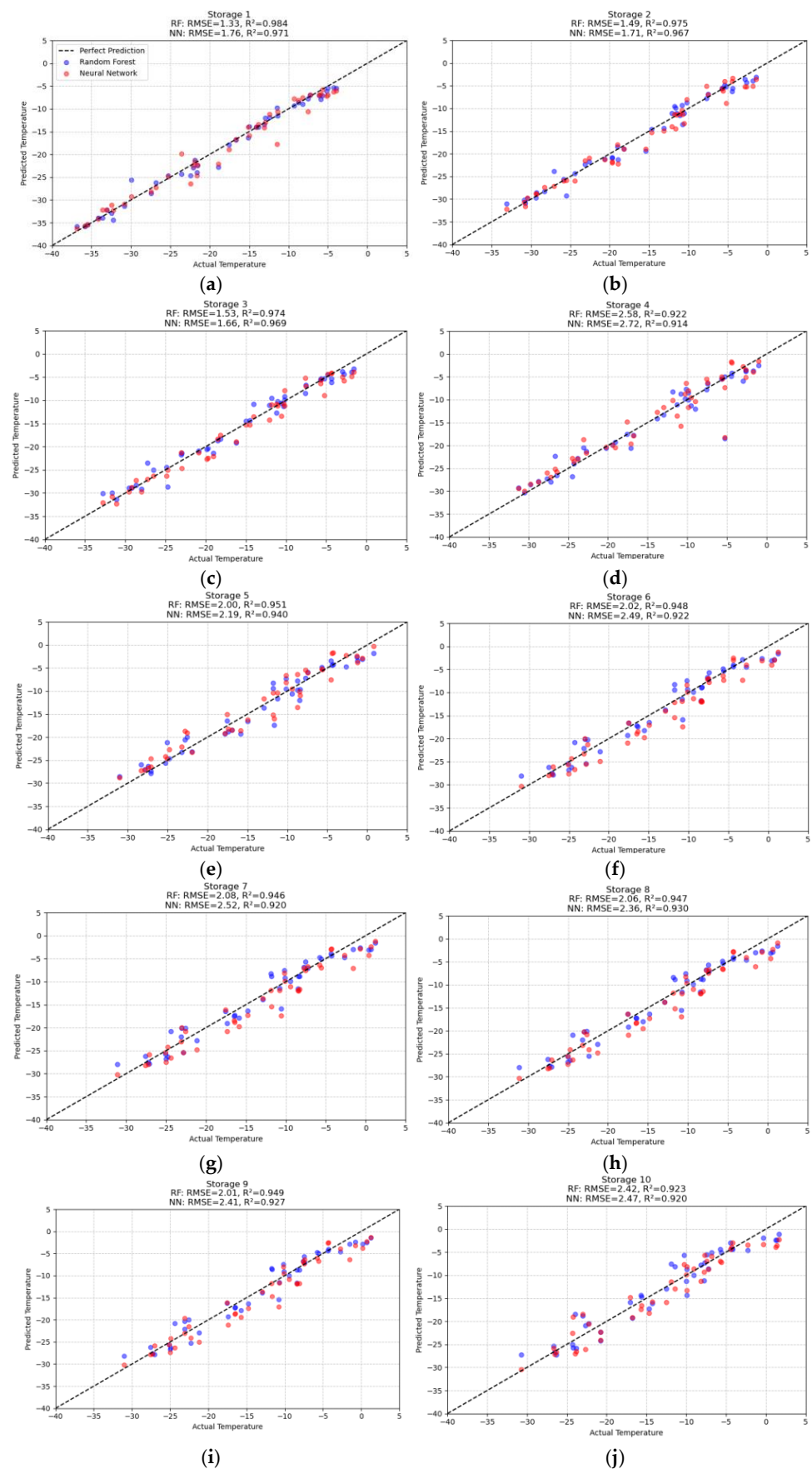


Figure 6. Comparison of RF and NN model performance in predicting steady-state storage temperatures: RMSE and R^2 analysis across ten different storage boxes, ranging from (a–j) for each storage box, respectively.

RF's better performance over the NN in storage temperature prediction may be attributed to three key factors: Firstly, RF's ensemble approach better captures the relatively stable patterns in steady-state temperature conditions without overcomplicating the relationships. Secondly, a decision tree structure effectively handles nonlinear relationships while remaining robust to outliers and noise, and its simpler architecture may be more appropriate for this specific prediction task, where the underlying temperature relationships are relatively straightforward. This is particularly evident in Storages 1–3, where RF maintains significantly better accuracy, suggesting that more complex NN architectures may be unnecessarily sophisticated for steady-state temperature prediction scenarios.

