www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 18(1), 2027, 237-255. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2027.v18i1.1465

A multi-output CNN1D–GRU–MLP Architecture for Thermal Time-to-Risk Classification in Cold Chain Logistics

Tania Ofelia López Zerón, Adrián Sánchez Ortega, *Marcelo Quezada Quiterio, Enrique de Jesús Mohedano Torres, Uriel Marrón Hernández, Gabriel Jiménez Zerón*

[1]Instituto Tecnológico de Pachuca

tania.lz@pachuca.tecnm.mx, l23200899@pachuca.tecnm.mx, l23200881@pachuca.tecnm.mx, enrique.mt@pachuca.tecnm.mx, uriel.mh@pachuca.tecnm.mx, jzeron@uaeh.edu.mx

Abstract. Maintaining an uninterrupted cold chain is essential for ensuring the safety and quality of food and pharmaceutical products. However, conventional monitoring systems mainly provide descriptive information and offer limited support for anticipating thermal risks. This study presents a hybrid CNN1D-GRU-MLP multi-output architecture for thermal time-to-risk classification in multizone cold-chain environments. Due to the limited availability of labeled operational data, the model was initially evaluated using a synthetic dataset generated from energy-balance formulations. Using 30-minute thermal observation windows, the architecture simultaneously performs four tasks: temperature forecasting, time-to-risk interval classification, thermal state classification, and critical zone identification. The interval-bases formulation provides decision-oriented time ranges that are more useful for logistics operations than exact minute-level predictions. Although validation with real-word data remains necessary, the proposes methodology demonstrates the potential to transform thermal trajectories into meaningful information that supports earlier interventions and more effective cold-chain management.

Keywords: Cold chain logistics, Thermal time series, Thermal time-to-risk classification, CNN1D-GRU neural network, Multi-output learning, Temperature forecasting, Physically informed synthetic data, Remaining Useful Life (RUL).

Article Info
Received Jun 26, 2026
Accepted Aug 11, 2026

1. Introduction

Cold-chain logistics are essential for preserving temperature-sensitive products from production to the final consumer. The scale of the problem extends beyond isolated temperature deviations: the lack of effective refrigeration was estimated to account for the loss of 526 million tonnes of food, equivalent to 12% of global food production in 2017 (UNEP & FAO, 2022). This evidence supports the need for monitoring approaches that can identify changes in thermal trajectories before operational limits are exceeded.

Chen and Shaw (2011) combined an exponentially weighted moving average (EWMA) control chart with a back-propagation neural network to monitor temperature variation and predict shifts and trends in data collected using active RFID temperature sensors. Their feasibility demonstration was conducted in a simulated environment built with LEGO® bricks. This study provides an early example of integrating real-time temperature acquisition with data-driven analysis in cold-chain management; the present work extends the predictive objective toward the joint analysis of three zonal temperature sequences and the classification of the remaining time interval before an operational threshold is crossed.

Deep-learning studies have extended refrigeration monitoring from descriptive records toward sequence-based forecasting and diagnosis. Hoang et al. (2021) evaluated traditional, stacked, bidirectional, and convolutional LSTM models using experimental cold-room data to predict temperature and power-demand perturbations associated with demand-response

events. In a different refrigeration application, Wang et al. (2020) combined a one-dimensional convolutional neural network with a GRU to diagnose faults in chiller time-series data, using the convolutional component to extract local features and the recurrent component to capture global and dynamic information. Together, these studies support the use of temporal deep learning for refrigeration-system data; however, neither evaluates multizone refrigerated transport nor classifies the remaining interval before a thermal threshold is crossed.

The aim of this study is more specific than general temperature forecasting, it is how to model thermal behavior across several cargo zones at once and translate that into something an operator can act on. Most published work stops at a single predicted value, which tells you how warm cargo is getting but not how much time remains before a safety threshold is crossed, the goal is to transform the data into something more useful for the logistics area. Duan et al. (2026) advanced temperature monitoring by using a BiLSTM-attention virtual-sensor model to estimate food temperatures at sensor-inaccessible pallet locations; with a fixed number of physical sensors, the reported RMSE remained below 0.3 °C when the virtual-to-physical sensor ratio increased from 16 to 32. Separately, Adefemi et al. (2025) reported up to 99.83% accuracy for a CNN-GRU intrusion-detection model on IoT network datasets; because its task and data were cybersecurity-oriented, that result does not establish performance for thermal forecasting or logistics time-to-risk classification.

Considering most of the variables that influence in thermal fluctuations is key in building a more accurate model, even though it can be more complex. At the same time combining the classifying power of a CNN seems to aide in the processing and learning of a GRU. The proposed CNN1D–GRU–MLP multi-output architecture combines CNN1D feature extraction, GRU temporal modeling, and an MLP shared representation. Its primary output is an interval class for the time remaining before the simulated thermal threshold is crossed, rather than an exact minute estimate; temperature forecasting, thermal-state classification, and critical-zone identification are auxiliary outputs. Here, thermal time-to-risk refers only to this time-to-threshold classification and not to classical equipment RUL or product shelf life. Operational usefulness remains to be evaluated with real data and users.

Instrumented cold chain data is difficult to obtain at the volumes that deep learning requires, the main reason being the amount of data needed from an individual event, on top of multiple events required. A single trip could or could not contain a significant event or multiple events useful to train the model in the different possible scenarios. Zhang, K. et al. (2023) combined JUST and STL decomposition with a two-layer Bi-LSTM to predict univariate daily mean temperatures for Chinese cities, reporting an RMSE of 0.2187. This provides a general antecedent for decomposed temperature-series prediction, but it does not validate preprocessing for multizone cold chain data. Xie et al. (2025) addressed real-time anomaly detection in multidimensional cold chain streams with an improved iForest model that uses sliding windows and cross-variable correlations, reporting an average F1-score of 0.8639 and an average AUC of 0.9064. Consequently, that study should not be characterized as static or single-snapshot analysis; its distinction from the present work is that it detects anomalies rather than forecasting future temperature and classifying the remaining interval before a thermal threshold is crossed. Together, these studies support temporal prediction and online multivariate anomaly detection, respectively, but they do not resolve the need for instrumented multizone sequences aligned with the four outputs evaluated here. Accordingly, the initial computational evaluation used physically informed synthetic sequences; the generator, energy-balance formulation, operational events, and zonal construction are described in the Experimental Procedures section.

Our study´s approach was shaped by three related areas of existing research. First, Bavaresco et al. (2021) mapped IoT applications for occupational well-being in Industry 4.0, including the interaction between sensor-based systems and workers. However, that study did not evaluate time-bounded thermal outputs or compare interval alerts with minute-level predictions; therefore, it is retained only as a broader IoT antecedent. Second, Mutambik (2026) integrated blockchain, IoT, and an Informer model in a pharmaceutical vaccine cold-chain framework and reported a MAPE below 3% for vaccine demand forecasting. This result supports forecasting within vaccine cold-chain management, but it does not demonstrate that prediction accuracy causes stakeholder trust or validate thermal time-to-threshold classification. Finally, Zhang, Y. et al. (2026) combined multisource temperature sensing and deep learning in a box-level digital twin to reconstruct in-box product temperatures and estimate dynamic shelf life. Compared with air-temperature-based modeling, their approach reduced RMSE and MAE by approximately 60%; however, its output was dynamic shelf life rather than the remaining interval before an operational temperature threshold. Together, these studies provide related precedents for connected monitoring, predictive modeling, and operational interpretation, but none directly evaluates the interval-based thermal time-to-threshold classification proposed here. Accordingly, this work retains the design premise that thermal sequences can be translated into bounded operational horizons, while their usefulness for operators remains to be validated with real deployments and users.

