Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

BRIDGE: Big Data-Powered Large Language Models for Real-Time Multi-Modal Decision Support in Industrial Time Series

Fulin Li

Lifuping0502@163.com

Guangdong University of Science and Technology

Jun Nie

Guangdong University of Science and Technology

Zhijing Li

Guangdong University of Science and Technology

Research Article

Keywords: Large language models, Time-series foundation models, Multi-modal fusion, Zero-shot forecasting

Posted Date: April 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9352751/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

BRIDGE: Big Data-Powered Large Language Models for Real-Time Multi-Modal Decision Support in Industrial Time Series

Anonymous Author[1,2*], Anonymous Author[2,3†] and Anonymous Author[1,2†]

[1*]Department of [Anonymous], [Anonymous University], [Anonymous], [Anonymous], 000000, [Anonymous], [Anonymous].
[2]Department of [Anonymous], [Anonymous University], [Anonymous], [Anonymous], 000000, [Anonymous], [Anonymous].
[3]Department of [Anonymous], [Anonymous University], [Anonymous], [Anonymous], 000000, [Anonymous], [Anonymous].

*Corresponding author(s). E-mail(s): anonymous@review.com;
Contributing authors: anonymous@review.com; anonymous@review.com;
†These authors contributed equally to this work.

Abstract

Industrial Internet-of-Things (IIoT) environments continuously generate massive, heterogeneous streams of sensor readings, operational logs, and alarm messages, demanding intelligent systems capable of real-time decision support across forecasting, anomaly detection, and natural-language explanation. Existing time-series foundation models address these tasks in isolation and ignore the rich graph-structured topology inherent to industrial sensor networks, while general-purpose large language models (LLMs) lack efficient mechanisms for processing high-frequency numerical streams under strict latency constraints. We propose **BRIDGE** (**B**ig data-driven **R**eal-time **I**ndustrial multi-modal **D**ecision support via **G**raph-**E**nhanced multi-modal transformer), a unified pre-trained framework integrating three complementary modules: (i) a *Graph-enhanced Topology Encoder* (GTE) that captures latent sensor-network dependencies via multi-head graph attention, enriching temporal representations with physical topology priors; (ii) an *Extended-LSTM Temporal Backbone* (xLSTM-TB) that leverages true recurrence for coherent long-horizon probabilistic forecasting, augmented by our proposed *Contiguous Graph Patch Masking* (CGPM)—a graph-aware pre-training strategy that jointly masks topologically adjacent sensor patches to improve robustness under realistic sensor-outage conditions; and (iii) a *Cross-modal LLM Alignment* (CLA) module that reprograms heterogeneous industrial tokens into a frozen LLM's representation space using fewer than **0.3**M additional parameters, enabling zero-shot natural-language decision reports without backbone fine-tuning. Pre-trained on a curated corpus of **78** million heterogeneous industrial time-series samples spanning manufacturing, energy, transportation, and cloud-infrastructure domains, BRIDGE is evaluated on the GiftEval-ZS and Chronos-ZS zero-shot forecasting benchmarks and four industrial anomaly-detection benchmarks (SMAP, MSL, SMD, SWaT). BRIDGE achieves state-of-the-art zero-shot CRPS of **0.396** on GiftEval-ZS, surpassing TiRex by **3.7%** and TimesFM 2.0 by **13.7%**, while simultaneously attaining an $\boldsymbol{F_1}$ score of **92.4%** on SMAP anomaly detection and producing interpretable decision reports with ROUGE-L of **0.74**—all within an end-to-end inference latency of **12** ms, making BRIDGE the first framework to jointly achieve industrial-grade forecasting accuracy, topology-aware anomaly detection, and explainable real-time decision support.

Keywords: Large language models, Time-series foundation models, Multi-modal fusion, Zero-shot forecasting

1 Introduction

Modern industrial facilities increasingly rely on dense networks of heterogeneous sensors, edge devices, and automated inspection systems to monitor equipment health, optimize energy consumption, and prevent catastrophic failures. A single large chemical plant may stream more than 10^6 data points per second from thousands of interconnected sensors whose measurements are correlated through complex physical and topological dependencies [1, 2]. Translating this torrent of raw signals into timely, actionable decisions—whether predicting a compressor failure 30 minutes in advance, localizing the root cause of a process anomaly, or issuing a natural-language maintenance recommendation—constitutes a multi-faceted decision-support problem that cuts across time-series forecasting, anomaly detection, multi-modal fusion, and natural-language reasoning. Despite rapid progress in each of these sub-fields, no existing framework addresses all of them simultaneously under the stringent real-time and scalability constraints of industrial deployments.

