UNIVERSITY OF WEST ATTICA

eISSN:3136-7658

Applied Manufacturing Intelligence Journal

Vol. 5, No. 3, July, 2026



Edge-Aware KAN-GraphMixer Forecasting Model for Abnormality Identification in Intelligent Logistics Cold Boxes

Norbert Brzeziński[1,*], Oliwier Majchrzak[1], Patryk Rybak[1]

[1] *Faculty of Computer Science and Information Technology, University of Economics in Wroclaw, Wroclaw, 50-231, Poland*

Received 07 April 2026; Revised 12 June 2026; Accepted 15 July 2026; Published 19 July 2026

Abstract

Abnormality identification in intelligent logistics cold boxes requires reliable interpretation of coupled sensor signals under transient handling, uneven thermal response, and intermittent communication. This study proposes an edge-aware KAN-GraphMixer forecasting model that combines nonlinear channel transformations with lightweight sensor-graph mixing. A Kolmogorov-Arnold network represents channel-specific response functions, while graph mixing propagates information among thermally related sensing locations. Multi-step forecasts are converted into node-level residuals, and abnormality decisions incorporate residual magnitude, persistence, and observation confidence. An illustrative evaluation configuration comprising 42,600 observation windows covers normal transport, short and prolonged door exposure, delayed pre-cooling, refrigeration interruption, and sensor drift. The reported configuration achieves a 15-min temperature forecasting mean absolute error of 0.286 °C and a root mean square error of 0.438 °C. Abnormality identification reaches 97.5% precision, 96.1% recall, and a 96.8% F1-score. Compared with the graph-mixing baseline, forecasting error decreases by approximately 10.1%, and F1-score increases by 2.4 percentage points. At a false alarm rate of 1.9%, the median warning lead time is 11.6 min. Local inference requires 8.7 ms per window, with a quantized parameter footprint of 0.84 MB. Component comparisons indicate that both nonlinear transformation and graph interaction contribute to performance, while packet loss and shared calibration reduce identification accuracy. The numerical findings describe a proof-of-concept configuration whose practical effectiveness requires independently collected operational records, reproducible training settings, and gateway profiling. They indicate potential for compact predictive monitoring within the stated sensing and calibration conditions.

Keywords: KAN-GraphMixer, Cold Chain Logistics, Edge Anomaly Identification, Temperature Forecasting, Sensor Graph Learning

1. Introduction

Temperature control and digital data exchange have been examined across food cold chains, providing a practical context for monitoring reusable cold boxes [1]. Integrated sensing and logging systems have been developed to support cold-chain monitoring applications [2]. Internet-of-Things-supported supervision has also been investigated for agricultural cold-chain logistics [3]. Frozen-food monitoring research includes compact chipless radio-frequency identification sensors as a complementary sensing approach [4]. These research directions become particularly relevant when a reusable box moves between sorting centers, line-haul vehicles, urban depots, and last-mile handover points. Assessing cargo risk from lid, wall, payload, ambient, humidity, vibration, and battery measurements requires interpretation of their joint thermal and temporal behavior. A short lid opening can produce rapid warming near the upper wall while the payload center remains stable. A similar opening after inadequate pre-cooling can create a different trajectory because the available thermal margin is smaller. The monitoring task therefore requires an interpretation of

* Corresponding author: norbert.br@ue.wroc.pl

https://doi.org/10.68406/amij.2026.vol5iss3np2:20-35

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

how the channels evolve together and how much confidence can be assigned to observations received during handling. Wireless interruptions introduce an additional difficulty: the interval in which local interpretation is most useful may coincide with delayed or incomplete transmission to the fleet platform. A practical forecasting model must retain enough spatial and temporal information to recognize deterioration while operating within the resources of the box gateway.

Existing abnormality identification methods in cold-chain monitoring usually follow threshold rules, statistical residuals, or sequence learning. Threshold rules provide a clear decision boundary, while predicting transient persistence requires temporal information. Statistical residual methods introduce an expected trajectory, although their behavior depends on assumptions about changing thermal response and measurement noise. Quantized autoencoder-based anomaly detection has been investigated for multivariate time series in communication networks, offering a cross-domain example of compact predictive processing [5]. Research on apple cooling in large bins examines heat transfer in a refrigerated product system and provides related thermal context [6]. Kolmogorov-Arnold network models have been investigated for nonlinear model predictive control [7]. This perspective is relevant to a cold box because lid, wall, payload, and ambient probes respond to the same disturbance at different rates. However, the usefulness of a relational representation must be assessed alongside its computational cost. Temporal Kolmogorov-Arnold networks have also been studied for stock-price forecasting, demonstrating interest in their application to sequence prediction outside cold-chain monitoring [8]. For a small sensor graph, attention complexity must be evaluated against its contribution to distinguishing normal handling from abnormal cross-channel responses. The unresolved problem is to combine nonlinear channel response, physically meaningful sensor interaction, and reliable inference from incomplete windows in a compact predictive model.

The present paper addresses this gap by developing an edge-aware KAN-GraphMixer forecasting model for abnormality identification in intelligent logistics cold boxes. The term edge-aware has two meanings in this study. First, the model is designed for local inference under bounded latency, memory, and packet availability, through compact graph mixing. Second, the method is aware of graph edges among sensing points, cargo zones, and handling events, so that residuals reflect coupled thermal behavior. A contemporary survey of green edge artificial intelligence provides a broader context for resource-constrained model design [9]. KAN channel transformations are motivated by the different response shapes of humidity, wall temperature, and payload temperature during exposure. The forecasting residual is converted into an abnormality score and calibrated against event persistence and observation confidence. Fully quantized lightweight object detection provides a separate application example of neural deployment on ultra-low-power edge systems [10]. The proposed formulation preserves node-level residual information until the decision stage, allowing a disturbance near the lid to be evaluated differently from a coordinated change across the return-air and payload channels. This design evaluates model quality through both forecasting accuracy and operational decision performance. The contribution lies in the cold-box-specific integration of node-preserving nonlinear transformations, event-conditioned graph mixing, and confidence-aware residual confirmation. The individual computational operators remain established modeling components, and the reported comparison assesses this integration within the illustrative configuration.

To connect the cold-box monitoring problem with a deployable predictive method, the paper proceeds from the limitations of existing approaches to the sensing representation, model design, and experimental assessment. Section 2 reviews cold-chain abnormality identification, inference on embedded devices, relational forecasting, and nonlinear function approximation, with attention to the assumptions that restrict their combined use. Section 3 defines the sensing graph and the observation window, describes how forecasting residuals become an abnormality score, and specifies the confidence and resource constraints governing local decisions. Section 4 develops the KAN-GraphMixer architecture, its channel transformations and graph mixing operations, and the training and quantization procedures used for edge execution. Section 5 examines the experimental configuration and comparative performance, then investigates sensor-specific errors, forecast horizons, probability calibration, packet loss, latency, resource use, and component removal. The same section analyzes residual contributions and difficult event cases to establish where the predictive decision remains sensitive to incomplete context. Section 6 states the conclusions, acknowledges the limits of the evaluation, and identifies directions for adaptation across payload arrangements and larger cold-chain fleets. This organization keeps the physical monitoring question, the computational design choices, and the interpretation of reported performance connected throughout the manuscript.

2. Related Work

Cold-chain abnormality identification has evolved from compliance-oriented recording toward predictive supervision. Early systems mainly used threshold alarms and manual audit trails, which remain useful when the monitored object has a clear regulatory temperature boundary. Conformal prediction has been investigated for uncertainty handling in online alarm-flood classification [11]. Several studies have therefore adopted statistical process control, Kalman filtering, and residual diagnosis to detect deviations from expected thermal trajectories. Wavelet- and Fourier-based multivariate time-series anomaly detection provides an additional example of data-driven representation for changing signal behavior [12]. Recent work has introduced data-driven models for route-conditioned temperature prediction, in which abnormality is identified when observed values diverge from predicted thermal behavior. Prediction-based monitoring can provide

warning lead time from sensor observations. Its suitability for reusable boxes depends on generalization across packing conditions and sensor layouts.

Edge intelligence provides another important direction for cold-chain monitoring. In a cloud-only architecture, raw sensor data are transmitted to a central platform for inference, storage, and dashboard visualization. A survey of adversarial-learning-based multivariate time-series imputation addresses the separate problem of incomplete observations [13]. Edge inference places part of the decision process on the box gateway or nearby handheld terminal, reducing data traffic and allowing local alarms when communication is delayed. The engineering challenge is that edge devices usually have limited processing capacity, limited memory, and strict battery constraints. Adaptive industrial alarm design offers a related perspective on selecting decision policies as operating conditions change [14]. For cold boxes, the edge model must maintain stable alarm behavior under irregular sampling, short missing segments, and asynchronous sensor updates.