Accordingly, this study aimed to develop and computationally evaluate a CNN1D–GRU–MLP multi-output proof of concept using physically informed synthetic multizone thermal sequences for 3.5-ton refrigerated units. The primary task was interval classification of the horizon until a simulated operational temperature threshold was crossed, while temperature forecasting, thermal-state classification, and critical-zone identification were evaluated as auxiliary outputs. The base model was compared with a class-weighted preventive variant using class-wise precision, recall, F1-score, and confusion matrices, without claiming validated performance in real refrigerated transport.

2. Experimental procedures

This study followed a two-component methodological design comprising (i) a structured literature review and evidence synthesis and (ii) a controlled simulation-based computational experiment. The first component established the architectural, application, and methodological basis of the proposal, whereas the second evaluated a CNN1D–GRU–MLP multi-output proof of concept under reproducible synthetic conditions. Model performance was quantified using mean absolute error (MAE) and root mean square error (RMSE) for temperature forecasting and accuracy, class-wise precision, recall, F1-score, and confusion matrices for the classification outputs.

This investigation begins with the operational gap as its starting point. Refrigerated units generate continuous thermal series in different cargo zones; however, in monitoring-centered implementations, these data primarily support retrospective records rather than predictive time-to-risk estimates. The methodological design therefore addressed the transition from thermal records to anticipatory decision support under controlled computational conditions.

The design of this study was organized into two components to address this gap. The first consisted of a systematic literature review conducted through an integrated SALSA–PRISMA approach. SALSA structured the sequential Search, Appraisal, Synthesis, and Analysis phases (Mengist et al., 2020), while PRISMA 2020 was applied complementarily to structure and transparently report record identification, duplicate removal, title-and-abstract screening, eligibility assessment, and final inclusion (Page et al., 2021a, 2021b). The review focused on recurrent and hybrid deep-learning architectures, thermal time-series analysis, temperature forecasting, anomaly detection, and time-to-threshold formulations in cold-chain logistics and related industrial monitoring contexts.

The controlled simulation-based computational experiment served as the second component of the study. The hybrid CNN1D–GRU–MLP multi-output architecture was trained and evaluated using physically informed synthetic multizone thermal sequences generated with the corrected v4.1 simulator for a 3.5-ton refrigerated compartment. All neural configurations used the same effective windowed dataset, trip-level partition, normalization parameters estimated exclusively from the training partition, output definitions, and final test partition.

The evaluated configurations comprised a base model without manual class weighting and two preventive class-weighted variants with w30–60 = 1.8 and w30–60 = 2.0. These variants were evaluated as an exploratory sensitivity analysis of class weighting rather than as an optimized cost matrix. The 1.8 configuration was close to the inverse-frequency estimate of 1.7495, whereas the 2.0 configuration represented a more conservative preventive weighting of the 30–60 min class. The comparison examined whether increasing the contribution of preventive samples during training altered the detection of early-warning intervals, particularly the misclassification of 30–60 min samples as >60 min / No near risk.

This experimental phase was developed as a direct response to the findings of the systematic review, rather than as an isolated design choice. The gaps identified in the literature guided the use of multizone inputs, temporal feature extraction, recurrent sequence modeling, interval-based thermal time-to-risk classification, complementary outputs, and physically informed synthetic data to establish an initial performance baseline. Consequently, the reported findings constitute a controlled computational proof-of-concept evaluation and are not interpreted as operational validation in an instrumented refrigerated vehicle.

2.1. Systematic Literature Review (Integrated SALSA–PRISMA Approach)

Phase 1 — Search and PRISMA identification/screening: A protocol-driven search was conducted across Scopus, Web of Science, and Google Scholar using structured Boolean operators. The search terms were organized into three primary semantic axes: architecture, application domain, and methodological focus.

The architecture axis focused on terms such as 'recurrent neural networks,' 'LSTM,' 'GRU,' and 'CNN-GRU.' The application-domain axis included 'cold chain,' 'cold chain logistics,' 'refrigerated transport,' and 'thermal monitoring.' Finally, the methodological axis employed 'time series,' 'anomaly detection,' 'temperature prediction,' and 'Remaining Useful Life' (RUL).

The search was restricted to peer-reviewed publications between 2010 and 2026, identifying an initial pool of 1,450 records across the three databases. After duplicate removal and screening of titles and abstracts, 1,050 records remained for the subsequent appraisal stage.

Phase 2 — Appraisal and PRISMA eligibility/inclusion: Inclusion and exclusion criteria were applied by two independent reviewers, with discrepancies resolved through consensus.

Studies were eligible for inclusion if they met three conditions: (a) explicit use of recurrent or hybrid deep-learning architectures; (b) application to thermal time series in supply-chain or industrial IoT contexts; and (c) reporting of at least one quantitative performance metric.

Review articles lacking original experimental results, conference abstracts, and studies focused exclusively on non-thermal logistics variables were excluded. Following this appraisal process, 29 high-relevance studies were retained for systematic synthesis and analysis.

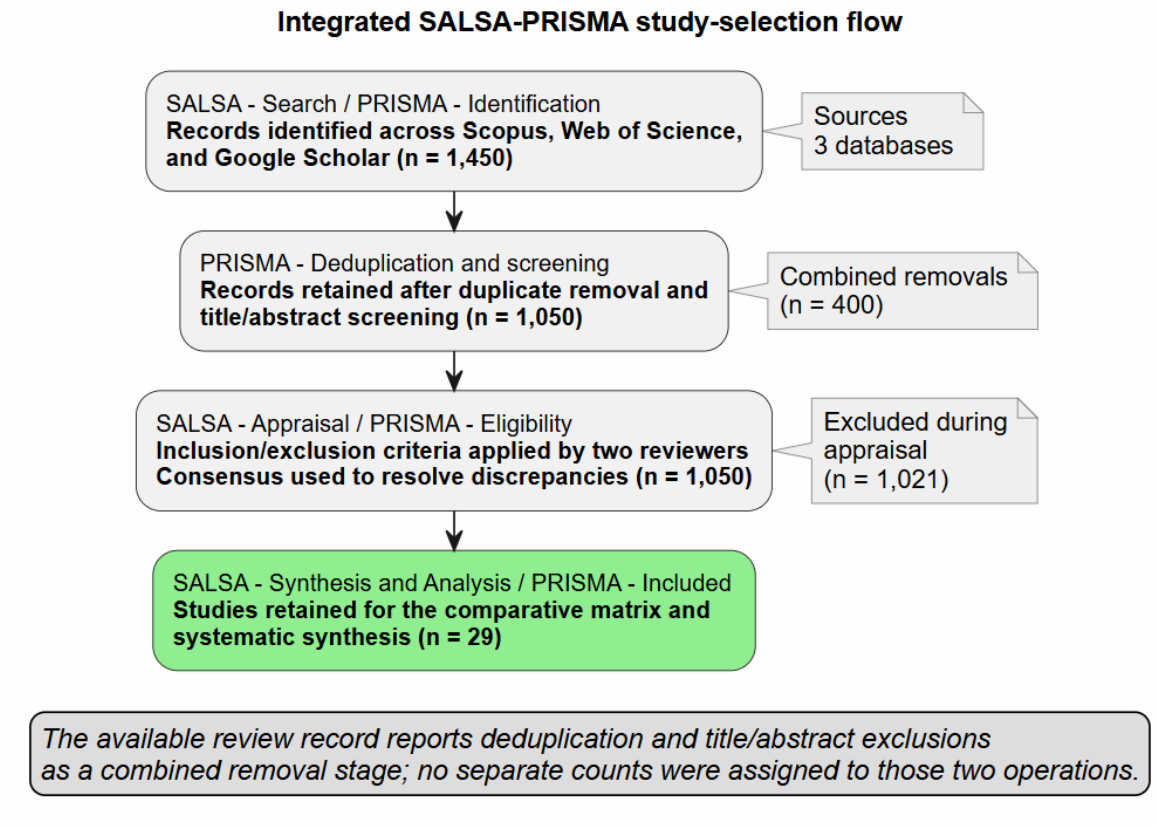

Figure 1 *Integrated SALSA–PRISMA study-selection flow used in the literature review.*

Phase 3 — Synthesis: Data from the 29 selected studies were organized into a structured comparative matrix documenting their objectives, methodologies, primary results, limitations, and specific relevance to this study.