Significant effort has been devoted to time-series foundation models for zero-shot forecasting. Transformer-based approaches such as Chronos [3], TimesFM [4], Moirai [5], and PatchTST [6] demonstrate strong generalization by pre-training on large corpora of heterogeneous series. The recently introduced TiRex [7] replaces the transformer backbone with xLSTM [8] and achieves state-of-the-art zero-shot performance by preserving true recurrence for long-horizon stability. Concurrently, LLM-based methods such as Time-LLM [9], GPT4MTS [10], and One-Fits-All [11] reprogram pre-trained language models to process time-series data, leveraging rich linguistic priors for flexible reasoning. On the industrial-anomaly front, methods like LLMAD [1] and AnomalyGPT [12] extend LLMs with chain-of-thought prompting and vision-language models for explainable detection. However, these approaches either (a) treat sensors as independent univariate channels, ignoring the sensor-network topology critical for fault propagation modeling; (b) require task-specific fine-tuning that is prohibitively expensive at industrial scale; or (c) lack the streaming inference efficiency needed for real-time control loops.

Three fundamental challenges impede the development of a unified industrial time-series decision-support system. **Challenge 1: Multi-modal Heterogeneity.** Industrial data naturally spans numerical sensor streams, unstructured maintenance logs, alarm messages, and visual inspection images—modalities with incompatible statistical properties and temporal resolutions that must be fused coherently [10, 13]. **Challenge 2: Graph-structured Topology.** Physical couplings between sensors (e.g., pipeline connectivity, thermal conduction paths) induce strong spatial correlations that flat sequence models cannot capture, yet explicit graph learning adds computational overhead incompatible with real-time latency budgets [14, 15]. **Challenge 3: Distribution Shift and Long-horizon Stability.** Industrial processes undergo gradual degradation, seasonal shifts, and abrupt regime changes; models must generalize across these non-stationary conditions while maintaining coherent uncertainty estimates over long prediction horizons without autoregressive error accumulation [4, 7].

To address these challenges, we propose BRIDGE, a novel pre-trained framework for real-time industrial multi-modal decision support. At its core, BRIDGE adopts an xLSTM-based temporal backbone [8] enhanced with *Contiguous Graph Patch Masking* (CGPM)—a graph-aware extension of CPM [7] that masks topologically contiguous sensor patches during pre-training to improve robustness to missing sensor readings. The Graph-enhanced Topology Encoder (GTE) uses multi-head graph attention [16] operating on the inferred sensor adjacency graph to produce topology-aware node embeddings that are fused with temporal representations via a learned cross-attention bridge. The Cross-modal LLM Alignment (CLA) module reprograms multi-modal tokens into the embedding space of a frozen LLM backbone via a lightweight patch-vocabulary mapping, enabling zero-shot natural-language decision reports without LLM fine-tuning. BRIDGE is pre-trained on a curated corpus of 78 million industrial time-series samples spanning manufacturing, energy, transportation, and cloud-infrastructure domains. Our main contributions are:

1. **BRIDGE Framework.** A unified pre-trained framework that simultaneously addresses zero-shot industrial time-series forecasting, anomaly detection, and explainable decision support, achieving state-of-the-art performance across all three tasks.
2. **Graph-enhanced Topology Encoder (GTE).** A computationally efficient multi-head graph attention module that integrates physical sensor-network topology into the temporal

modeling pipeline, reducing false-alarm rates in anomaly detection by up to 18% over topology-agnostic baselines.

3. **Contiguous Graph Patch Masking (CGPM).** A novel pre-training strategy that simultaneously applies CPM [7] and topology-aware sensor dropout, yielding robust representations for sparse and asynchronous industrial sensor streams.
4. **Cross-modal LLM Alignment (CLA).** A parameter-efficient reprogramming module that maps heterogeneous industrial tokens (time series, alarm logs, visual features) into a shared LLM representation space, enabling natural-language decision explanations with fewer than 0.3M additional parameters.