Graph neural networks and mixer-style architectures provide modeling methods for modeling coupled sensing systems. Prediction of uncertainty reduction in online alarm-flood classification connects available evidence with the timing of an operational decision [15]. However, graph attention may be unnecessarily expensive for the small but dense sensing topology of a cold box, and learned adjacency matrices can become unstable when the number of labeled abnormal events is limited. Mixer-style networks reduce computational overhead by separating token mixing and channel mixing through simple matrix operations, supporting their use in edge deployment. Research on human-system interfaces examines how support tools assist the interpretation of monitoring information [16]. Combining KAN functions with lightweight graph mixing offers an approach to cold-box abnormality identification whose applicability requires further assessment. These adjacent research directions motivate uncertainty handling and operator support; cold-box-specific performance remains a separate empirical question. The present literature coverage supplies application context, while a comprehensive architectural comparison would require additional direct studies of graph mixers and nonlinear channel transformations.

3. Edge Aware Problem Formulation

3.1. Cold Box Sensing Graph

An intelligent cold box is represented as a compact sensing graph that explicitly preserves relationships among measurement channels. The graph includes temperature measurements near the lid, upper wall, payload center, return-air path, side wall, and ambient boundary, together with auxiliary observations of humidity, vibration, lid state, and battery condition. Each node retains a recent observation window, while its connections describe thermal coupling, physical proximity, or the influence of handling events. After a lid opening, the lid-state indicator increases the weights assigned to connections involving upper-wall temperature and humidity. The connection between ambient temperature and payload-center temperature remains comparatively weak because the payload responds more slowly. Figure 1 illustrates how these observations enter the local forecasting and abnormality identification workflow. The nine-node representation used for residual reporting comprises lid, upper-wall, payload-center, return-air, side-wall, ambient-temperature, humidity, vibration, and battery channels. Lid state is an event-conditioning attribute supplied to the graph update and input embedding. This convention distinguishes event metadata from residual-bearing sensing nodes.

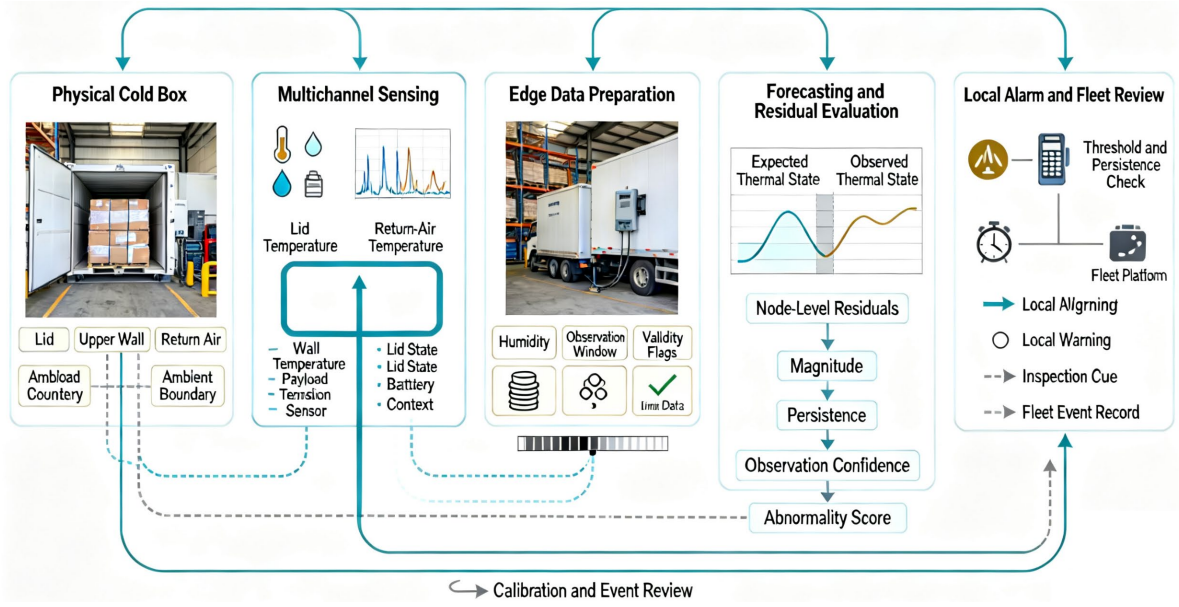


Figure 1. Edge sensing workflow for intelligent logistics cold-box abnormality identification

The graph preserves the location and propagation of a disturbance. During door exposure, the upper wall may warm before the payload center; during refrigeration interruption, the return-air path may change before a substantial payload response becomes visible. A flat feature representation can obscure these differences by combining observations before their physical relationships have been evaluated. The proposed representation retains node identity and allows event attributes to modify a limited set of connections. Edge weights are initialized from sensing positions and adjusted through lightweight coefficients. A lid-state transition strengthens the connections associated with exposure, whereas a vibration burst identifies a handling interval whose measurements may require additional scrutiny. These updates adjust the interpretation of observations while preserving the graph dimension. The graph specification currently states its size and physical rationale. Exact edge endpoints, update coefficients, self-connection handling, and weight normalization require a reproducible adjacency specification; these details remain implementation requirements for independent replication.

Let $X_t \in \mathbb{R}^{N\times C}$ denote the observation matrix at sampling time t, where N is the number of sensing nodes and C is the number of features associated with each node. The adjacency matrix $A_t \in \mathbb{R}^{N\times N}$ describes the current connection weights. Physical layout provides its initial structure, while handling indicators adjust selected entries. The future thermal state is denoted by Y_{t+h}, with the horizon h chosen according to the required warning interval.

3.2. Forecasting Residual and Abnormality Score

Abnormality identification begins with a forecast of the thermal state expected under the current observations and handling context. The model learns ordinary thermal evolution from normal transport and mild transitions, then evaluates deviations when the corresponding future measurements become available. This approach reduces exclusive reliance on rare abnormality labels. The alarm decision has a measurable basis in the deviation between observed behavior and the expected thermal trajectory. Direct classification can place excessive emphasis on frequent door-opening events while providing weaker discrimination for delayed pre-cooling or gradual sensor drift. Forecasting provides a common reference against which these different disturbances can be examined. At any decision time, a realized residual compares the current observation with a forecast issued one horizon earlier. The newly issued forecast concerns a later target and can support a separate prospective risk estimate. Warning lead time is measured against a subsequent observed event boundary, with decision and boundary timestamps recorded separately.

The forecasting equation maps a window of L observations, the current sensing graph, and locally available context to the horizon target:

$$\hat{Y}_{t+h} = F_\theta(X_{t-L+1:t}, A_t, u_t) \tag{1}$$

Here, F_θ is the forecasting model; its parameters and input observation sequence are $\theta, X_{t-L+1:t}$, respectively, and u_t contains route and handling attributes. When the horizon is reached, the new measurement is compared with the stored forecast. The residual equation retains channel-specific deviations:

$$r_{t+h} = W_r\left|Y_{t+h} - \hat{Y}_{t+h}\right| \tag{2}$$

The absolute value is applied elementwise. The diagonal weighting matrix W_r assigns importance according to the sensing location and its relevance to thermal risk. Payload-center, upper-wall, and auxiliary deviations receive weights corresponding to their respective relevance to thermal risk. Preserving the residual vector also allows a localized lid disturbance to be distinguished from a coordinated change across the return-air, side-wall, and payload nodes.

Alarm confirmation depends jointly on residual magnitude, persistence, and observation confidence. A short residual peak during handling differs from a sustained departure, and incomplete observations can alter confidence in either interpretation. These factors enter the abnormality-score equation:

$$s_t = \sigma(\alpha\|r_t\|_1 + \beta d_t - \gamma c_t) \tag{3}$$

In this equation, σ is the sigmoid function, d_t measures residual persistence, and c_t represents confidence in the available observations and graph state. The coefficients α, β, and γ control their respective contributions. Because reduced confidence raises the unadjusted score through the negative confidence term, missing-data handling requires joint calibration of the confidence contribution and the alarm threshold to control false alarms. An abnormal event is confirmed only when the score exceeds the calibrated threshold and the residual satisfies the persistence requirement. This separates the detection of a deviation from its confirmation as an operational event. Confidence describes observation freshness and validity. Each observation-quality condition requires validation of the score and threshold together. Evidence of false-alarm suppression depends on that validation. Confidence construction, score coefficients, and the persistence interval remain necessary calibration metadata.