To understand the relative importance of different variable inputs, it is important to display relative feature importance values, showing how much each input variable contributes to determining the temperature of each box. Figure 7 presents the feature importance heatmap by revealing the key dominant importance of cold air temperature as the primary influential factor across all storage box temperatures, with importance values consistently exceeding 0.86 (86%). This relationship demonstrates a robust hierarchical pattern, where Box 1 exhibits the strongest correlation (0.982) with cold air temperature, followed by a gradual decrease in importance through Box 10 (0.863). The remaining environmental parameters, including air velocity, ambient temperature, external wall heat transfer coefficient, food thermal conductivity, and PUR insulator thermal conductivity, demonstrate notably lower importance values (generally below 0.044), suggesting their relatively minimal impact on storage temperature variations. This shows the importance values indicate that the thermal management system's performance is predominantly governed by cold air temperature regulation, while other thermal and fluid dynamic parameters play subsidiary roles in maintaining desired storage conditions.

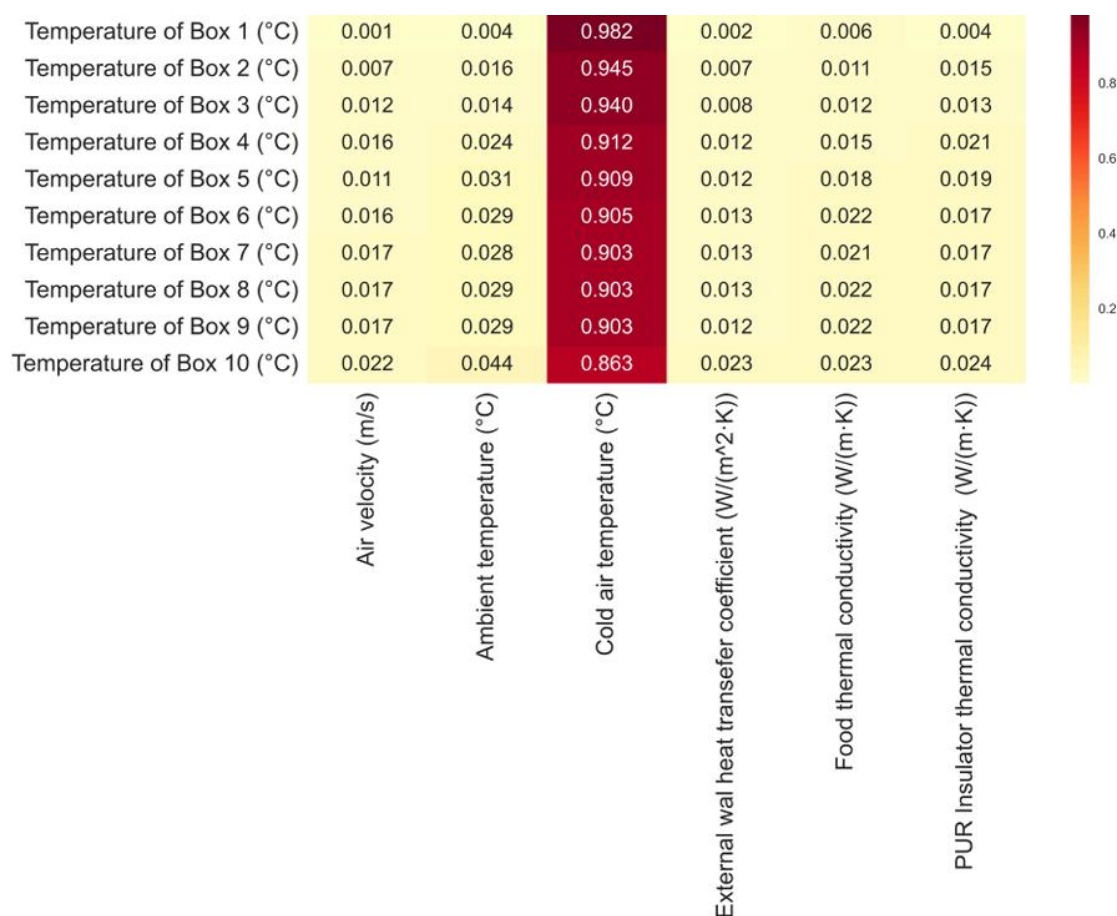


Figure 7. Relative feature importance variable inputs on temperature of ten boxes in cold container.

Figure 8 confirms observations from Figure 7, establishing cold air temperature as the dominant factor in storage temperature control (importance: 0.917), which far exceeds ambient temperature, PUR insulation, food thermal conductivity, air velocity, and wall heat transfer. This demonstrates the system's primary dependence on cold air temperature regulation, suggesting optimization efforts should prioritize precise control of this parameter while treating others as secondary factors in thermal management.