The articles were categorized into three architectural families: (a) pure recurrent models, such as LSTM or GRU; (b) hybrid CNN–RNN models; and (c) multi-output or multi-task architectures. This synthesis indicated that single-output temperature forecasting is prevalent, whereas multi-output architectures specifically applied to refrigerated transport remain scarce. Table 1 presents the 10 studies with the most direct theoretical and technical relevance to the RNN-CF-IM framework.

Phase 4 — Analysis: The cross-study comparison identified three recurring constraints within the reviewed corpus: (a) reliance on univariable or single-zone inputs, which do not represent internal spatial gradients; (b) absence of interval-based thermal time-to-risk formulations tailored to logistics decision windows; and (c) dependence on instrumented real-world datasets that are difficult to obtain at transportation scale.

These gaps directly informed the proposed approach: a multizone input architecture, an interval-based thermal time-to-risk output providing bounded decision horizons, and a physically informed synthetic dataset for the controlled initial evaluation and performance baseline.

Table *1*. Summary of the 10 most relevant articles included in the literature review

Title / Source (APA)	Technical Objective	Key Methodology & Architecture	Main Results & Specific Metrics	Identified Gaps & Refinement for Current Study
Applying back propagation network... *(Chen & Shaw, 2011)*	Monitor collected temperature data and predict temperature shifts and trends.	Back-propagation neural network for shift and trend prediction, combined with an EWMA control chart for monitoring temperature variation.	Demonstrated the feasibility of a temperature-monitoring system using active RFID temperature sensors and a simulated environment built with LEGO® bricks.	Scope relative to the present study: the stated objective addresses temperature monitoring and shift/trend prediction; it does not report multi-zone threshold-interval classification.
Data-driven virtual sensor systems... *(Duan et al., 2026)*	Estimate food temperature in sensor-inaccessible locations within pallets.	BiLSTM with an attention mechanism using multisource data (initial temp, ambient, and pallet sensors).	With a fixed number of physical sensors, RMSE remained below 0.3 °C when the virtual-to-physical sensor ratio increased from 16 to 32; a one-sensor-per-pallet BiLSTM-attention configuration outperformed shallow networks.	Scope relative to the present study: estimates temperature at sensor-inaccessible pallet locations; it does not classify the remaining interval before a thermal threshold is crossed or evaluate a CNN-GRU model.

Title / Source (APA)	Technical Objective	Key Methodology & Architecture	Main Results & Specific Metrics	Identified Gaps & Refinement for Current Study
A Hybrid CNN–GRU Deep Learning Model... *(Adefemi et al., 2025)*	Improve anomaly detection in high-volume industrial IoT datasets.	Hybrid CNN-GRU with SMOTE for handling extreme class imbalance.	Achieved 99.83% accuracy on IoTID20, surpassing baseline models by 2–3% margin.	Scope relative to the present study: evaluates intrusion detection in IoT network datasets; it does not test thermal forecasting, multizone cargo signals, or logistics time-to-risk classification.
Temperature Time Series Prediction Model... *(Zhang, K. et al., 2023)*	Predict temperature series by reducing random noise and fluctuations.	STL and JUST algorithms for decomposition into trend/seasonal components + 2-layer Bi-LSTM.	Average RMSE of 0.2187, significantly lower than standard Bi-LSTM (RMSE 4.3997). Proved superior stability in 60 Chinese cities.	Purely univariable model; does not account for operational factors like door openings or localized hotspots.
An anomaly detection scheme... *(Xie et al., 2025)*	Detect real-time abnormal events in high-dimensional cold chain streams.	Improved iForest (Isolated Forest) using cross factors and sliding windows (length = 200).	F1-score of 0.8639 and AUC of 0.9064 in 150-dimensional data per carriage.	Scope relative to the present study: uses sliding windows and temporal-spatial correlations for online anomaly detection; it does not forecast future temperature or classify the remaining interval before a thermal threshold is crossed.
A Blockchain–IoT–ML Framework... *(Mutambik, 2026)*	Secure and sustainable management of pharmaceutical cold chains.	Integration of Blockchain for traceability, IoT, and Informer model for long-sequence forecasting.	The Informer model achieved a MAPE below 3% for vaccine demand forecasting; BERT sentiment analysis achieved 80% accuracy.	Scope relative to the present study: evaluates vaccine demand forecasting and stakeholder sentiment; it does not model thermal trajectories, classify time-to-threshold, or compare the computational efficiency of Informer and GRU.

Title / Source (APA)	Technical Objective	Key Methodology & Architecture	Main Results & Specific Metrics	Identified Gaps & Refinement for Current Study
Development of deep learning... models... *(Hoang et al., 2021)*	Predict temperature and power demand for Demand Response (DR).	Comparative evaluation of traditional, stacked, BiLSTM, and Convolutional LSTM.	Conv-LSTM achieved the best performance in predicting product temperatures under sudden thermal variations.	Focused on fixed warehouses; does not address multi-zone complexity in urban light distribution (3.5-ton trucks).
A digital twin–driven dynamic shelf life... *(Zhang, Y. et al., 2026)*	Mitigate food loss through digital twin-driven real-time shelf life.	Multisource sensing coupled with LSTM and Arrhenius kinetic models.	Approximately 60% reductions in RMSE and MAE relative to air-temperature-based modeling; uncertainty in baseline deterioration kinetics dominated shelf-life prediction error.	Scope relative to the present study: reconstructs product temperature and estimates dynamic shelf life; it does not classify the remaining interval before an operational temperature threshold.
Internet of Things and occupational well-being... *(Bavaresco et al., 2021)*	Analyze IoT impact on worker interaction with smart environments.	Systematic mapping and taxonomy of 67 articles, 11 segments, and 32 sensor types.	Systematic mapping and taxonomy of IoT applications associated with occupational well-being in Industry 4.0.	Scope relative to the present study: does not evaluate cold-chain thermal forecasting, time-bounded temperature alerts, or interval-based time-to-threshold classification.
A Novel Fault Diagnosis Approach... *(Wang et al., 2020)*	Hybrid fault diagnosis for chillers based on time-series sequences.	1D-CNN for local features + GRU for global/dynamic features (128 filters, 256 GRU units).	Superior identification rate for minor faults compared to state-of-the-art algorithms on RP-1043.	Primarily focused on industrial data center chillers rather than the dynamic environments of refrigerated transport.

Note. Own elaboration (2026) based on the 10 studies listed in the table.

The studies summarized in Table 1 show that hybrid CNN–RNN architectures have been applied successfully to several time-series monitoring tasks; however, the reviewed corpus does not support a universal claim that they consistently outperform every single-architecture baseline. Across the selected studies, multizone inputs, interval-based thermal time-to-risk classification for logistics decision windows, and reproducible initial evaluation using physically informed synthetic transport data were not jointly addressed. These combined gaps informed the design decisions of RNN-CF-IM.

2.2. Controlled Computational Experiment and Implementation Phases

The controlled simulation-based computational experiment was implemented in five stages: (1) physically informed synthetic data generation; (2) operational delimitation, window construction, and trip-level partitioning; (3) interval-based thermal time-

to-risk and auxiliary-label construction; (4) CNN1D–GRU–MLP multi-output modeling; and (5) training, preventive class-weight sensitivity analysis, baseline comparison, and evaluation.

Phase 1: Physically informed synthetic data generation.

Because an instrumented multizone dataset containing the four labels required by the model was not available to this study, the initial proof of concept used the corrected v4.1 synthetic generator. It generated minute-level thermal trajectories for a 3.5-ton refrigerated compartment serving light urban dairy-distribution routes. The generator represented a compartment-level base trajectory with zone-specific offsets, localized perturbations, and stochastic sensor noise for the Front, Middle, and Door zones. It was not designed as a calibrated digital twin, a computational-fluid-dynamics model, or three independent thermodynamic balances.