3.3. Edge Constraints

The gateway imposes computational and observational constraints on the identification process. It must complete inference before the next processing cycle, retain recent measurements within limited memory, and produce stable decisions when individual channels are delayed or missing. The design targets an inference latency below 12 ms and a quantized parameter footprint below 1.0 MB. Packet-loss robustness is evaluated separately through changes in identification performance. These targets define the intended deployment envelope, with performance assessed for each gateway and sensing configuration. Resource feasibility must therefore be assessed using the complete inference path, including graph updates, input preparation, and score computation.

Local evidence also differs from the information available to a central fleet platform. A cloud service can incorporate inspection records received after a route, whereas a gateway must act on observations available at the decision time. The formulation consequently separates continuous forecasting, residual scoring, and alarm confirmation. Forecasting is updated with each observation window. Scoring evaluates the available deviation together with persistence and confidence. Confirmation requires sufficient temporal evidence to prevent a single corrupted packet from producing an operational alarm. Missing observations are represented through validity indicators so that an imputed value is distinguishable from a fresh measurement.

The local output includes both a risk score and a residual-based inspection cue. A wall-dominant deviation may direct attention toward lid sealing or ambient exposure, whereas an isolated channel deviation may justify checking measurement quality. These cues localize the supporting evidence, while causal fault attribution requires additional inspection. Retaining validity indicators with the alarm record is particularly important when a disturbance coincides with communication loss. The reliability of an explanation depends on the completeness and consistency of the evidence available to the gateway.

4. KAN GraphMixer Forecasting Model

4.1. Architecture Overview

The proposed model comprises graph-aware input embedding, a KAN channel transformation, lightweight graph mixing, and a forecasting head followed by residual scoring. Input embedding maps each node window into a compact representation and incorporates indicators of lid state, route dwell, ambient exposure, and observation validity. The KAN transformation models nonlinear channel response through learnable univariate functions. Graph mixing then exchanges information across connected sensing nodes using the current adjacency matrix. The forecasting head produces the future thermal state required by the residual pathway. Figure 2 presents the relationship between these stages.

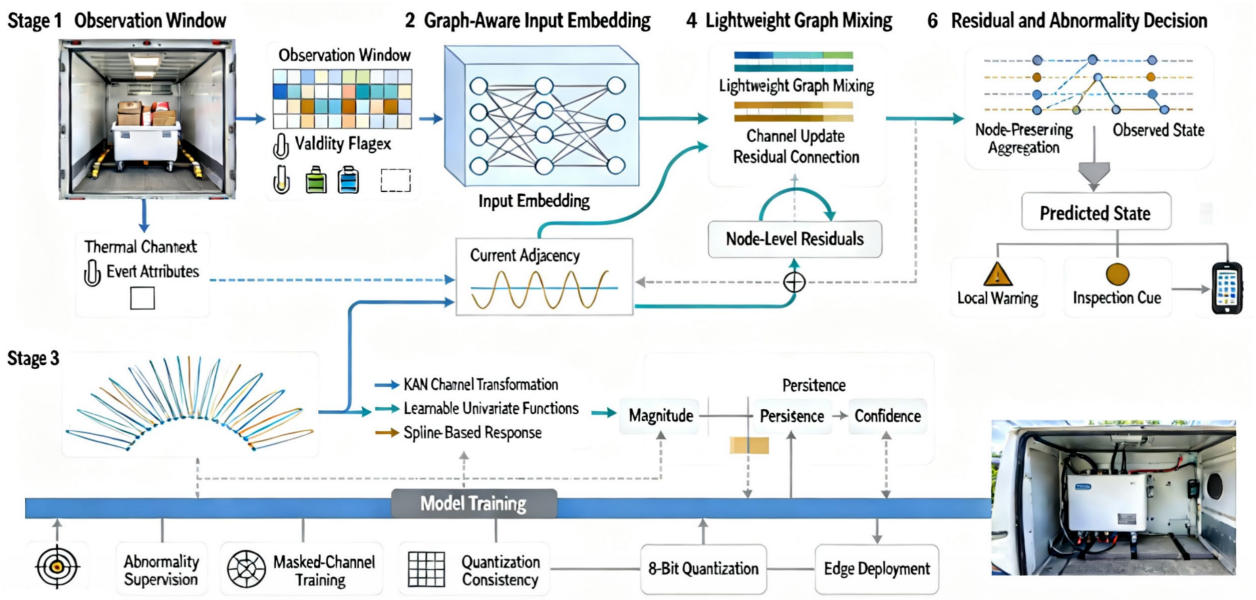


Figure 2. Architecture of the edge-aware KAN-GraphMixer forecasting model

The architecture uses the partially known sensing topology to constrain relational computation. Thermal connections are influenced by physical layout and handling events, whereas nonlinear response arises within the individual channels and their interaction with recent thermal history. Separate operators model nonlinear channel response and graph-based information exchange using the available sensing topology. The graph mixer propagates evidence between related locations, while the KAN functions transform the incoming features before that evidence is combined. This distinction

is relevant during warm stops, when the wall response may become visible well before the payload-center temperature changes. A mixer state has one row per sensing node and one column per latent feature. The adjacency operator acts on the node axis, while the low-rank operators act on the latent-feature axis; every additive branch must preserve the state shape. Rank eight specifies a factorization constraint, and the latent width and number of blocks still require an explicit configuration record.

For node i, the channel transformation follows the KAN equation:

$$z_i = \sum_{k=1}^{C} \phi_{ik}(x_{ik}) + b_i \tag{4}$$

The function ϕ_{ik} is represented by a compact spline basis, and b_i is a learnable bias. Applied componentwise, the transformation produces the latent representation used by subsequent mixer blocks. Different channel functions can represent different response shapes, allowing humidity and payload-temperature transitions to retain their distinct nonlinear responses. The scalar expression describes one output coordinate; a latent vector uses separately indexed functions for its output coordinates. Eight control points specify the reported spline capacity. Spline degree, knot placement, boundary extrapolation, and feature scaling determine the realized transformation and remain required implementation details.

Information exchange is governed by the graph-mixing equation, which combines a graph-dependent update with a channel update and a residual connection:

$$H^{(\ell+1)} = H^{(\ell)} + \rho\left(A_t H^{(\ell)} P_\ell\right) + H^{(\ell)} Q_\ell \tag{5}$$

Here, $H^{(\ell)}$ is the node representation at block ℓ, while P_ℓ and Q_ℓ are dimension-compatible learnable operators implemented through low-rank factors. The gated activation ρ controls the graph-dependent contribution. The residual connection retains the incoming representation, so evidence from an individual node remains available after neighboring information has been introduced.

After K mixer blocks, the forecasting-head equation converts the retained node representations and context into a thermal prediction:

$$\hat{Y}_{t+h} = G_\omega\left(Pool\left(H^{(K)}\right), u_t\right) \tag{6}$$

The head G_ω has parameters ω. The pooling operator aggregates temporal or feature information while preserving the node identity needed for node-specific output. The resulting representation retains the sensing-node distinctions required for residual localization. The predicted thermal vector can subsequently be compared with the observed vector to obtain the location-specific residuals used by the identification stage.

4.2. Edge Aware Training and Inference

Training balances forecasting accuracy, abnormality discrimination, and sensitivity to quantization. The forecasting term uses windows with available future thermal observations, including windows without event annotations. The abnormality term is evaluated where event labels are available. Random channel masking and packet-delay simulation introduce incomplete observations during training, while validity indicators identify the affected inputs. This procedure exposes the model to both complete sequences and changes in observation quality. Threshold selection remains a validation-stage decision so that the deployed alarm policy can be assessed independently of parameter fitting. Normalization statistics should be estimated from training observations or a documented calibration period preceding test operation. The forecasting subset and annotated classification subset require separate masks, because target availability and event-label availability describe different supervision conditions. The present description specifies this separation at the formulation level.

The training objective is specified by the composite-loss equation:

$$\mathcal{L} = \lambda_1 MAE(Y, \hat{Y}) + \lambda_2 BCE(a, \hat{a}) + \lambda_3 \left\| \Theta_q - \Theta \right\|_2 \tag{7}$$

The first term measures thermal forecasting error, and the second measures binary abnormality discrimination. Here, a denotes the annotated event label and $\hat{a}$ the predicted abnormality probability. The final term penalizes parameter drift between the full-precision parameters Θ and their dequantized deployment approximation Θ_q. The weights λ_1, λ_2, and λ_3 regulate the balance between these objectives. Deployment evaluation measures forecasting error and identification performance using the quantized model to assess the predictive consequences of parameter approximation. Independent reconstruction requires numerical loss weights, the optimizer, learning-rate schedule, stopping rule, and seed policy. These settings still require explicit specification. The objective therefore defines a model family whose exact training procedure requires the original configuration record.