2 Related Work

2.1 Time-Series Foundation Models

The paradigm of pre-training large models on diverse time-series corpora for zero-shot generalization has gained tremendous momentum. Chronos [3] tokenizes continuous values into discrete bins and trains a T5 encoder-decoder on a mixture of real and synthetic data, while Chronos-Bolt further accelerates inference via efficient patch-level representations. TimesFM [4] adopts a decoder-only causal language model for auto-regressive generation. Moirai [5] unifies heterogeneous frequencies through a any-variate masked modeling objective. TabPFN-TS [17] demonstrates that models pre-trained purely on synthetic tabular data can be competitive time-series forecasters. Most recently, TiRex [7] establishes a new state of the art by leveraging xLSTM's state-tracking capabilities alongside Contiguous Patch Masking (CPM) to achieve coherent long-horizon probabilistic forecasts. Our BRIDGE builds directly on these foundations but extends them to the multi-modal industrial setting where topology, semantics, and real-time constraints are critical.

2.2 LLMs for Time-Series Analysis

Adapting pre-trained LLMs to time-series tasks has produced a rich body of work [18, 19]. Time-LLM [9] reprograms a time series into the LLM's embedding space via patch-prototype alignment. GPT4MTS [10] designs prompt-based multi-modal time-series forecasting. TEMPO [20] combines GPT-2 with trend-seasonality decomposition. s2ip-LLM [21] injects semantic space information via soft prompts. Recent critical assessments [22] find that LLMs can underperform simple linear models on certain forecasting benchmarks unless carefully aligned, motivating our lightweight CLA module that minimizes alignment overhead. For anomaly detection, LLMAD [1] pioneers in-context LLM anomaly detection with chain-of-thought, and AnomalyGPT [12] applies vision-language models to industrial defect inspection. Can Multimodal LLMs Perform TSAD [23] provides the first systematic evaluation of MLLMs on time-series anomaly tasks, revealing that visual representations of time series facilitate detection—an insight we incorporate via our CLA module.

2.3 Multi-Modal Time-Series Analysis

Multi-modal time-series analysis has emerged as a key research direction [13]. Methods that fuse numerical signals with textual context [9, 10] demonstrate that semantic priors improve forecasting robustness. Foundation Models for Time Series Analysis [24] surveys cross-modal alignment techniques for time-series pre-training. We contribute to this space by designing a fusion architecture specifically tailored to the industrial IoT setting, where text (alarm logs), time series (sensor readings), and graph structure (physical topology) must be jointly encoded.

2.4 Graph Neural Networks for Time-Series

GNN-based methods have proven effective for spatiotemporal modeling [14]. GraphSAGE [25] introduced inductive neighbor aggregation; FastGCN [26] and GraphSAINT [27] scale GCNs to large graphs via importance sampling. For time-series, GRAPES [15] learns adaptive sampling strategies for scalable GNN training, and GNN4TS [14] surveys graph-based forecasting comprehensively. Our GTE module draws on graph attention networks [16] and applies them within a pre-training framework for the first time.