The base thermal trajectory followed the simplified energy balance shown in Eq. (1), and its one-minute discrete implementation is shown in Eq. (2)**.**

$$C\frac{dT}{dt} = UA\,[T_{amb}(t) - T(t)] - Q_{cold}(t) + Q_{event}(t) \tag{1}$$

$$T(t+1) = T(t) + \frac{\Delta t}{C}\{UA[T_{amb}(t) - T(t)] - Q_{cold}(t) + Q_{event}(t)\} + \varepsilon(t) \tag{2}$$

Here, T is the compartment-level base temperature, Tamb is ambient temperature, C is the equivalent thermal capacity, UA represents global heat transfer, Qcold(t) is effective cooling capacity, Qevent(t) represents event-related thermal perturbations, Δt = 60 s, and ε(t) is stochastic noise. For each trip, U, A, C, and nominal cooling capacity were sampled from 0.35–0.85 W m−2 K−1, 18–32 m2, 2.5 × 105–2.2 × 106 J K−1, and 120–850 W, respectively. Initial compartment temperature ranged from 2.2 to 5.2 °C, ambient temperature from 24 to 38 °C, and the synthetic control target from 3.2 to 4.8 °C. Gaussian sensor noise used a standard deviation of 0.05 °C, increased to 0.10 °C in the thermal-spike scenario.

The conceptual form of the balance and expected direction of the simulated thermal responses were reviewed with support from a physics specialist as a plausibility check. This review did not constitute empirical calibration or operational validation of a specific vehicle.

Twelve scenario families were generated: stable operation, gradual warming, weak compressor, compressor shutdown, short and repeated door openings, localized hotspots, thermal recovery, thermal spikes, ambient variation, localized overcooling, and global overcooling. The sampling distribution intentionally emphasized progressive warming and refrigeration-loss scenarios; each overcooling family had a generation probability of 0.5%. In v4.1, repeated door openings were constrained to non-overlapping intervals separated by at least one closed minute. Door-open states took precedence over recovery states, and door_event_id and planned_door_event_count were retained only for traceability and temporal audits, not as neural-network inputs or targets. Dataset export was conditional on passing recovery-causality, door-sequence, and trip-overlap audits.

Phase 2: Operational delimitation, window construction, and partitioning.

Synthetic trip durations represented short (70–120 min), medium (120–210 min), and long (210–300 min) light-distribution routes, sampled with probabilities of 0.45, 0.40, and 0.15, respectively. Temperature was sampled once per minute. Each model input contained 30 consecutive observations from the Front, Middle, and Door zones, producing the input tensor in Eq. (3).

$$\mathbf{X} \in \mathbb{R}^{30 \times 3} \tag{3}$$

Windows were generated with a one-minute stride and therefore overlapped by 29 observations. A window was retained only when all four future forecast labels at +5, +10, +15, and +30 min were available within the same trip. The regression target forecast_temp was the maximum future temperature across the three zones at each horizon. Minimum future temperatures were exported for auditability but were not used to train the final regression head. All windows with a thermal time-to-risk of 60 min or less were retained; windows labeled >60 min / No near risk were retained with probability 0.35 to limit dominance of long stable segments.

A total of 1,000 synthetic trips produced 145,689 minute-level records. After window construction and controlled subsampling, 999 trips contributed at least one retained window, yielding 53,270 30 × 3 samples. One trip generated candidate windows but contributed none after subsampling of the >60 min / No near risk class. The effective trips were partitioned by trip_id with a fixed seed of 42: 699 trips for training, 150 for validation, and 150 for testing. These partitions contained 38,140, 7,471, and 7,659 windows, respectively. Trip-level partitioning kept temporally correlated windows from the same trip within a single partition.

Normalization statistics for the three input channels and the four regression horizons were estimated exclusively from the training partition. The resulting means and standard deviations were then applied unchanged to training, validation, and test data. Forecast predictions were transformed back to degrees Celsius before MAE and RMSE were calculated.

Phase 3: Interval-based thermal time-to-risk and auxiliary-label construction.

For each retained window, thermal time-to-risk was defined as the earliest future minute at which any zone exceeded the upper synthetic operational threshold of 7.0 °C or fell below the lower threshold of −0.510 °C. Values were capped at 120 min. These thresholds were used as labeling assumptions for the controlled synthetic experiment and were not presented as universal cold-chain limits for all dairy products, jurisdictions, vehicles, or operating conditions. The continuous value was converted into four operational intervals as shown in Table 2.

Table 2. Definition of thermal time-to-risk intervals and operational interpretation

Class / interval	Operational interpretation
0: 0–10 min (≤10)	**Immediate risk**
1: >10–30 min	**High preventive risk**
2: >30–60 min	**Moderate preventive risk**
3: >60 min	**>60 min / No near risk**

The thermal_state target combined the current three-zone temperatures with the time-to-risk label. A sample was Critical when any current temperature was outside the labeling thresholds or when time-to-risk was ≤10 min; it was Vigilance when any zone was within 1.0 °C of the upper threshold, within 0.5 °C of the lower threshold, or when time-to-risk was ≤30 min; otherwise it was Normal. For the output retained under the implementation name critical_zone, the target denoted the first zone expected to cross a threshold when a future excursion existed. When no future excursion was identified within the capped horizon, it denoted the zone currently nearest to either threshold. It is therefore interpreted in the manuscript as an associated-zone classification rather than as evidence that every sample contains an already critical zone.

The generator also recorded risk direction (warming, cooling, or no future excursion) and zone-label type (risk zone or nearest-limit zone) for auditing. These variables were not additional output heads. The final dataset was dominated by warming-related windows, and cold-risk windows were absent from the final test partition; consequently, the evaluation protocol did not assess cold-risk detection.

Phase 4: CNN1D–GRU–MLP multi-output modeling.

The network received a 30 × 3 thermal window and used two one-dimensional convolutional layers, each with 24 filters, kernel size 3, same padding, and ReLU activation, to extract local variations and short perturbations. A GRU layer with 48 units, dropout of 0.15, and recurrent dropout of 0.10 summarized the temporal evolution of the feature maps. GRU was

selected to reduce structural complexity while preserving temporal-sequence modeling capacity for moderate-length windows; the study did not perform an empirical GRU-versus-LSTM architecture comparison.

The shared multilayer perceptron contained a 48-unit ReLU layer, dropout of 0.30, and a 24-unit ReLU layer. Four heads were connected to the shared representation: forecast_temp, a four-value linear regression output for maximum future temperature across zones; rul_class, a four-class softmax output for thermal time-to-risk; thermal_state, a three-class softmax output for Normal, Vigilance, and Critical; and critical_zone, a three-class softmax output for the associated Front, Middle, or Door zone. The complete model contained 16,526 trainable parameters. Figure 2 presents the processing flow and output definitions.