The trained model is exported with 8-bit weights for execution on the ARM gateway. Inference uses a sliding observation window and retains the latest graph state in memory. A missing packet is represented by an imputed observation accompanied by a validity flag that identifies its acquisition status. The associated confidence is reduced, and the threshold policy is calibrated for that observation condition. Joint validation of the score and confidence adjustment evaluates the balance between false-alarm suppression and timely event detection. Persistence is checked after scoring, allowing a short disturbance to be distinguished from a sustained deviation before an alarm is confirmed.

4.3. Implementation Details

The implementation uses a 12-step observation window and a 15-min forecasting horizon. Temperature, humidity, lid-state, vibration, battery, and ambient variables are normalized using box-specific calibration statistics. The sensing graph contains nine nodes and twenty-six directed connections after event-related expansion. The KAN block uses eight spline control points per channel, and the mixer operators use rank-eight factors. The reported parameter footprint is 0.91 MB before quantization and 0.84 MB afterward. These settings define the illustrative configuration assessed in the experimental chapter. The inference routine returns a predicted thermal vector, a node-specific residual vector, an abnormality score, and an inspection cue for the operator interface.

Missing and delayed samples are handled within a fixed graph dimension. If a channel misses one sampling step, the gateway retains its latest valid observation and marks the input as imputed. When consecutive observations are unavailable, the node remains in the graph but receives reduced confidence. Retaining the node preserves the input shape and avoids rebuilding the deployed model around changing sensor availability. Validity indicators enter the channel transformation so that the model can distinguish a stable measurement from a repeated value caused by packet loss. During quantization calibration, the validation subset includes both complete and corrupted windows to expose differences between ideal sensing and the observation conditions expected during transport.

The local record also preserves the information required to review a decision. Alongside the score, it stores the sensing validity state and the channels contributing most strongly to the residual. This enables an inspection cue to be interpreted in relation to the evidence available when the alarm was raised. A return-air deviation supported by neighboring thermal channels warrants a different review from a deviation associated with stale ambient measurements. When observations provide limited support for an explanation, the output records the associated diagnostic uncertainty. The inspection cue is an evidence-localization output. A causal diagnosis requires corroborating inspection or measurement-quality evidence, and a low-confidence record should retain its uncertainty category for operator review.

5. Experimental Data and Analysis

5.1. Dataset and Experimental Configuration

Adaptive temperature scaling addresses probability calibration in neural models and supplies calibration-related methodological context [17]. The illustrative evaluation specifies separation by route group to reduce overlap between temporally related training and test observations. The illustrative dataset specification contains 42,600 synchronized windows, including 27,800 normal windows and 14,800 abnormal windows. Each window contains nine sensing nodes, twelve recent sampling steps, and attributes describing ambient temperature, dwell time, handling mode, and battery level. Five event classes are represented: short door exposure, prolonged door exposure, delayed pre-cooling, refrigeration interruption, and slow sensor drift. Baselines comprise ARIMA residual monitoring, LSTM, a temporal convolution network, Transformer, Graph WaveNet, and GraphMixer without KAN transformations. The alarm threshold is selected using validation observations and held fixed during test evaluation. Temperature MAE and RMSE describe forecast quality; precision, recall, F1-score, false alarm rate, and lead time describe abnormality identification. The numerical comparison is reported in Table 1. The numerical values are illustrative configuration results. Operational verification requires raw acquisition files, collection dates, event adjudication records, and source-to-window mappings, which remain outside the documented specification. Exact split counts and route membership require further documentation. Reproducible evaluation requires these records, baseline search spaces, common preprocessing, and repeated-run uncertainty estimates. The tabulated point estimates support a descriptive comparison within the stated configuration.

Table 1. Comparative performance on the cold-box abnormality identification test set

Model	MAE (*deg C*)	RMSE (*deg C*)	Precision (*%*)	Recall (*%*)	F1 (*%*)	Latency (*ms*)
ARIMA residual	0.526	0.742	86.4	82.7	84.5	2.1
LSTM	0.412	0.604	91.8	89.2	90.5	7.9
TCN	0.389	0.566	92.6	90.1	91.3	6.8

Transformer	0.365	0.541	93.3	90.7	92.0	15.6
Graph WaveNet	0.337	0.506	94.5	92.9	93.7	12.4
GraphMixer	0.318	0.481	95.1	93.8	94.4	9.5
Proposed KAN-GraphMixer	0.286	0.438	97.5	96.1	96.8	8.7

The proposed model has a 15-min MAE of 0.286 degrees C and an RMSE of 0.438 degrees C. Relative to GraphMixer, these values decrease by 0.032 and 0.043 degrees C, respectively. The MAE reduction is approximately 10.1%, whereas the F1-score increases by 2.4 percentage points, from 94.4% to 96.8%. SensorDBSCAN provides a related example of sensor-based anomaly detection and diagnosis [18]. Inference latency is 8.7 ms for the proposed model, 9.5 ms for GraphMixer, and 2.1 ms for ARIMA residual monitoring. ARIMA reaches an F1-score of 84.5%, indicating a trade-off between inference cost and identification performance in this configuration.

Sampling irregularity is retained in the evaluation because removing incomplete windows would change the operating conditions represented by the dataset. The illustrative configuration assigns timestamp-gap rates of 3.8% and 1.1% to its field-like and laboratory-like subsets, respectively, with a median interval deviation of 4.2 s after alignment. Channel-validity flags distinguish fresh observations from imputed values. The four ambient bands used for the graphical comparison have representative temperatures of 5, 15, 25, and 35 degrees C. Two payload masses are included to expose differences in thermal inertia. At the plotted ambient endpoints of 5 and 35 degrees C, the MAE reductions relative to GraphMixer are 0.05 and 0.06 degrees C, respectively. These differences are computed from the displayed band values and refer specifically to those representative temperatures.

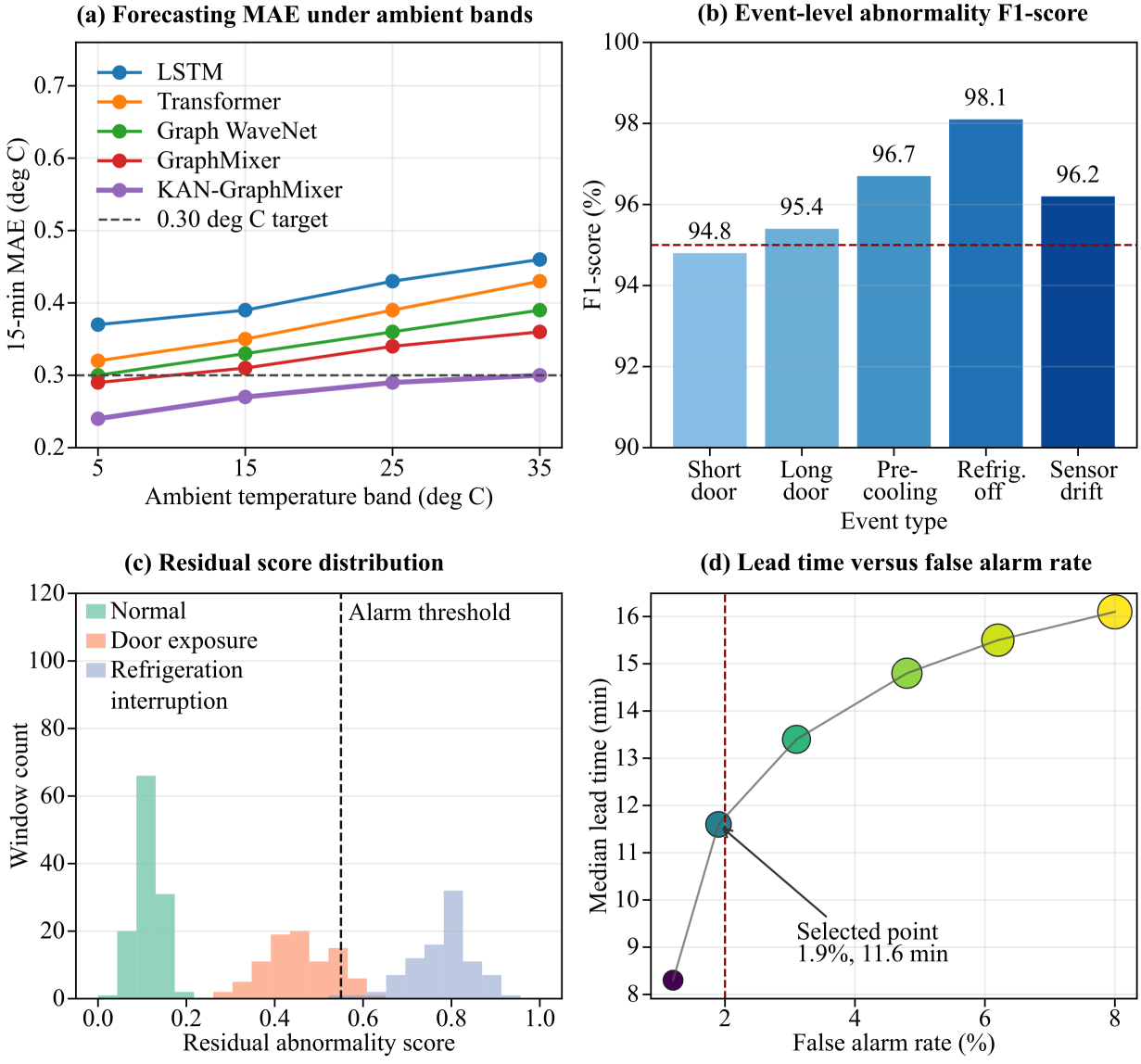


Figure 3. Forecasting accuracy and abnormality identification performance: (a) ambient-band MAE; (b) event-specific F1-score; (c) residual-score distributions; (d) lead-time and false-alarm operating points