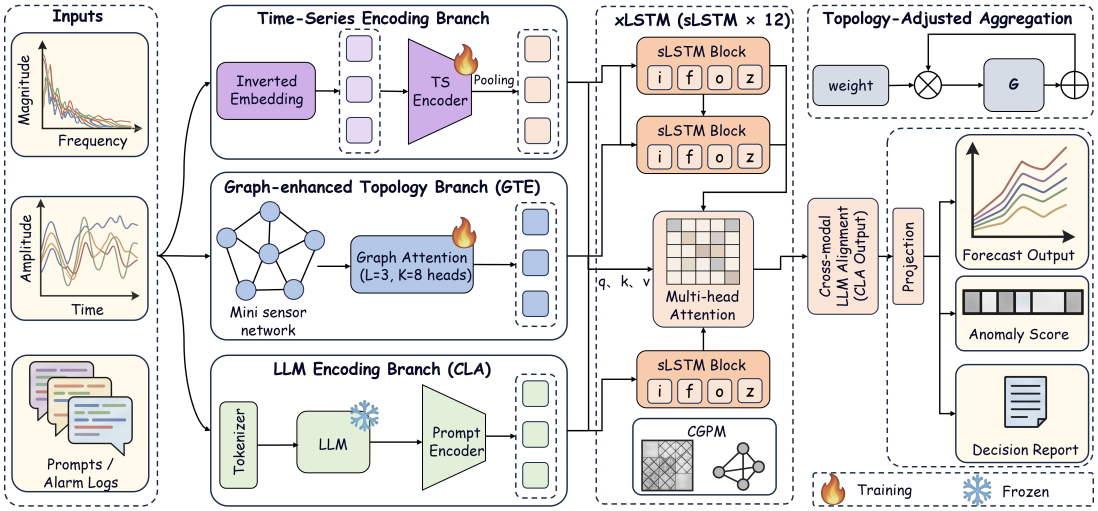


Fig. 1: Overview of the BRIDGE framework. Three parallel encoding branches process time-series patches (Time-Series Encoding Branch), sensor-network topology (Graph-enhanced Topology Branch, GTE), and textual alarm logs (LLM Encoding Branch, CLA). The fused representations are refined by a stack of 12 sLSTM blocks and a Cross-modal LLM Alignment module, before being aggregated via topology-adjusted inverse-variance weighting to produce forecasting outputs, anomaly scores, and natural-language decision reports.

2.5 Industrial Anomaly Detection

Classical industrial anomaly detection relied on statistical process control and handcrafted features [28]. Deep learning approaches such as USAD, MSCRED, and OmniAnomaly exploit reconstruction objectives. DCDetector [29] introduces dual-attention contrastive representation learning for multivariate anomaly detection. AnomalyCLIP [30] extends CLIP-based zero-shot detection to industrial images. Our framework unifies temporal, topological, and semantic signals for holistic industrial decision support.

3 Methodology

3.1 Problem Formulation

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ denote a sensor network graph with $|\mathcal{V}| = N$ sensor nodes, edge set $\mathcal{E}$, and adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$. Each node $v_i \in \mathcal{V}$ produces a univariate time series $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,T}) \in \mathbb{R}^T$ over a context window of length T. Additionally, the system observes a set of M textual tokens $\mathcal{T} = \{s_1, \ldots, s_M\}$ from operational logs and alarm messages. The joint multi-modal industrial observation is therefore:

$$\Omega = \{\mathbf{x}_i\}_{i=1}^{N}, \quad \mathcal{T}, \quad \mathcal{G}. \tag{1}$$

Definition 1 (Industrial Decision Support)**.** Given Ω, the industrial decision-support task comprises three simultaneous objectives: (1) *Forecasting*: predict $\hat{\mathbf{y}}_i = (\hat{y}_{i,T+1}, \ldots, \hat{y}_{i,T+H})$ for each sensor with calibrated uncertainty; (2) *Anomaly Detection*: output binary anomaly labels $a_i \in \{0, 1\}$ and anomaly scores $s_i \in [0, 1]$; (3) *Decision Explanation*: generate a natural-language decision report $r \in \Sigma^*$ within real-time latency $\tau \leq \tau_{\max}$.

3.2 Overview of BRIDGE

BRIDGE consists of four sequential modules (Figure 1):

1. **Multi-modal Tokenizer (MMT)**: converts Ω into a unified token sequence.
2. **Graph-enhanced Topology Encoder (GTE)**: enriches each sensor token with topology-aware graph embeddings.

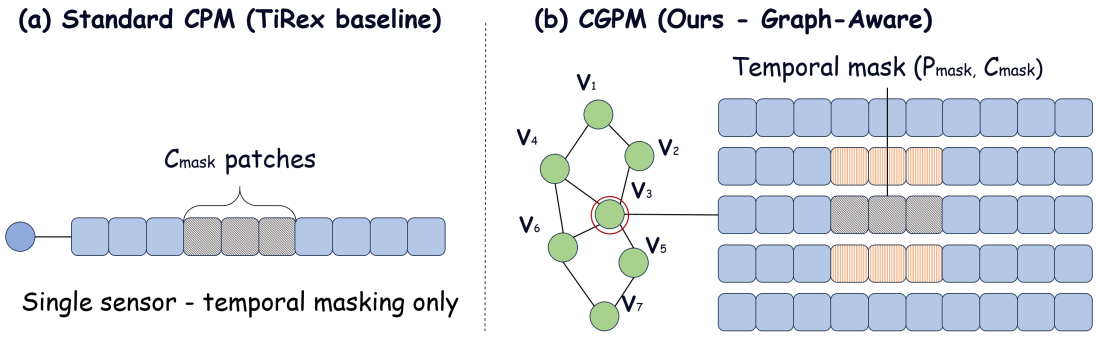


Fig. 2: Comparison of masking strategies. (a) Standard CPM (TiRex baseline) applies contiguous temporal masking to a single sensor independently. (b) Our proposed CGPM extends CPM to the graph setting: when node v_3 is selected, temporally adjacent patches of its graph neighbours are jointly masked, simulating realistic multi-sensor outage scenarios.