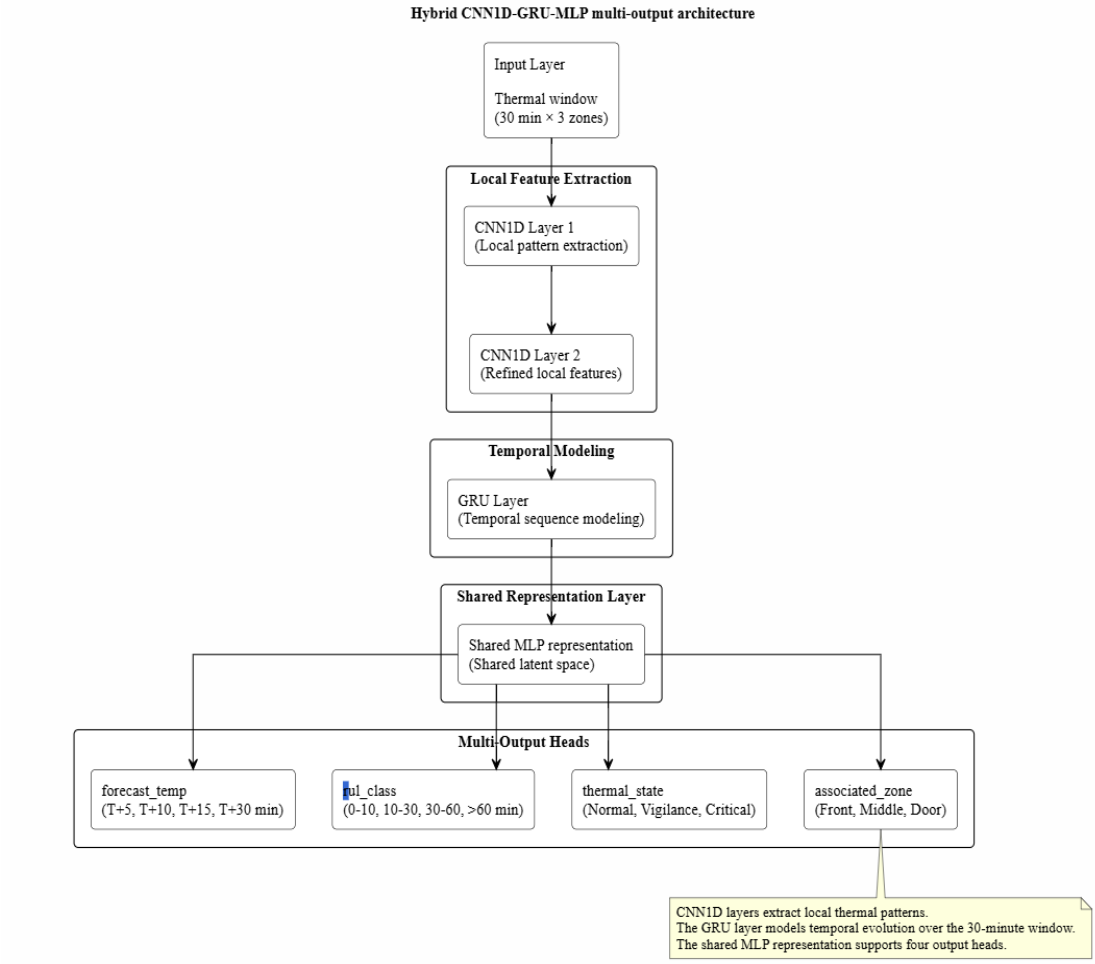

***Figure 2**. CNN1D–GRU–MLP multi-output architecture and output definitions.*

Phase 5: Training, preventive sensitivity analysis, baseline, and evaluation.

All configurations were implemented in TensorFlow/Keras 2.20.0 with seed 42 assigned to Python, NumPy, TensorFlow, and the trip-level split. Deterministic TensorFlow operations were enabled. The Adam optimizer used an initial learning rate of 5 × 10−4. Mean squared error was used for forecast_temp, and sparse categorical cross-entropy was used for the three classification heads. The task-loss weights were 0.7 for forecast_temp, 2.0 for rul_class, 1.5 for thermal_state, and 1.0 for critical_zone. Training used a maximum of 50 epochs and batch size 64. Early stopping monitored validation loss with patience 8 and restored the best weights; ReduceLROnPlateau used factor 0.5, patience 4, and a minimum learning rate of 1 × 10−5.

Three training configurations used the same architecture, data partitions, normalization, optimizer, loss functions, task-loss weights, callbacks, and final test set. The base model used no manual sample weights. Both preventive configurations applied sample weights only to rul_class during training: class weights were 1.0, 1.7, 1.8, and 1.0 for the 0–10, 10–30, 30–60, and

>60 min classes in the first variant, and 1.0, 1.7, 2.0, and 1.0 in the second. Validation and test samples were evaluated with unit weights.

Inverse-frequency weights were first calculated from the 38,140 training samples using wi = N/(Kni), where N is the number of training windows, K = 4 classes, and ni is the class-specific count. Training counts of 14,092, 5,735, 5,450, and 12,863 produced weights of 0.6766, 1.6626, 1.7495, and 0.7413, respectively. The 1.7 weight rounded the estimate for the 10–30 min class. For the 30–60 min class, 1.8 was evaluated as a value close to 1.7495 and 2.0 as a more conservative preventive setting. This comparison was an exploratory sensitivity analysis and was not designed to identify an optimized cost matrix.

A deterministic non-neural baseline was recalculated using the final v4.1 dataset. For each zone, it combined the recent five-minute slope with a 30-minute linear-regression slope, extrapolated the maximum future temperature at the same four horizons, and estimated time to the upper or lower threshold from the most adverse positive or negative slope. Baseline calculations were confined to each trip.

Evaluation was performed once on the fixed test partition after model selection based on validation loss. Forecasting was assessed with MAE and RMSE in degrees Celsius. Each classification head was assessed with accuracy, and rul_class was additionally examined through class-wise precision, recall, F1-score, support, and confusion matrices. The primary comparison focused on missed early-warning cases, particularly 30–60 min samples assigned to >60 min / No near risk, together with the reciprocal increase in false preventive assignments.

2.3. Experimental Configuration and Reproducibility Controls

Table 3 consolidates the fixed experimental conditions used across all neural configurations. The generator, windowed dataset, trip split, normalization parameters, and test partition were frozen before the preventive weight comparison; only the training sample weights applied to rul_class differed among configurations.

Table 3. Final v4.1 experimental configuration

Element	Final configuration
Generator	Corrected v4.1 physically informed synthetic generator
Generated / effective trips	1,000 generated; 999 contributed retained windows
Minute-level records	145,689
Retained samples	53,270 windows; 30 min × 3 zones; one-minute stride
Trip-level split	Train 699; validation 150; test 150; seed 42
Window split	Train 38,140; validation 7,471; test 7,659
Input normalization	Mean and standard deviation estimated on train only
Forecast target	Maximum future temperature across zones at +5, +10, +15, and +30 min
Architecture	Conv1D(24)–Conv1D(24)–GRU(48)–Dense(48)–Dropout(0.30)–Dense(24); 16,526 parameters
Training	TensorFlow/Keras 2.20.0; Adam 5 × 10−4; batch 64; maximum 50 epochs
Model configurations	Base; preventive w30–60 = 1.8; preventive w30–60 = 2.0
Evaluation	Fixed test set; MAE/RMSE; accuracy; class-wise precision, recall, F1-score; confusion matrices

The resulting evidence is interpreted as a reproducible computational proof of concept under the defined synthetic conditions. It does not establish calibrated thermodynamic fidelity, operational performance in an instrumented vehicle, or validated detection of cold-risk events.

2.4. Analytical Hypotheses and Decision Criteria

The revised analysis was organized around three explicit hypotheses. These hypotheses concern predictive performance and class-weight sensitivity under the controlled synthetic experiment; they do not assert operational effectiveness in a real refrigerated vehicle or a causal synergy among the neural-network components.

H1 — Joint predictive performance: On the same fixed test partition, the unweighted multi-output model is expected to reduce temperature-forecast MAE and RMSE and to increase the overall accuracy of thermal time-to-risk, thermal-state, and associated-zone classification relative to the deterministic trend-based baseline.

H2 — Interval discrimination: The thermal time-to-risk head is expected to distinguish all four operational intervals on unseen trips, evidenced by non-zero class-wise F1-scores and an overall accuracy above the empirical majority-class reference. This criterion evaluates computational discrimination and not user-perceived actionability.

H3 — Preventive sensitivity: Relative to the unweighted model, increasing the training weight of the 30–60 min class is expected to increase its recall and F1-score and to reduce 30–60 min samples assigned to >60 min / No near risk. Changes in global accuracy and reciprocal false preventive assignments are treated as the corresponding trade-off rather than as evidence of global superiority.