Figure 3(a) compares five neural models across the four ambient bands. KAN-GraphMixer MAE rises from 0.24 to 0.30 degrees C, while Transformer increases from 0.32 to 0.43 degrees C and GraphMixer from 0.29 to 0.36 degrees C. The proposed error remains at or below the 0.30-degree target across all four ambient bands. The increase across the ambient range is 0.06 degrees C for the proposed model and 0.11 degrees C for Transformer. Figure 3(b) reports F1-scores of 94.8%, 95.4%, 96.7%, 98.1%, and 96.2% for short door exposure, prolonged door exposure, delayed pre-cooling, refrigeration interruption, and sensor drift. The 3.3-percentage-point difference between short door exposure and refrigeration interruption indicates that event duration and cross-node coherence affect recognition even when the overall score is high.

Figure 3(c) displays residual-score distributions centered near 0.12 for normal windows, 0.46 for door exposure, and 0.79 for refrigeration interruption. The 0.55 threshold lies above the center of the door-exposure distribution but below the center of the interruption distribution. Door events that end before the persistence requirement is met may produce residual deviations below the alarm-confirmation criterion. Graph-convolutional anomaly detection with node and line sensor data offers a related multichannel example from smart grids [19]. The distribution centers describe score separation, while the tails identify windows whose classification depends on additional temporal evidence under the persistence rules.

Figure 3(d) examines six operating points with false alarm rates of 1.2%, 1.9%, 3.1%, 4.8%, 6.2%, and 8.0%. Their corresponding median lead times are 8.3, 11.6, 13.4, 14.8, 15.5, and 16.1 min. Moving from 1.2% to the selected 1.9% point adds 3.3 min of warning, whereas increasing the false alarm rate from 6.2% to 8.0% adds only 0.6 min. At the selected point, the aggregate median warning interval is 11.6 min, with intervention time varying across event types. The refrigeration-interruption median is 13.2 min, with a 25th percentile of 9.5 min, while delayed pre-cooling and short door exposure have medians of 10.4 and 6.8 min. The selected point represents one validation-chosen alarm policy within the illustrative comparison. Operational selection requires the costs of missed events and unnecessary interventions, together with a precise event-boundary definition. Reported lead time measures the interval from alarm activation to that boundary; the forecast horizon specifies the separate interval between forecast issuance and its prediction target.

5.2. Forecasting and Identification Results

Figure 4(a) reports sensor-specific MAE for five temperature locations. KAN-GraphMixer errors are 0.331, 0.302, 0.241, 0.287, and 0.355 degrees C at the lid, upper wall, payload, return-air path, and ambient boundary, compared with 0.381, 0.352, 0.291, 0.332, and 0.402 degrees C for GraphMixer. The absolute reductions range from 0.045 to 0.050 degrees C. Their relative significance differs: the payload reduction is approximately 17.2%, whereas the ambient reduction is approximately 11.7%. Figure 4(b) extends the comparison over six horizons from 5 to 30 min. The proposed RMSE grows from 0.213 to 0.492 degrees C, compared with 0.239 to 0.571 degrees C for GraphMixer and 0.276 to 0.681 degrees C for Transformer. At 30 min, the proposed advantage is 0.079 degrees C over GraphMixer and 0.189 degrees C over Transformer, indicating that the residual quality depends on the prediction horizon used by the alarm policy.

Graph-neural-network detection of malicious Internet-of-Things communication illustrates anomaly identification in a different sensing domain [20]. Probability calibration separately evaluates agreement between predicted event likelihood and observed event frequency. Figure 4(c) contains ten calibration points; observed frequencies increase from 0.04 at a predicted probability of 0.05 to 0.94 at 0.95. At intermediate points, a prediction of 0.35 corresponds to a frequency of 0.31, and a prediction of 0.75 corresponds to 0.77. The reported expected calibration error is 0.024 for the proposed model and 0.061 for Graph WaveNet. Calculating these aggregate values requires the evaluation bin counts and sample weights in addition to the plotted calibration coordinates. The two calibration-error values remain illustrative reported aggregates. The displayed coordinates support visual comparison. Independent verification of the aggregates requires bin definitions, bin populations, and the underlying probability-label pairs, which remain to be supplied.

Figure 4(d) relates residual magnitude to ambient exposure duration for 160 windows. The residual trend increases with exposure, with a plotted reference level of 0.55 and a duration reference of 8 min. Overlap near these guides identifies the 8-min duration line as an evaluation reference whose risk interpretation depends on thermal and operating conditions. Two windows with similar exposure may have different residuals because payload inertia and the initial cooling state differ. Time-series cross-validation research examines promises and pitfalls in model selection, motivating careful separation of temporally related evaluation observations [21]. In this evaluation, the selected policy yields 97.5% precision and 96.1% recall, with a 1.9% false alarm rate, compared with 4.8% for Graph WaveNet and 6.2% for Transformer. During short exposure, a strong lid residual may coexist with a stable payload residual, requiring joint consideration of alarm sensitivity and payload-level persistence.