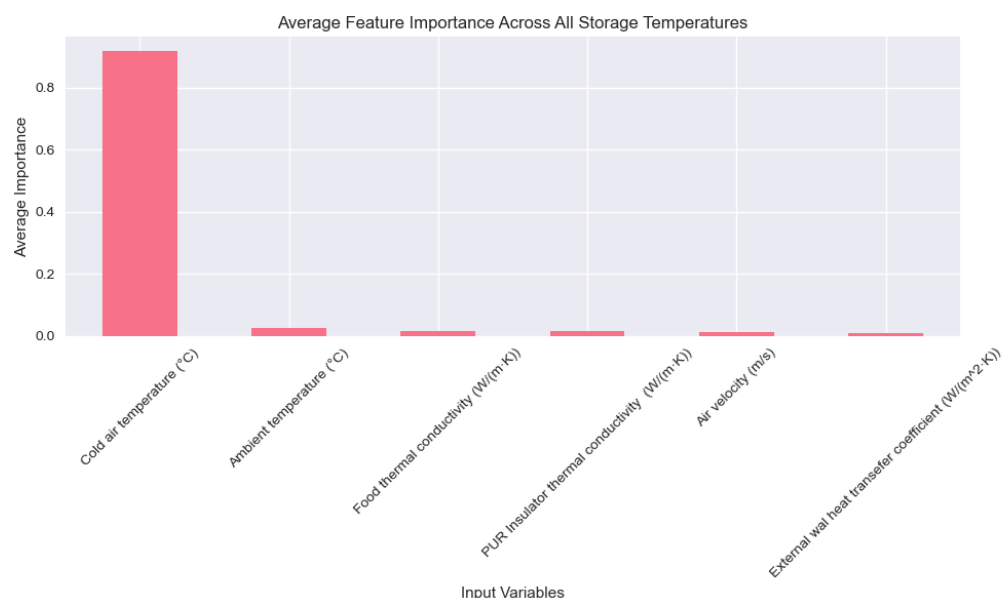


Figure 8. Average feature importance across all storages for different variable inputs.

Figure 9 reveals cold air temperature has the strongest positive correlation with storage temperature, decreasing from Box 1 to 10. Ambient temperature shows a moderate positive correlation, increasing across the boxes. Thermal conductivity and wall heat transfer show weak negative correlations, while air velocity effects appear distance-dependent. These findings confirm cold air temperature as the primary factor for maintaining desired storage conditions.

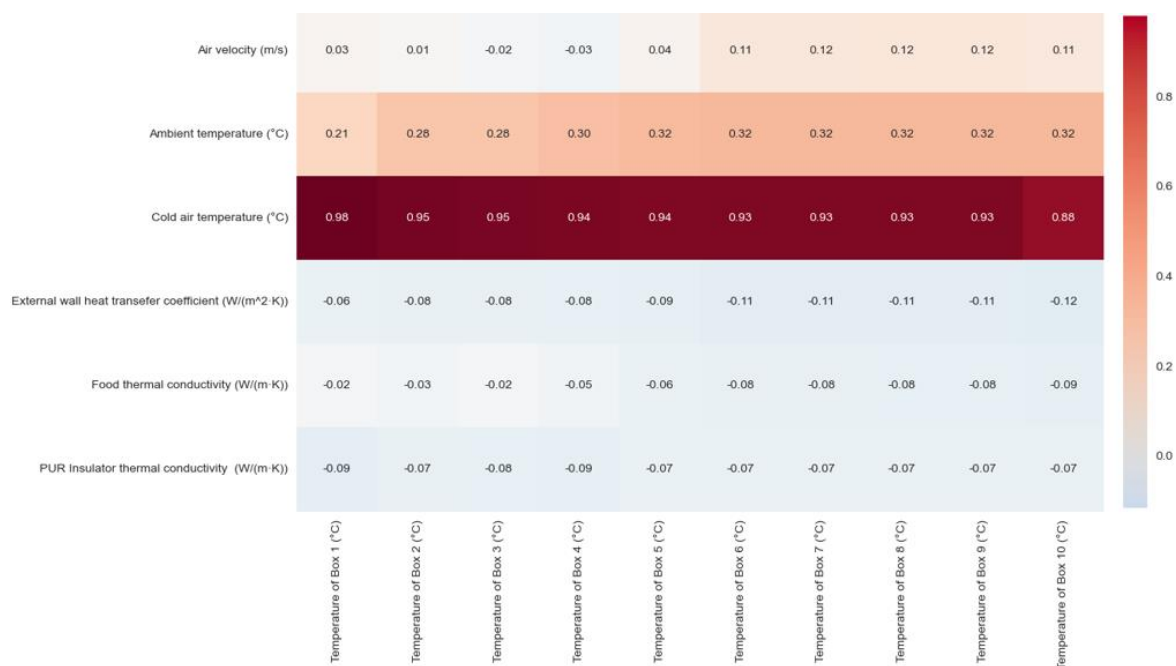


Figure 9. Correlation between input variables and ten storage boxes in cold container.