3. Results

3.1. Final Test Set and Deterministic Baseline

The final test partition contained 7,659 windows from 150 trips that were not used for model training or validation. The thermal time-to-risk supports were 3,193 samples for 0–10 min, 1,133 for >10–30 min, 918 for >30–60 min, and 2,415 for >60 min / No near risk. The empirical majority-class reference for this output was therefore 0.4169. No cold-risk window was present in the test partition; the results in this section primarily characterize warming-related threshold events.

The deterministic trend-based baseline and the unweighted multi-output model were aligned to the same 7,659 test windows. The baseline produced complete predictions for every test sample. As shown in Table 4, the neural model reduced forecast error and increased the overall accuracy of the three classification outputs. These comparisons are descriptive results from the fixed test partition.

Table 4. Deterministic baseline and unweighted multi-output model on the final test partition

Metric	Deterministic baseline	Unweighted multi-output model
Forecast MAE (°C)	0.7452	0.5167
Forecast RMSE (°C)	1.8857	0.8255
Thermal time-to-risk accuracy	0.6685	0.6861
Thermal-state accuracy	0.8207	0.8390
Associated-zone accuracy	0.7329	0.9227

3.2. Performance of the Base Multi-output Model

The unweighted model achieved a thermal time-to-risk accuracy of 0.6861. Its strongest class-wise result was obtained for the 0–10 min interval (F1 = 0.9107), followed by >60 min / No near risk (F1 = 0.6650). Performance was lower for >10–30 min (F1 = 0.3617) and >30–60 min (F1 = 0.2390). Table 5 reports one authoritative value, to four decimal places, for every class metric.

Table 5. Class-wise thermal time-to-risk performance of the unweighted model

Thermal time-to-risk interval	Precision	Recall	F1-score	Support
0–10 min	**0.9669**	**0.8606**	**0.9107**	**3,193**
>10–30 min	**0.5113**	**0.2798**	**0.3617**	**1,133**
>30–60 min	**0.3171**	**0.1917**	**0.2390**	**918**
>60 min / No near risk	**0.5530**	**0.8340**	**0.6650**	**2,415**

The confusion pattern in Figure 3 shows that 601 of the 918 true >30–60 min samples were assigned to >60 min / No near risk, whereas 176 were classified correctly. For the >10–30 min interval, 669 of 1,133 samples were assigned to >60 min / No near risk. These counts identify missed early-warning assignments as the principal weakness of the unweighted time-to-risk output.

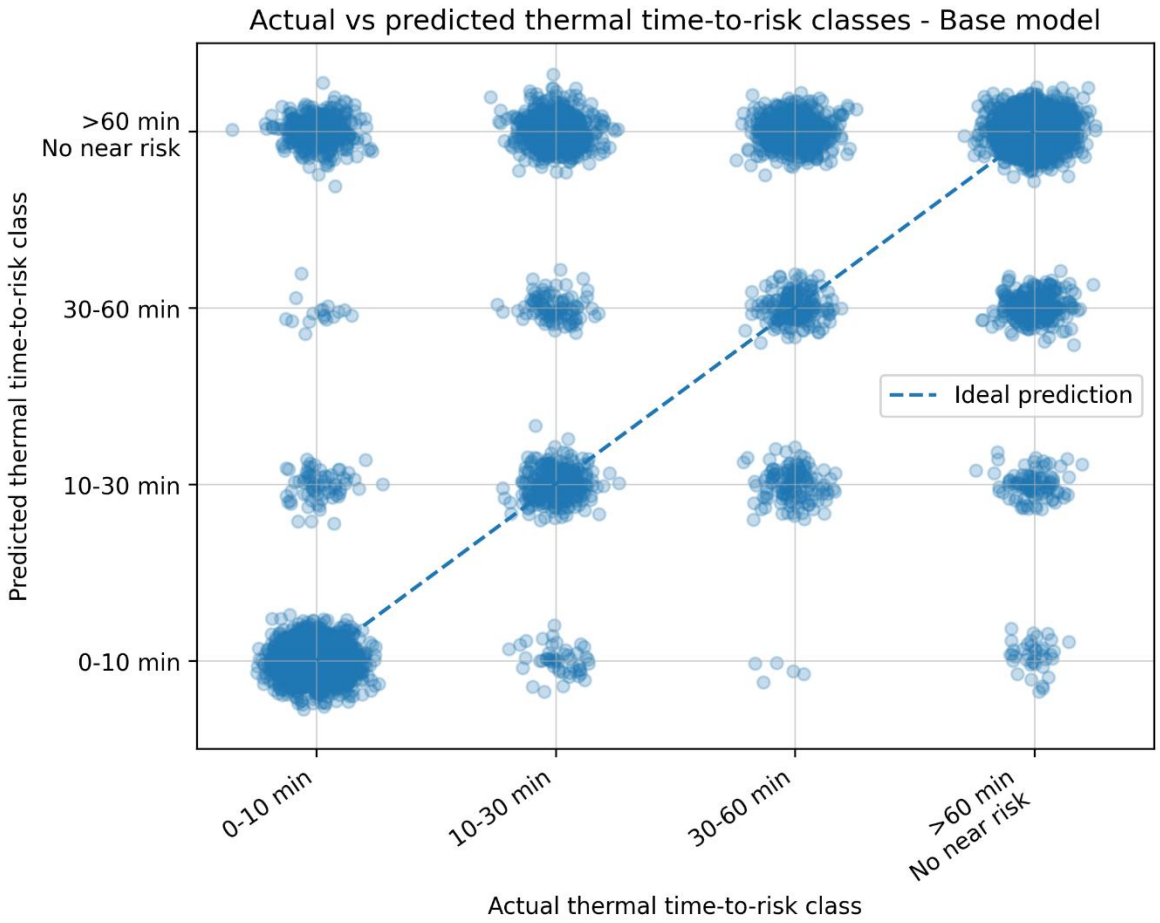

Figure 3. *Actual versus predicted thermal time-to-risk classes for the unweighted multi-output model on the final test partition.*

For the auxiliary outputs, overall thermal-state accuracy was 0.8390 and associated-zone accuracy was 0.9227. Table 6 shows that these aggregate values were not uniform across classes. The Vigilance state had an F1-score of 0.6878, and the Middle associated zone had an F1-score of 0.6828, compared with 0.9262 for Front and 0.9518 for Door. The associated-zone result is interpreted according to the labeling rule defined in Methodology and is not evidence that every window contained an already critical physical zone.

Table 6. Class-wise performance of the auxiliary classification outputs in the unweighted model

Output / class	Precision	Recall	F1-score	Support
Thermal state — Normal	0.7303	0.9818	0.8376	2,752
Thermal state — Vigilance	0.8538	0.5758	0.6878	1,714
Thermal state — Critical	0.9765	0.8572	0.9129	3,193
Associated zone — Front	0.9748	0.8823	0.9262	1,402
Associated zone — Middle	0.8678	0.5629	0.6828	851
Associated zone — Door	0.9166	0.9898	0.9518	5,406

3.3. Preventive Weighting Analysis

Two preventive configurations were compared with the unweighted model using the same dataset, split, normalization, architecture, and test partition. Table 7 shows that global thermal time-to-risk accuracy remained within 0.0020 of the unweighted result across the two weighted configurations. The other output metrics also changed only modestly; therefore, neither weighted configuration is interpreted as globally superior.

Table 7. Overall performance across the three neural training configurations

Metric	Unweighted	Preventive 1.8	Preventive 2.0
Forecast MAE (°C)	0.5167	0.5064	0.5013
Forecast RMSE (°C)	0.8255	0.8433	0.8170
Thermal time-to-risk accuracy	0.6861	0.6853	0.6842
Thermal-state accuracy	0.8390	0.8367	0.8361
Associated-zone accuracy	0.9227	0.9149	0.9256

The class-wise F1 comparison in Table 8 and Figure 4 shows the principal sensitivity effect. For >30–60 min, F1 increased from 0.2390 in the unweighted model to 0.3214 with a weight of 1.8 and to 0.3476 with a weight of 2.0. Recall for this class increased from 0.1917 to 0.3039 and 0.3573, respectively. The 0–10 min F1-score remained approximately unchanged, whereas the >60 min / No near risk F1-score decreased from 0.6650 to 0.6597 and 0.6518.