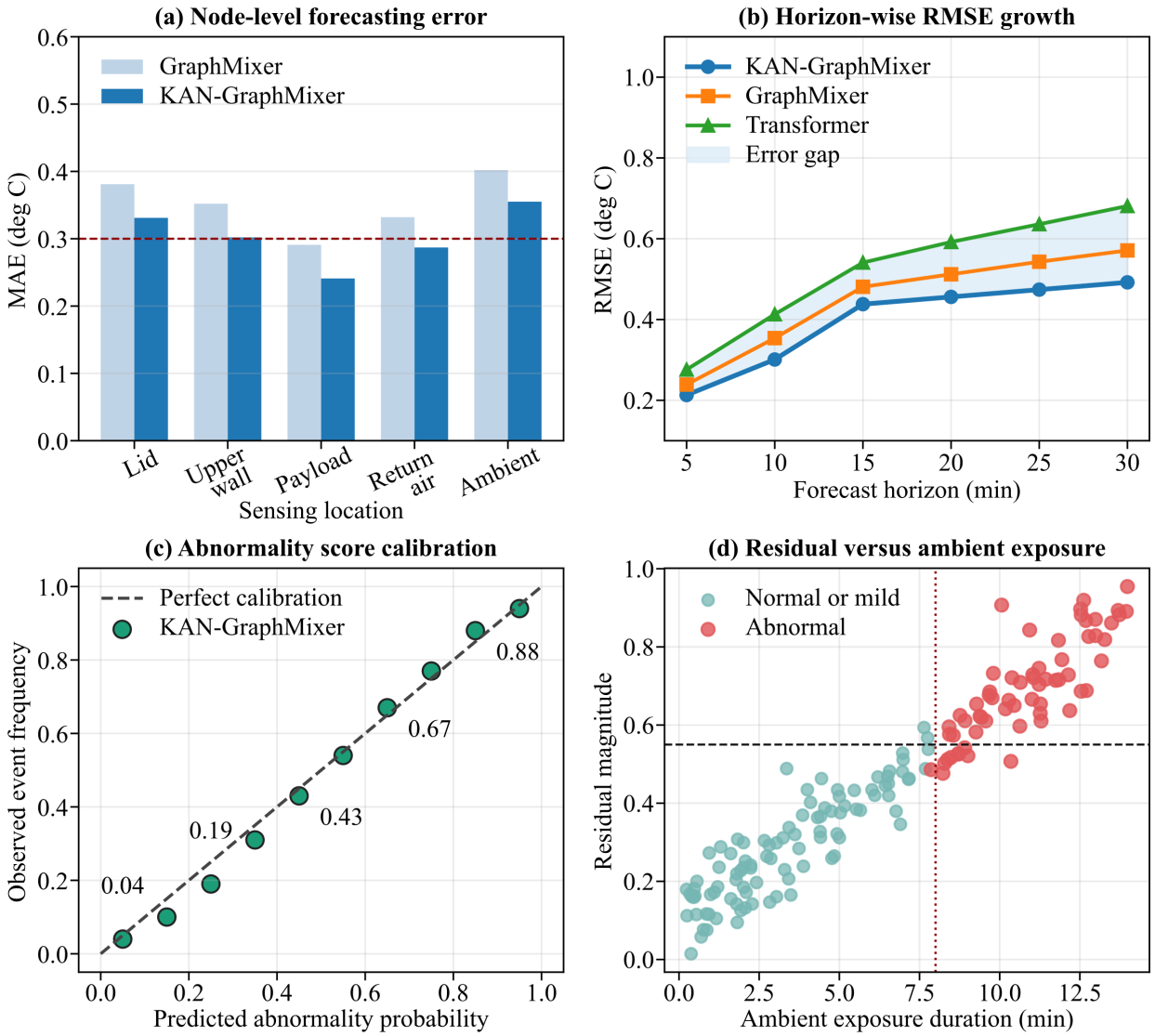


Figure 4. Sensor and decision behavior: (a) node-specific MAE; (b) horizon-dependent RMSE; (c) probability calibration; (d) residual magnitude versus exposure duration

Figure 5(a) evaluates five missing-packet levels of 0%, 3%, 5%, 10%, and 15%. F1-score decreases from 96.8% to 96.1%, 95.6%, 94.0%, and 92.7%, respectively. At 10% packet loss, performance reaches the 94.0% reference. The first five percentage points of packet loss reduce F1 by 1.2 points, while the next ten reduce it by another 2.9 points. A narrative guide to advanced cross-validation provides methodological context for evaluation design [22]. Figure 5(b) reports latencies of 6.1, 8.7, and 11.4 ms for windows of 8, 12, and 16 steps. The associated margins below the 12-ms inference budget are 5.9, 3.3, and 0.6 ms. A 16-step window approaches the inference budget and provides a limited margin for execution variability; data acquisition and communication impose additional processing costs. The latency values describe the reported model-inference stage. Sampling, graph updates, input preparation, scoring, and communication require separate timing records before the 12-ms target can be interpreted as an end-to-end guarantee. Processor configuration, software backend, warm-up policy, and latency percentiles are also needed to reproduce the timing comparison.

Figure 5(c) compares four abnormality-score distributions. Normal scores center near 0.106, door exposure near 0.51, delayed pre-cooling near 0.63, and refrigeration interruption near 0.842. The 0.55 guide intersects the door and pre-cooling distributions, revealing why these two event types are more sensitive to the threshold and persistence settings. Figure 5(d) compares two resource measures before and after quantization: parameter storage decreases from 0.91 to 0.84 MB, and peak memory decreases from 4.6 to 3.9 MB. These reductions are approximately 7.7% and 15.2%. The larger peak-memory saving reflects changes in runtime memory as well as parameter storage; attributing the saving to specific allocations requires separate profiling of intermediate activations and other runtime components. The reported storage reduction concerns the aggregate deployment artifact. Interpreting the storage reduction under an 8-bit weight configuration also requires metadata, spline storage, mixed-precision components, and serialization overhead to be

specified. A component-wise byte inventory is required to explain the modest artifact reduction, while memory attribution requires runtime profiling.

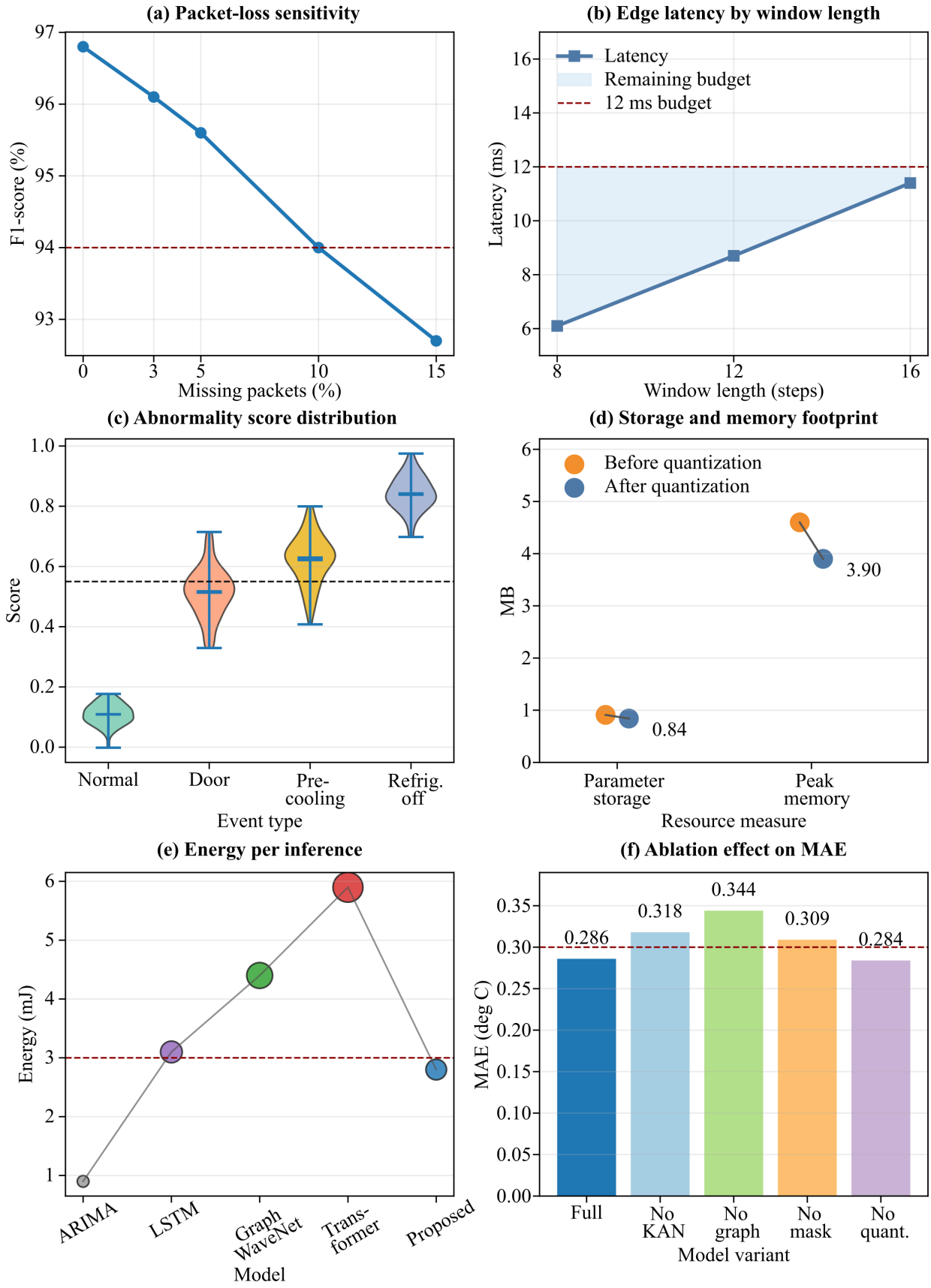


Figure 5. Deployment and component analysis: (a) packet-loss sensitivity; (b) inference latency; (c) score distributions; (d) storage and peak memory; (e) inference energy; (f) component-removal MAE

Figure 5(e) reports per-inference energies of 0.9, 3.1, 4.4, 5.9, and 2.8 mJ for ARIMA, LSTM, Graph WaveNet, Transformer, and the proposed method. Relative to Transformer, the proposed energy is lower by approximately 52.5%, while its F1-score is higher by 4.8 percentage points. The 3.0-mJ reference identifies the inference energy margin, while estimating route battery life requires the combined energy use of inference, sensing, idle operation, and radio activity. Cost-aware classifier calibration provides related context for balancing decision errors with their operational consequences [23]. That distinction affects how an energy-efficient alarm can be interpreted: reducing transmission may increase missing observations, which can erode the discrimination performance shown at the clean-data operating point.

Figure 5(f) compares five component variants. MAE is 0.286 degrees C for the full model, 0.318 without KAN, 0.344 without graph mixing, 0.309 without masking, and 0.284 without quantization. Removing graph mixing creates the largest increase, approximately 20.3%, whereas removing KAN increases MAE by approximately 11.2%. These findings indicate sensitivity to both relational representation and nonlinear channel response; retraining each variant requires their contributions to be interpreted at the model level. Disabling quantization improves MAE by 0.002 degrees C and increases the storage and memory requirements measured in the resource comparison. The masking variant is particularly dependent on test corruption: the reported F1-score falls to 94.1%, despite a smaller clean-data MAE change than that observed after removing graph mixing. These point estimates describe component sensitivity. Retraining changes each variant's optimized parameters; isolated causal attribution requires matched controls. Matched tuning budgets and repeated runs are needed to quantify uncertainty, particularly for the 0.002-degree quantization difference.