Referring to both Figures 7 and 9 provides valuable understanding because feature importance reveals causal influence within a model (which factors drive predictions), while correlation simply shows statistical relationships without implying causation, and altogether these highlight that cold air temperature is both highly important (0.86–0.98) and strongly correlated (0.88–0.98) with box temperature. Thus, this provides complementary insights for understanding the thermal dynamics of the cold container system.

Based on the above analysis, it shows that cold air temperature and ambient temperature are the two most important variables that determine the thermal performance of storage boxes inside cold containers. To further understand the role of these two input variables, Figure 10 presents a dual scatter plot analysis to reveal distinctly different relationships between storage box temperatures and the two input variables. The left panel demonstrates a strong positive linear correlation between cold air temperature ($-40\text{ }^{\circ}\text{C}$ to $-5\text{ }^{\circ}\text{C}$) and box temperatures, with correlation coefficients consistently above 0.88 across all boxes. Box 1 (pink markers) consistently maintains the lowest temperatures, with subsequent boxes showing incrementally higher temperatures, indicating a systematic thermal gradient across the storage system. In contrast, the right panel illustrates a notably weaker and more dispersed relationship between ambient temperature ($10\text{ }^{\circ}\text{C}$ to $40\text{ }^{\circ}\text{C}$) and box temperatures, characterized by substantial scatter and limited correlation (correlation coefficients ranging from 0.21 to 0.32). This comparative visualization effectively highlights the dominant influence of cold air temperature on the storage system's thermal behavior, while ambient temperature exhibits minimal impact on internal box temperatures, likely due to effective insulation. This shows that PUR insulation material is very effective at providing sufficient thermal insulation between the cold air and the ambient environment in general. The outliers in Figure 10 may be due to recirculation patterns creating isolated thermal environments that differ significantly from the general trend. These recirculation zones resulting from pressure differentials cause warmer air to circulate back into certain box locations despite the overall cold air temperature setting.

Interpretability is an important aspect of machine learning models. One of the most used methods for interpretability evaluation is SHapley Additive exPlanation (SHAP). The SHAP analysis reveals distinct patterns of feature importance and their impacts on machine learning model predictions for outputs of interest (e.g., temperature of boxes in the cold container in the present study). It provides a mathematical framework for interpreting machine learning model predictions by assigning each feature an importance value based on game theory principles [25]. It addresses the “black box” nature of machine learning models by quantifying both individual and interactive effects of input variables on model outputs, enabling model validation behavior against physical principles and deriving actionable insights for system optimization.

Storage box 5 is chosen for SHAP analysis, and the results are presented in Figure 11. From Figure 11, SHAP dependence analysis of Box 5's temperature control system reveals complex thermal interactions across six key parameters. Cold air temperature (e.g., ranging from -40 to $-5\text{ }^{\circ}\text{C}$ for frozen food) demonstrates the strongest linear correlation, establishing itself as the primary control variable, while ambient temperature (e.g., 10 – $40\text{ }^{\circ}\text{C}$, which covers typical environment temperatures) shows a robust positive linear relationship, indicating its significant external influence. Air velocity exhibits optimal cooling effectiveness between 0.2 and 0.4 m/s , with diminishing returns at higher velocities, suggesting a critical operational range for system efficiency. The external wall heat transfer coefficient shows clustered behavior below $20\text{ W}/(\text{m}^2\cdot\text{K})$, indicating a potential optimization threshold. Food thermal conductivity (e.g., $<100\text{ W}/(\text{m}\cdot\text{K})$) displays a negative correlation, with considerable scatter, highlighting the importance of product-specific considerations. Thermal conductivity of PUR insulation material (e.g., $<0.200\text{ W}/(\text{m}\cdot\text{K})$) demonstrates a

weak positive correlation with high variability, suggesting its role as a secondary factor in temperature control.