Table 8. Class-wise thermal time-to-risk F1-scores across training configurations

Interval	Unweighted	Preventive 1.8	Preventive 2.0
0–10 min	0.9107	0.9099	0.9107
>10–30 min	0.3617	0.3845	0.3894
>30–60 min	0.2390	0.3214	0.3476
>60 min / No near risk	0.6650	0.6597	0.6518

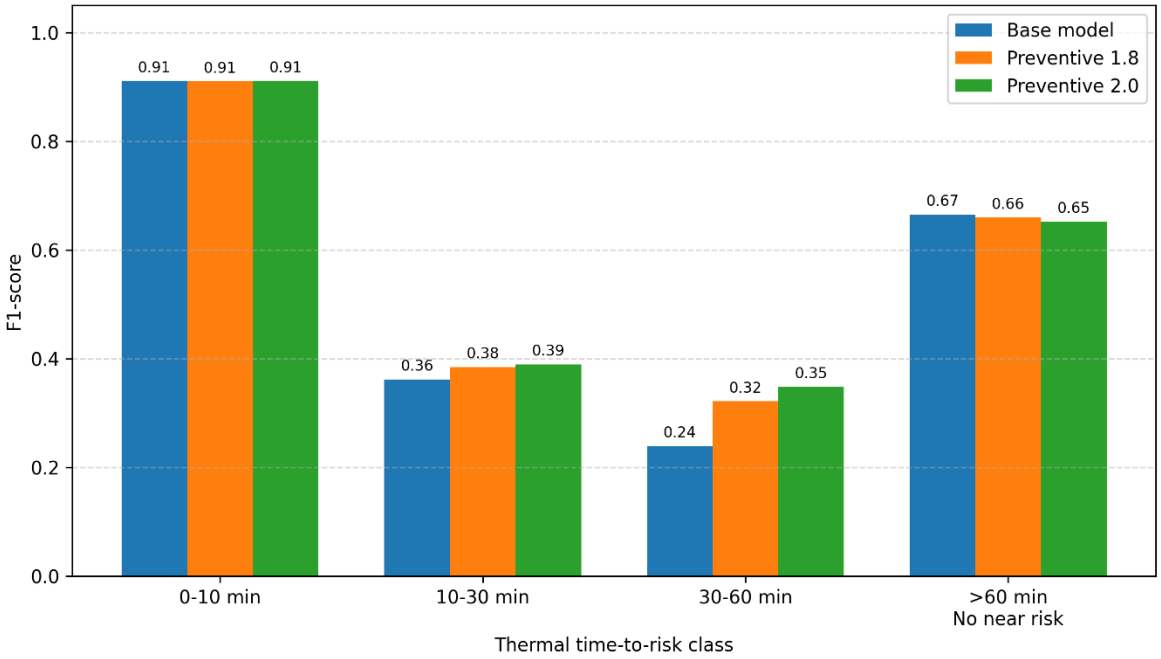

Figure 4. *Comparison of F1-scores per class between the base model and the preventive variant*

Figure 5 and Table 9 show the corresponding error trade-off. Correct >30–60 min predictions increased from 176 to 279 and 328, while >30–60 min samples assigned to >60 min / No near risk decreased from 601 to 482 and 472. Conversely, >60 min / No near risk samples assigned to >30–60 min increased from 265 to 371 and 436. The mean absolute ordinal class error decreased from 0.5204 to 0.5099 and 0.5069 classes. Thus, the increased preventive sensitivity was accompanied by more false preventive assignments.

Table 9. Operational error trade-off in the >30–60 min sensitivity analysis

Configuration	>30–60 correct	>30–60 → >60	>60 → >30–60	Mean absolute class error
Unweighted	176	601	265	0.5204
Preventive 1.8	279	482	371	0.5099
Preventive 2.0	328	472	436	0.5069

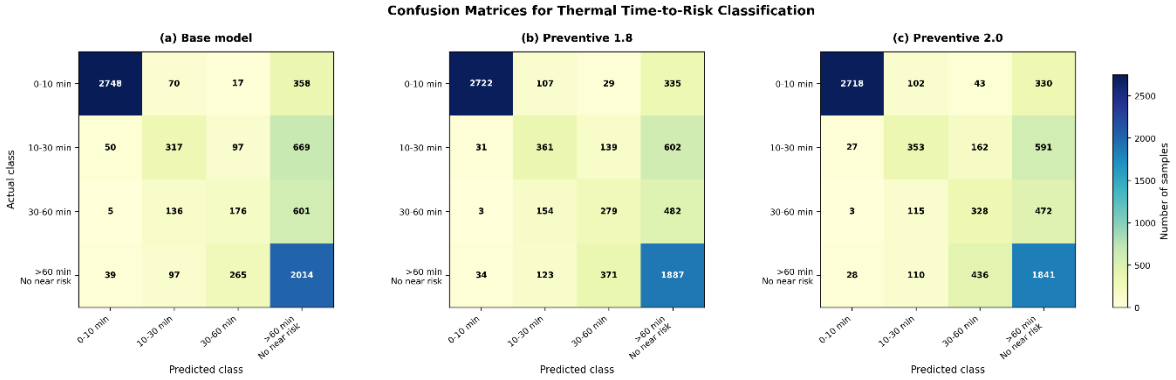

Figure 5*. Thermal time-to-risk confusion matrices for (a) the unweighted model, (b) the preventive 1.8 configuration, and (c) the preventive 2.0 configuration.*

3.4. Hypothesis-Oriented Interpretation

H1 was supported within the defined synthetic test conditions: relative to the deterministic baseline, the unweighted multi-output model reduced forecast MAE and RMSE and increased overall accuracy for thermal time-to-risk, thermal state, and associated zone. This comparison supports joint predictive performance, but it does not isolate the contribution of CNN1D, GRU, shared MLP, or multi-output training.

H2 was supported as a computational discrimination criterion: thermal time-to-risk accuracy (0.6861) exceeded the majority-class reference (0.4169), and every interval had a non-zero F1-score. Nevertheless, the >30–60 min interval remained weak (F1 = 0.2390), so the result does not establish equivalent performance across horizons or externally validated operational actionability.

H3 was supported descriptively: both weighted configurations increased recall and F1-score for >30–60 min and reduced its assignment to >60 min / No near risk, while global time-to-risk accuracy changed by less than 0.0020. The increase in reciprocal false preventive assignments confirms that the result is a sensitivity trade-off rather than global superiority.

All three hypothesis evaluations are based on one fixed, deterministic training run per configuration. Accordingly, the reported differences are descriptive for the frozen split and do not constitute statistical significance tests or estimates of variability across random initializations.

4. Conclusions

This study developed and evaluated, under controlled synthetic conditions, a hybrid multi-output CNN1D–GRU–MLP architecture for classifying interval-based thermal time-to-risk from 30-minute sequences representing the Front, Middle, and Door zones of a refrigerated compartment. The input construction represented a compartment-level thermal trajectory, zone-specific offsets, event-related perturbations, and stochastic sensor noise. Accordingly, this evaluation characterized computational multizone pattern modeling and did not constitute inference from measurements collected in an operational vehicle.

The implemented architecture comprised CNN1D layers, a GRU layer, a shared MLP, and four output heads. Local-feature extraction and temporal-sequence modeling were the design roles assigned to the convolutional and recurrent components, respectively. Because the experiment did not include component ablations, equivalent single-output controls, or a GRU-versus-LSTM comparison, it did not evaluate the individual contribution of each component, an advantage attributable to joint training, or superiority of the selected recurrent unit.