5.3. Ablation and Error Analysis

The component comparison is examined further through threshold and persistence sensitivity. Reducing the score threshold from 0.55 to 0.48 increases recall from 96.1% to 97.4%, but raises the false alarm rate from 1.9% to 3.6%. Raising the threshold to 0.62 increases precision to 98.4% while reducing recall to 93.7%. The low-threshold setting adds 1.3 percentage points of recall at a cost of 1.7 percentage points of false alarms; the high-threshold setting misses more short door exposures and mild pre-cooling delays. Removing persistence doubles false alarms during high-vibration loading, whereas requiring three consecutive windows delays refrigeration-interruption detection by 2.4 min on average. The test therefore exposes two separate controls: a score boundary determines which deviations are eligible for an alarm, and persistence determines how much temporal evidence is required before an eligible deviation becomes an event.

Robustness across box identity is also affected by calibration. Four box configurations are represented in the illustrative comparison to examine sensitivity to calibration and thermal-response variation. Sharing calibration statistics across boxes increases MAE from 0.286 to 0.304 degrees C and reduces F1-score from 96.8% to 95.5%. Restoring individual calibration recovers most of the reported loss within the evaluated box configurations; transfer to substantially different geometries requires additional assessment. Calibration quality depends on the availability of stable reference observations and the extent of loading-induced disturbance during the calibration interval.

Multiscale transformer-pyramid networks provide a related approach to multivariate time-series forecasting [24]. Residual localization here relates a prediction deviation to its contributing channels. Figure 6(a) displays a matrix with five event types and nine sensing nodes. Refrigeration interruption has normalized contributions of 0.86 at the return-air path, 0.81 at the side wall, and 0.78 at the payload center. These values are accompanied by 0.67 at the upper wall and 0.54 at the ambient node, giving the event a distributed thermal signature. Short door exposure instead has contributions of 0.69 at the lid and 0.58 at the upper wall, with only 0.31 at the payload center and 0.22 at the return-air path. Residual location and magnitude jointly distinguish a localized lid disturbance from a distributed deviation across thermally linked nodes. The heat-map values quantify normalized residual contributions; estimating component-failure probabilities requires a separately calibrated diagnostic model.

Prolonged door exposure has a lid contribution of 0.81, an upper-wall contribution of 0.74, and a humidity contribution of 0.74. Humidity is therefore an additional discriminating channel in this case, while its value during sensor drift is only 0.20. Delayed pre-cooling produces a payload-center contribution of 0.61 and a return-air contribution of 0.53, exceeding its lid contribution of 0.35. The matrix provides one contribution value per event-node pair: the 0.39 ambient value and the 0.61 payload value describe distinct sensing locations within the delayed pre-cooling pattern. A review and evaluation of elastic distance functions addresses time-series clustering as a separate approach to comparing temporal patterns [25]. Operating metadata help distinguish delayed pre-cooling from warmer-payload loading, since both conditions can produce moderate interior residuals accompanied by a weak lid signal.

The sensor-drift row differs from the thermal-event rows: battery contribution is 0.58, while lid, upper wall, payload, return air, and side wall contributions remain between 0.16 and 0.21. Localizing the drifting probe and assessing a possible electrical cause require measurement-quality review of sensor offset, timestamp continuity, and voltage-related acquisition instability. The vibration contribution is 0.13 during drift and 0.24 during refrigeration interruption, indicating a secondary contribution from the vibration channel. A reason code derived from these patterns should retain the distinction between a suspected thermal event and uncertain measurements. Reporting the diagnostic category

explicitly helps operators distinguish measurement-quality warnings from cargo-risk alarms when neighboring temperature channels remain consistent.

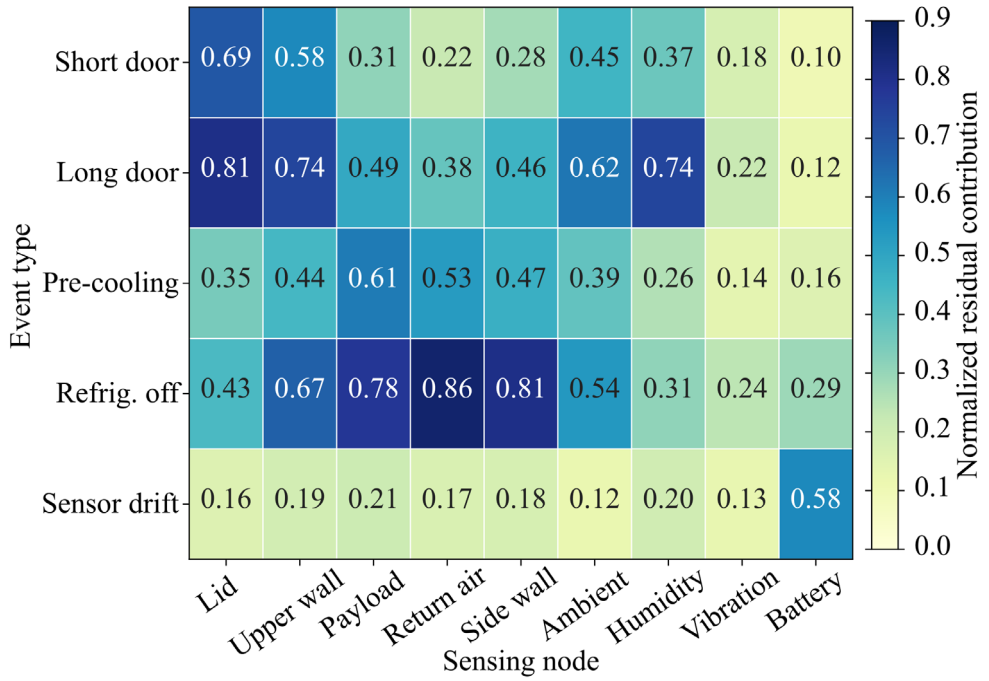


Figure 6. Residual localization: (a) normalized contributions for five event types across nine sensing nodes

The illustrative error inventory contains 73 false positives: 41 during inspection stops longer than 4 min, 19 during high-vibration loading, and 13 following temporary battery-voltage drops. The 58 false negatives comprise 24 door exposures shorter than 90 s, 18 gradual pre-cooling delays under low ambient temperature, and 16 mixed events involving door opening shortly before refrigeration recovery. These counts localize failure conditions; calculating aggregate precision, recall, and false alarm rate also requires the corresponding event and window denominators. A more permissive threshold changes the balance between false-positive and false-negative cases, making route and handling context relevant to decision calibration. The false-positive and false-negative inventories describe selected error cases. Their aggregation unit, positive-event denominator, negative-window denominator, and event-merging rule require further specification. The counts consequently provide a failure taxonomy, while reconstruction of the overall metrics requires a consistently defined confusion matrix and the full evaluation population.

Approximate-computing research surveys architectures, applications, and implementation challenges relevant to resource-aware deployment [26]. Analysis of ambiguous inspection cases requires a local record of handling timestamps and node contributions at alarm activation. A lid-dominant residual can direct inspection toward an opening or seal condition, whereas a coordinated return-air and payload residual can prompt a check of the refrigeration source. Verification of the associated fault causes requires inspection informed by these cues. The 3.3-ms inference margin at the selected 12-step window provides a smaller allowance for processing variability than the shorter window, and route-level battery use remains sensitive to radio scheduling. During a mixed event with missing ambient observations, the gateway may receive evidence supporting deviation detection; assigning a stable reason code still requires additional observations. Retaining validity flags with the alarm record supports retrospective review with explicit consideration of diagnostic confidence.

6. Conclusion

This study developed an edge-aware KAN-GraphMixer forecasting model for cold-box abnormality identification, with an illustrative evaluation indicating the potential of nonlinear channel transformations and graph mixing to support thermal prediction and local alarm decisions within the stated configuration.

1. The model achieved a 15-min temperature forecasting MAE of 0.286 °C and an RMSE of 0.438 °C, with abnormality identification precision, recall, and F1-score of 97.5%, 96.1%, and 96.8%, respectively. At a false alarm rate of 1.9%, the median warning lead time was 11.6 min, describing the reported interval between alarm activation and the subsequent event boundary. Establishing an actionable intervention benefit requires timestamped operational validation.