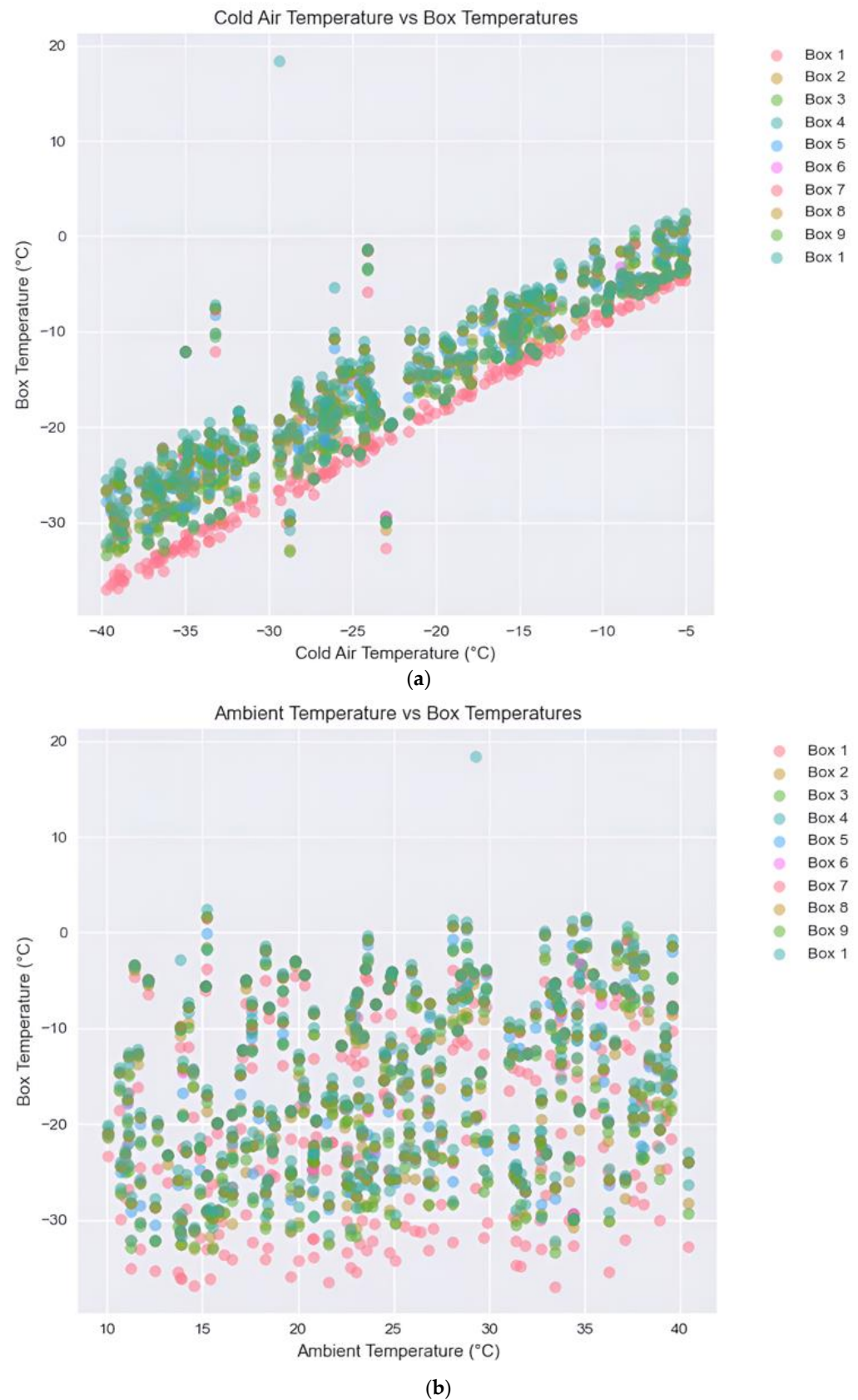


Figure 10. Comparative analysis of (a) cold air and (b) ambient temperature effects on storage boxes in cold container.