One methodological element of the study was the operationalization of thermal time-to-risk as four non-overlapping intervals: 0–10 min, >10–30 min, >30–60 min, and >60 min / No near risk. On synthetic test trips excluded from training and validation, the unweighted model achieved an overall accuracy of 0.6861, above the 0.4169 empirical majority-class reference, and produced non-zero F1-scores for all four intervals. These results met the predefined H2 computational-discrimination criterion on the fixed test partition; the study did not compare the actionability of these intervals with exact-minute estimates in an operational setting.

Across the three fixed training runs, the class-weight sensitivity analysis produced a specific descriptive trade-off rather than evidence of a statistically significant or globally superior configuration. Overall thermal time-to-risk accuracy changed from 0.6861 in the unweighted model to 0.6853 and 0.6842 with weights of 1.8 and 2.0, respectively. In parallel, the >30–60 min F1-score increased from 0.2390 to 0.3214 and 0.3476, and its assignments to >60 min / No near risk decreased from 601 to 482 and 472. The reciprocal >60 min / No near risk assignments to >30–60 min increased from 265 to 371 and 436. Thus, the higher training weight for the >30–60 min class was accompanied by improved class-specific detection and more reciprocal false preventive assignments on the fixed test partition.

On the same 7,659 test windows, the unweighted neural model produced lower forecast MAE and RMSE and higher overall accuracy for thermal time-to-risk, thermal-state, and associated-zone classification than the deterministic trend-based baseline. This descriptive comparison met the predefined H1 criterion for the fixed synthetic test partition. Because equivalent single-output controls and component ablations were not included, the comparison does not establish that the observed differences were caused by the shared representation, joint training, CNN1D layers, or GRU layer.

The evaluation used synthetic sequences produced with a simplified, physically motivated thermal balance for controlled neural-model development; it did not use a calibrated thermodynamic digital twin of a specific vehicle. The final test partition contained no cold-risk windows, and each configuration was evaluated through one fixed deterministic run. Consequently, the reported results characterize performance only for the defined synthetic test conditions and do not estimate variability across initializations, cold-risk detection, or operational performance. Future work should include repeated-seed evaluation, component and single-output ablations, broader overcooling scenarios, systematic cost-matrix analysis, and external validation using real multizone measurements from the Front, Middle, and Door zones of refrigerated vehicles.

References

Adefemi, K. O., Mutanga, M. B., & Alimi, O. A. (2025). A hybrid CNN–GRU deep learning model for IoT network intrusion detection. *Journal of Sensor and Actuator Networks, 14*(5), Article 96. https://doi.org/10.3390/jsan14050096

Artuso, P., Rossetti, A., Minetto, S., Marinetti, S., Moro, L., & Del Col, D. (2019). Dynamic modeling and thermal performance analysis of a refrigerated truck body during operation. *International Journal of Refrigeration, 99*, 288–299. https://doi.org/10.1016/j.ijrefrig.2018.12.014

Bavaresco, R., Arruda, H., Rocha, E., Barbosa, J., & Li, G.-P. (2021). Internet of Things and occupational well-being in Industry 4.0: A systematic mapping study and taxonomy. *Computers & Industrial Engineering, 161*, Article 107670. https://doi.org/10.1016/j.cie.2021.107670

Ben Taher, M. A., Kousksou, T., Zeraouli, Y., Ahachad, M., & Mahdaoui, M. (2021). Thermal performance investigation of door opening and closing processes in a refrigerated truck equipped with different phase change materials. *Journal of Energy Storage, 42*, Article 103097. https://doi.org/10.1016/j.est.2021.103097

Chen, K.-Y., & Shaw, Y.-C. (2011). Applying back propagation network to cold chain temperature monitoring. *Advanced Engineering Informatics, 25*(1), 11–22. https://doi.org/10.1016/j.aei.2010.05.003

Duan, F., Meng, X., Wu, W., Zou, Y., & Zeng, X. (2026). Data-driven virtual sensor systems for dynamic temperature monitoring along food supply chains. *Journal of Stored Products Research, 115*, Article 102844. https://doi.org/10.1016/j.jspr.2025.102844

U.S. Food and Drug Administration. (2023). *Grade "A" Pasteurized Milk Ordinance (PMO): 2023 revision*. U.S. Department of Health and Human Services.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hoang, H. M., Akerma, M., Mellouli, N., Le Montagner, A., Leducq, D., & Delahaye, A. (2021). Development of deep learning artificial neural networks models to predict temperature and power demand variation for demand response application in cold storage. *International Journal of Refrigeration, 131*, 857–873. https://doi.org/10.1016/j.ijrefrig.2021.07.029

Hu, J., Sari, O., Eicher, S., & Rakotozanakajy, A. R. (2009). Determination of specific heat of milk at different fat content between 1 °C and 59 °C using micro DSC. *Journal of Food Engineering, 90*(3), 395–399. https://doi.org/10.1016/j.jfoodeng.2008.07.009

International Organization for Standardization. (2009). *Milk—Determination of freezing point—Thermistor cryoscope method (Reference method) (ISO Standard No. 5764:2009).* ISO official record

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data, 6*, Article 27. https://doi.org/10.1186/s40537-019-0192-5

Khan, S. H., Hayat, M., Bennamoun, M., Sohel, F. A., & Togneri, R. (2018). Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems, 29*(8), 3573–3587. https://doi.org/10.1109/TNNLS.2017.2732482

KRESS Fahrzeugbau GmbH. (n.d.). *Refrigerated truck bodies using an IVECO Daily chassis—Our CoolerBox2.0 has a lot to offer*. Retrieved August 7, 2026, from KRESS Iveco Daily page

Maiorino, A., Petruzziello, F., & Aprea, C. (2021). Refrigerated transport: State of the art, technical issues, innovations and challenges for sustainability. *Energies, 14*(21), Article 7237. https://doi.org/10.3390/en14217237

Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX, 7*, Article 100777. https://doi.org/10.1016/j.mex.2019.100777

Mutambik, I. (2026). A blockchain–IoT–ML framework for sustainable vaccine cold chain management in pharmaceutical supply chains. *Systems, 14*(5), Article 467. https://doi.org/10.3390/systems14050467

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ, 372*, n160. https://doi.org/10.1136/bmj.n160

Slaghuis, B. A. (2001). The freezing point of authentic and original farm bulk tank milk in the Netherlands. *International Dairy Journal, 11*(3), 121–126. https://doi.org/10.1016/S0958-6946(01)00043-7

Thermo King. (n.d.). *V-600*. Retrieved August 7, 2026, from Thermo King V-600

United Nations Economic Commission for Europe. (2024). *ATP handbook 2023: To be used with the Agreement on the International Carriage of Perishable Foodstuffs and on the Special Equipment to be Used for Such Carriage (ATP) as amended on 22 June 2024*. United Nations. UNECE ATP Handbook 2023 PDF

United Nations Environment Programme, & Food and Agriculture Organization of the United Nations. (2022). *Sustainable food cold chains: Opportunities, challenges and the way forward*. https://doi.org/10.4060/cc0923en

Wang, Z., Dong, Y., Liu, W., & Ma, Z. (2020). A novel fault diagnosis approach for chillers based on 1-D convolutional neural network and gated recurrent unit. *Sensors, 20*(9), Article 2458. https://doi.org/10.3390/s20092458

Xie, Z., Long, H., Ling, C., Zhou, Y., & Luo, Y. (2025). An anomaly detection scheme for data stream in cold chain logistics. *PLOS ONE, 20*(3), e0315322. https://doi.org/10.1371/journal.pone.0315322

Zhang, K., Huo, X., & Shao, K. (2023). Temperature time series prediction model based on time series decomposition and Bi-LSTM network. *Mathematics, 11*(9), Article 2060. https://doi.org/10.3390/math11092060

Zhang, Y., Zou, Y., Liu, L., Manzardo, A., Wang, X., & Wu, J. (2026). A digital twin–driven dynamic shelf life approach for mitigating food loss across fruit supply chains. *Journal of Stored Products Research, 117*, Article 103016. https://doi.org/10.1016/j.jspr.2026.103016