2. Compared with GraphMixer, the proposed model reduced MAE by approximately 10.1% and increased F1-score by 2.4 percentage points. Removing graph mixing increased MAE to 0.344 °C, whereas removing KAN transformations

increased it to 0.318 °C, indicating contributions from both sensor interaction and nonlinear channel response. Edge inference required 8.7 ms per window, with a quantized parameter footprint of 0.84 MB; energy consumption was 2.8 mJ per inference, approximately 52.5% lower than that of Transformer.

3. Performance remained sensitive to observation quality and box-specific calibration. F1-score declined to 92.7% at 15% packet loss, while sharing calibration statistics across boxes increased MAE to 0.304 °C. Brief door exposures and legitimate inspection stops remained difficult to distinguish, and residual localization provided inspection cues requiring subsequent verification of fault causes. These conditions limit direct transfer to different payload arrangements, sensing layouts, and operating environments.

Future work should validate the model using independently collected fleet data and investigate adaptive calibration and context-sensitive alarm policies under changing payload and communication conditions.

7. Declarations

7.1 Author Contributions

Norbert Brzeziński contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Oliwier Majchrzak contributes to validation, analysis, investigation, data collection, draft preparation. Patryk Rybak contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Date Availability

Not applicable.

7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Qullibot for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Waldhans, C., Albrecht, A., Ibald, R., Wollenweber, D., Sy, S.-J., & Kreyenschmidt, J. (2024). Temperature control and data exchange in food supply chains: Current situation and the applicability of a digitalized system of time–temperature-indicators to optimize temperature monitoring in different cold chains. Journal of Packaging Technology and Research, 8(1), 79–93. https://doi.org/10.1007/s41783-024-00165-2

[2] Baghel, L. K., Raina, R., Kumar, S., & Catarinucci, L. (2024). IoT-based integrated sensing and logging solution for cold chain monitoring applications. IEEE Journal of Radio Frequency Identification, 8, 837–846. https://doi.org/10.1109/jrfid.2024.3488534

[3] Yu, M., Zhang, H., Ma, J., Duan, X., Kang, S., & Li, J. (2025). Cold chain logistics supervision of agricultural products supported using internet of things technology. IEEE Internet of Things Journal, 12(4), 3502–3511. https://doi.org/10.1109/jiot.2024.3420949

[4] Bui, C. D., Quinn, A., Iacopino, D., & Narbudowicz, A. (2024). Compact chipless RFID sensor for frozen food monitoring. IEEE Sensors Journal, 24(9), 14205–14212. https://doi.org/10.1109/jsen.2024.3378678

[5] Trappolini, G., Purificato, A., Siciliano, F., D'Addona, L., Spagnolo, A. M., Dato, D., & Silvestri, F. (2025). Quantized auto encoder-based anomaly detection for multivariate time series data in 5G networks. IEEE Access, 13, 82668–82679. https://doi.org/10.1109/access.2025.3568133

[6] Hoffmann, T. G., Linke, M., Praeger, U., Sonawane, A. D., Büchele, F., Neuwald, D. A., Jedermann, R., Sturm, B., & Mahajan, P. V. (2025). Heat transfer in large bins during the apples cool-down process. International Journal of Refrigeration, 170, 60–69. https://doi.org/10.1016/j.ijrefrig.2024.11.023

[7] Schwedersky, B. B., & Flesch, R. C. C. (2025). Practical nonlinear model predictive control with Kolmogorov–Arnold network models. IEEE Access, 13, 146745–146756. https://doi.org/10.1109/access.2025.3600308

[8] Haryono, A. T., Sarno, R., Anggraini, R. N. E., & Sungkono, K. R. (2024). Permuted temporal Kolmogorov-Arnold networks for stock price forecasting using generative aspect-based sentiment analysis. IEEE Access, 12, 178672–178689. https://doi.org/10.1109/access.2024.3506658

[9] Mao, Y., Yu, X., Huang, K., Zhang, Y. J. A., & Zhang, J. (2024). Green edge AI: A contemporary survey. Proceedings of the IEEE, 112(7), 880–911. https://doi.org/10.1109/jproc.2024.3437365

[10] Moosmann, J., Müller, H., Zimmerman, N., Rutishauser, G., Benini, L., & Magno, M. (2024). Flexible and fully quantized lightweight TinyissimoYOLO for ultra-low-power edge systems. IEEE Access, 12, 75093–75107. https://doi.org/10.1109/access.2024.3404878

[11] Manca, G., Kunze, F. C., & Fay, A. (2024). Addressing uncertainty in online alarm flood classification using conformal prediction. IEEE Access, 12, 165626–165652. https://doi.org/10.1109/access.2024.3492348

[12] Huang, Y., Liu, P., Han, R., Yin, T., Dou, Q., Song, Y., & Luo, M. (2025). FreqWave-TranDuD: A multivariate time series anomaly detection method based on wavelet and Fourier transforms. IEEE Access, 13, 68384–68397. https://doi.org/10.1109/access.2025.3557571

[13] Richter, A., Ijaradar, J., Wetzker, U., Jain, V., & Frotzscher, A. (2024). A survey on multivariate time series imputation using adversarial learning. IEEE Access, 12, 148167–148189. https://doi.org/10.1109/access.2024.3473540

[14] Asaadi, M., Aslanseefat, K., Izadi, I., & Yang, F. (2024). Adaptive design of alarm systems in industrial processes. IFAC-PapersOnLine, 58(14), 841–846. https://doi.org/10.1016/j.ifacol.2024.08.442

[15] Manca, G., Kunze, F. C., & Fay, A. (2025). Predicting uncertainty reduction in online alarm flood classification. IFAC-PapersOnLine, 59(25), 119–124. https://doi.org/10.1016/j.ifacol.2025.11.935

[16] Amazu, C. W., Mietkiewicz, J., Abbas, A. N., Briwa, H., Alonso-Perez, A., Baldissone, G., Fissore, D., Demichela, M., & Leva, M. C. (2025). Exploring the influence of human system interfaces: Introducing support tools and an experimental study. International Journal of Human–Computer Interaction, 41(10), 6300–6317. https://doi.org/10.1080/10447318.2024.2376354

[17] Balanya, S. A., Maroñas, J., & Ramos, D. (2024). Adaptive temperature scaling for robust calibration of deep neural networks. Neural Computing and Applications, 36(14), 8073–8095. https://doi.org/10.1007/s00521-024-09505-4

[18] Ivanov, P., Shtark, M., Kozhevnikov, A., Golyadkin, M., Botov, D., & Makarov, I. (2025). SensorDBSCAN: Semi-supervised active learning powered method for anomaly detection and diagnosis. IEEE Access, 13, 25186–25197. https://doi.org/10.1109/access.2025.3537649

[19] Zarif, M., & Moghaddass, R. (2025). Advanced anomaly detection in smart grids using graph convolutional networks with integrated node and line sensor data. IEEE Access, 13, 141170–141186. https://doi.org/10.1109/access.2025.3595431

[20] Carletti, V., Foggia, P., Rosa, F., & Vento, M. (2025). Detecting malicious IoT network communication through Graph Neural Networks in real-world conditions. Pattern Recognition Letters, 189, 92–98. https://doi.org/10.1016/j.patrec.2025.01.010

[21] Liu, S., & Zhou, D. J. (2024). Using cross-validation methods to select time series models: Promises and pitfalls. British Journal of Mathematical and Statistical Psychology, 77(2), 337–355. https://doi.org/10.1111/bmsp.12330

[22] Allgaier, J., & Pryss, R. (2024). Cross-validation visualized: A narrative guide to advanced methods. Machine Learning and Knowledge Extraction, 6(2), 1378–1388. https://doi.org/10.3390/make6020065

[23] Yang, M., & Bi, X. (2025). Cost-aware calibration of classifiers. INFORMS Journal on Data Science, 4(2), 101–113. https://doi.org/10.1287/ijds.2024.0038

[24] Zhang, Y., Wu, R., Dascalu, S. M., & Harris, F. C. (2024). Multi-scale transformer pyramid networks for multivariate time series forecasting. IEEE Access, 12, 14731–14741. https://doi.org/10.1109/access.2024.3357693

[25] Holder, C., Middlehurst, M., & Bagnall, A. (2024). A review and evaluation of elastic distance functions for time series clustering. Knowledge and Information Systems, 66(2), 765–809. https://doi.org/10.1007/s10115-023-01952-0

[26] Dalloo, A. M., Humaidi, A. J., Al Mhdawi, A. K., & Al-Raweshidy, H. (2024). Approximate computing: Concepts, architectures,

challenges, applications, and future directions. IEEE Access, 12, 146022–146088. https://doi.org/10.1109/access.2024.3467375