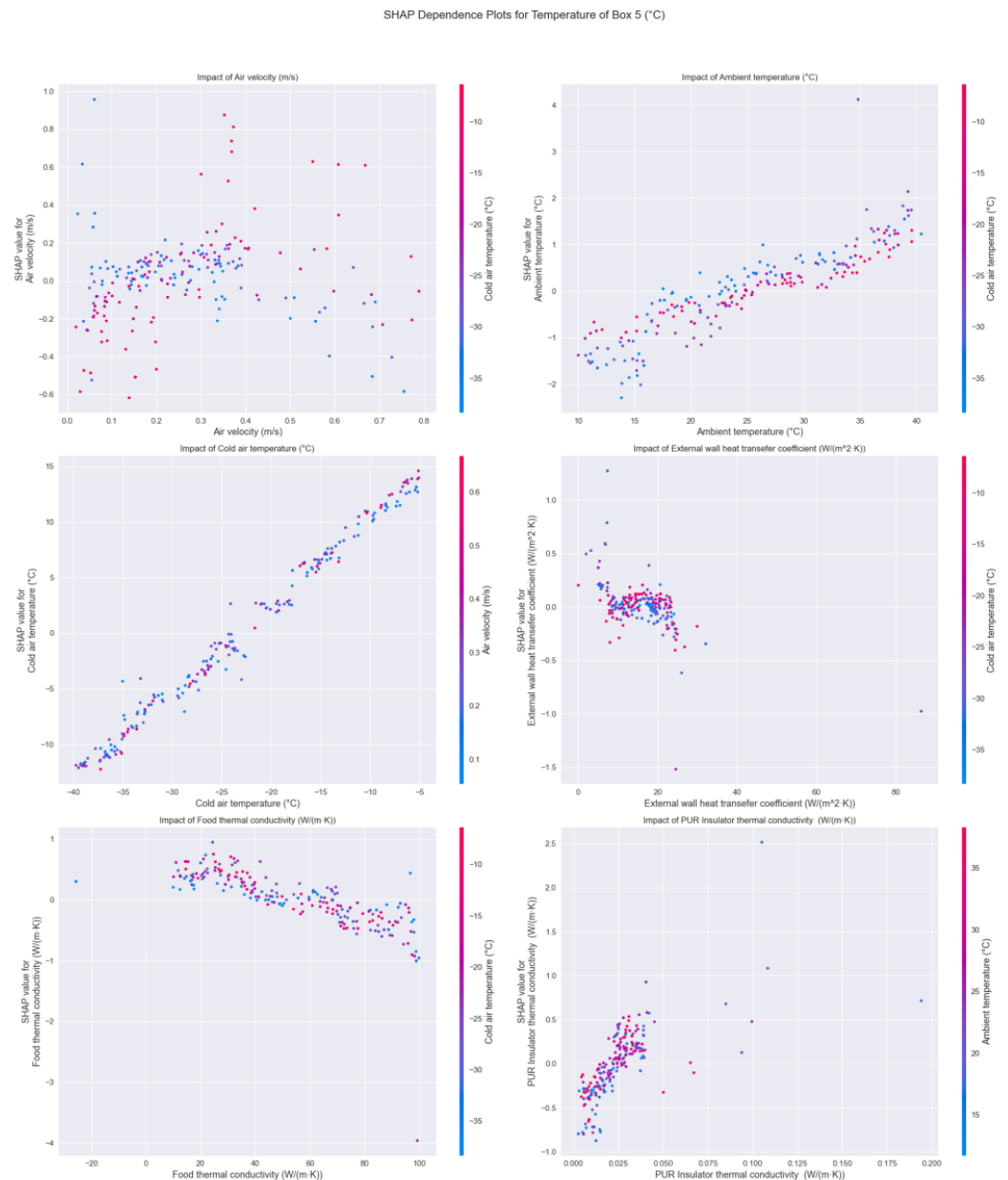


Figure 11. SHAP dependence results of variable inputs Influencing temperature of storage Box 5.

In Figure 12, the SHAP summary plot analysis of Box 5's temperature control system reveals a distinct hierarchical influence of thermal and physical parameters on system performance. Cold air temperature emerges as the dominant control variable, exhibiting the most extensive SHAP value distribution (−10 to +10) with a pronounced bimodal pattern, indicating its critical role in system control. Ambient temperature demonstrates a secondary but significant influence, with moderate SHAP values (−5 to +5) and clear temperature-dependent effects, highlighting the importance of environmental conditions. Food thermal conductivity, PUR insulator conductivity, air velocity, and external wall heat transfer coefficient display progressively diminishing impacts, with their SHAP values clustered tightly around zero, suggesting their roles as less important factors. These findings indicate that system optimization should prioritize precise cold air temperature control mechanisms and robust ambient temperature compensation while maintaining cost-effective management of secondary parameters. This hierarchical understanding enables targeted design improvements and operational strategies, potentially leading to enhanced system efficiency through focused control system development and resource allocation.

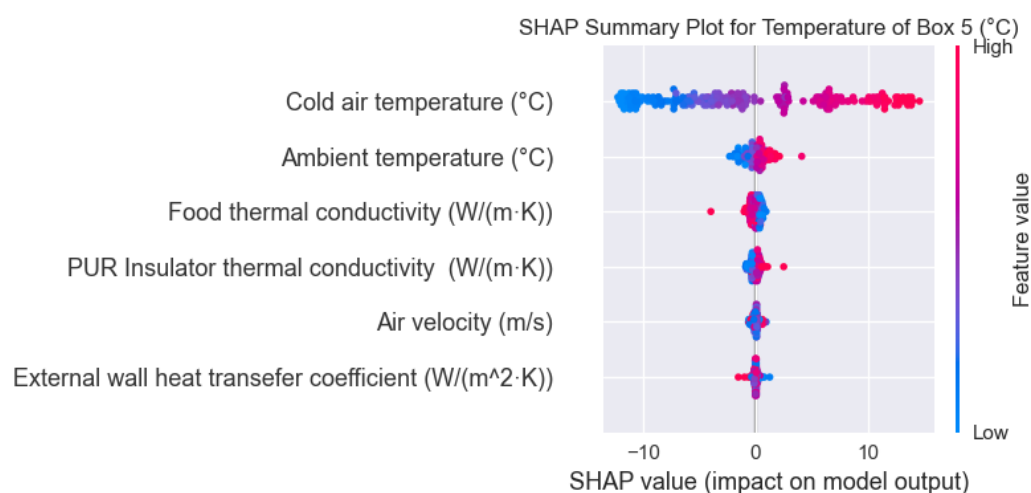


Figure 12. SHAP summary of hierarchical impact assessment of process input variables on the temperature of Storage Box 5 in the cold container.

4. Conclusions

In this study, a full-scale CFD model was developed and validated against experimental data from the literature. Machine learning models were subsequently developed to provide a comprehensive analysis of temperature management for storage boxes in cold containers. The integration of these computational approaches yielded several significant findings in the following aspects:

- Leveraging the symmetrical nature of the cold room along its longitudinal axis proved to be computationally advantageous, enabling a more efficient modeling approach without compromising the accuracy of temperature and airflow predictions. This symmetry-based modeling strategy demonstrates how geometric characteristics can be effectively utilized to reduce computational overhead in complex thermal systems.
- The hybrid model combining CFD and machine learning demonstrates superior efficiency in cold storage temperature prediction, with the RF model consistently outperforming the NN across all storage boxes (RMSE improvement of up to 19% and R^2 values exceeding 0.948).
- Cold air temperature emerges as the dominant control variable (importance value > 0.86) for storage temperature management, while other parameters, including ambient temperature, air velocity, and thermal conductivity properties, show significantly lower influence (importance values < 0.044), suggesting optimization efforts should primarily focus on cold air temperature control.
- SHAP analysis reveals optimal operational ranges for key parameters, notably air velocity (0.2–0.4 m/s) and external wall heat transfer coefficient ($< 20 \text{ W/m}^2\cdot\text{K}$), providing valuable insights for system design and optimization.
- The PUR insulation material proves highly effective in thermal management, evidenced by the weak correlation between ambient temperature (10–40 °C) and storage box temperatures, indicating robust insulation performance in maintaining desired storage conditions.

The present study innovatively provides a hybrid approach that integrates CFD with machine learning to overcome critical limitations in cold storage temperature management. By leveraging symmetrical domain modeling, the research simultaneously reduces computational demands while maintaining prediction accuracy. The present work also identifies optimal operational thresholds, which provide actionable engineering guidelines that balance product preservation with energy efficiency. This hybrid methodology bridges the gap between high-fidelity physics-based modeling and practical prediction ca-

pabilities needed for real-world cold chain optimization, ultimately advancing sustainable temperature management in critical supply chains.

Author Contributions: Methodology, L.G. and X.H.; Software, L.G. and X.H.; Resources, Y.L.; Writing—original draft, Y.L.; Writing—review & editing, M.Z.; Supervision, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Negi, S.; Anand, N. Issues and Challenges in the Supply Chain of Fruits & Vegetables Sector in India: A Review. *Int. J. Manag. Value Supply Chain*. **2015**, *6*, 47–62.
2. Mercier, S.; Villeneuve, S.; Mondor, M.; Uysal, I. Time–Temperature Management Along the Food Cold Chain: A Review of Recent Developments. *Compr. Rev. Food Sci. Food Saf.* **2017**, *16*, 647–667. [[PubMed](#)]
3. Raab, V.; Bruckner, S.; Beierle, E.; Ilg, Y.; Petersen, B.; Kreyenschmidt, J. Generic model for the prediction of remaining shelf life in support of cold chain management in pork and poultry supply chains. *J. Chain Netw. Sci.* **2008**, *8*, 59–73. [[CrossRef](#)]
4. Derens, E.; Palagos, B.; Guilpart, J. The cold chain of chilled products under supervision in France. In Proceedings of the 13th World Congress of Food Science & Technology 2006, Nantes, France, 17–21 September 2006.
5. Norton, T.; Sun, D.-W. Computational fluid dynamics (CFD)—An effective and efficient design and analysis tool for the food industry: A review. *Trends Food Sci. Technol.* **2006**, *17*, 600–620.
6. Umeno, Y.; Hung, D.V.; Tanaka, F.; Hamanaka, D.; Uchino, T. The use of CFD to simulate temperature distribution in refrigerated containers. *Eng. Agric. Environ. Food* **2015**, *8*, 257–263.
7. Daniel, L.; Jakhar, S.; Dasgupta, M.S. Optimizing cold storage for uniform airflow and temperature distribution in apple preservation using CFD simulation. *Sci. Rep.* **2024**, *14*, 25402.
8. Yang, B.; Melikov, A.K.; Kabanshi, A.; Zhang, C.; Bauman, F.S.; Cao, G.; Awbi, H.; Wigö, H.; Niu, J.; Cheong, K.W.D.; et al. A review of advanced air distribution methods—theory, practice, limitations and solutions. *Energy Build.* **2019**, *202*, 109359. [[CrossRef](#)]
9. Ben Taher, M.; Kousksou, T.; Zeraouli, Y.; Ahachad, M.; Mahdaoui, M. Thermal performance investigation of door opening and closing processes in a refrigerated truck equipped with different phase change materials. *J. Energy Storage* **2021**, *42*, 103097.
10. Carneiro, R.; Gaspar, P.D.; Silva, P.D. 3D and transient numerical modelling of door opening and closing processes and its influence on thermal performance of cold rooms. *Appl. Therm. Eng.* **2017**, *113*, 585–600.
11. Catlow, R.; Burke, P.; Goodfellow, J.; Tildesley, D.; Wilson, M.; Cant, S. High-performance computing in computational fluid dynamics: Progress and challenges. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **2002**, *360*, 1211–1225.
12. Eze, J.; Duan, Y.; Eze, E.; Ramanathan, R.; Ajmal, T. Machine learning-based optimal temperature management model for safety and quality control of perishable food supply chain. *Sci. Rep.* **2024**, *14*, 27228.
13. Loisel, J.; Cornuéjols, A.; Laguerre, O.; Tardet, M.; Cagnon, D.; Duchesne de Lamotte, O.; Duret, S. Machine learning for temperature prediction in food pallet along a cold chain: Comparison between synthetic and experimental training dataset. *J. Food Eng.* **2022**, *335*, 111156.
14. Alkhulaifi, N.; Bowler, A.L.; Pekaslan, D.; Serdaroglu, G.; Closs, S.; Watson, N.J.; Triguero, I. Machine Learning Pipeline for Energy and Environmental Prediction in Cold Storage Facilities. *IEEE Access* **2024**, *12*, 153935–153951.
15. Kim, T.H.; Kim, J.H.; Kim, J.Y.; Oh, S.E. Egg Freshness Prediction Model Using Real-Time Cold Chain Storage Condition Based on Transfer Learning. *Foods* **2022**, *11*, 3082. [[CrossRef](#)]
16. Zhang, M.; Sun, J.; Fricke, B.; Nawaz, K.; Gluesenkamp, K.; Shen, B.; Munk, J.; Liu, X. A study on computational fluid dynamics modeling of a refrigerated container for COVID-19 vaccine distribution with experimental validation. *Int. Commun. Heat Mass Transf.* **2022**, *130*, 105749.
17. Shaheed, R.; Mohammadian, A.; Kheirkhah Gildeh, H. A comparison of standard $k-\epsilon$ and realizable $k-\epsilon$ turbulence models in curved and confluent channels. *Environ. Fluid Mech.* **2019**, *19*, 543–568.
18. Smale, N.J.; Moureh, J.; Cortella, G. A review of numerical models of airflow in refrigerated food applications. *Int. J. Refrig.* **2006**, *29*, 911–930.
19. Seeger, M. Gaussian processes for machine learning. *Int. J. Neural Syst.* **2004**, *14*, 69–106.

20. Carson, J.K.; Wang, J.; North, M.F.; Cleland, D.J. Effective thermal conductivity prediction of foods using composition and temperature data. *J. Food Eng.* **2016**, *175*, 65–73.
21. Verma, S.; Singh, H. Vacuum insulation in cold chain equipment: A review. *Energy Procedia* **2019**, *161*, 232–241.
22. Ferrandez-Garcia, C.; Ferrández-García, A.; Ferrández-Villena, M.; Hidalgo-Cordero, J.F.; García-Ortuño, T.; Ferrández-García, M.-T. Physical and Mechanical Properties of Particleboard Made from Palm Tree Prunings. *Forests* **2018**, *9*, 755. [[CrossRef](#)]
23. Fawagreh, K.; Gaber, M.M.; Elyan, E. Random forests: From early developments to recent advancements. *Syst. Sci. Control Eng.* **2014**, *2*, 602–609. [[CrossRef](#)]
24. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[PubMed](#)]
25. Gharaee, H.; Erfanimatin, M.; Bahman, A.M. Machine learning development to predict the electrical efficiency of photovoltaic-thermal (PVT) collector systems. *Energy Convers. Manag.* **2024**, *315*, 118808. [[CrossRef](#)]
26. Comito, C.; Pizzuti, C. Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. *Artif. Intell. Med.* **2022**, *128*, 102286.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.