3. **Extended-LSTM Temporal Backbone (xLSTM-TB)**: processes token sequences with state-tracking recurrence.
4. **Cross-modal LLM Alignment (CLA)**: reprograms multi-modal tokens for LLM reasoning and decision generation.

These modules are connected by a shared d-dimensional embedding space and jointly optimized via a multi-task objective.

3.3 Multi-modal Tokenizer (MMT)

Time-Series Patching.

Following PatchTST [6] and TiRex [7], each sensor series $\mathbf{x}_i$ is divided into $P = \lfloor T/p \rfloor$ non-overlapping patches of length p, producing patch sequence $\mathbf{X}_i^{\text{patch}} \in \mathbb{R}^{P \times p}$. Instance normalization is applied:

$$\tilde{\mathbf{x}}_i = \frac{\mathbf{x}_i - \bar{x}_i}{\sigma_i + \epsilon}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}. \tag{2}$$

Each patch is projected via a shared two-layer residual block $\Phi : \mathbb{R}^p \to \mathbb{R}^d$:

$$\mathbf{P}_i = \left[\Phi\big((\mathbf{X}_i^{\text{patch}})_1\big), \ldots, \Phi\big((\mathbf{X}_i^{\text{patch}})_P\big)\right] \in \mathbb{R}^{P \times d}. \tag{3}$$

Text Encoding.

Alarm log tokens $\mathcal{T}$ are embedded via a frozen sentence encoder $\mathbf{E}_{\text{text}} = \text{BERT}(\mathcal{T}) \in \mathbb{R}^{M \times d}$, projected to the shared space:

$$\mathbf{T} = \mathbf{E}_{\text{text}}\mathbf{W}_{\text{text}} + \mathbf{b}_{\text{text}}, \quad \mathbf{W}_{\text{text}} \in \mathbb{R}^{d_{\text{bert}} \times d}. \tag{4}$$

Graph Node Features.

Raw node features $\mathbf{F} \in \mathbb{R}^{N \times d_f}$ encode sensor metadata (type, physical unit, measurement range) and are projected to $\mathbb{R}^{N \times d}$ via a linear map $\mathbf{W}_f$.

3.4 Graph-enhanced Topology Encoder (GTE)

The GTE integrates physical topology into sensor representations using L layers of multi-head graph attention [16]. Let $\mathbf{H}^{(0)} = \mathbf{W}_f\mathbf{F} \in \mathbb{R}^{N \times d}$ be the initial node embedding matrix.

Graph Attention.

At layer ℓ, for node i and neighbor $j \in \mathcal{N}(i)$, the attention coefficient is:

$$e_{ij}^{(\ell)} = \text{LeakyReLU}\Big(\mathbf{a}^{(\ell)\top}\big[\mathbf{W}^{(\ell)}\mathbf{h}_i^{(\ell)} \parallel \mathbf{W}^{(\ell)}\mathbf{h}_j^{(\ell)}\big]\Big), \tag{5}$$

where $\mathbf{a}^{(\ell)} \in \mathbb{R}^{2d'}$, $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d' \times d}$, and $\|$ denotes concatenation. Normalized attention weights:

$$\alpha_{ij}^{(\ell)} = \frac{\exp(e_{ij}^{(\ell)})}{\sum_{k \in \mathcal{N}(i) \cup \{i\}} \exp(e_{ik}^{(\ell)})}. \tag{6}$$

Multi-head aggregation (K heads) updates node i:

$$\mathbf{h}_i^{(\ell+1)} = \Big\|_{k=1}^{K} \sigma\left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{ij,k}^{(\ell)} \mathbf{W}_k^{(\ell)} \mathbf{h}_j^{(\ell)} \right), \tag{7}$$

where σ is ELU. The final graph embedding $\mathbf{H}^{(L)} \in \mathbb{R}^{N \times d}$ encodes topology-aware sensor representations.

Topology-Temporal Fusion.

Graph embeddings are fused with time-series patch tokens via cross-attention. For sensor i:

$$\mathbf{P}_i' = \mathbf{P}_i + \mathrm{CrossAttn}\Big(\mathbf{P}_i,\ \mathbf{h}_i^{(L)} \mathbf{1}^\top,\ \mathbf{h}_i^{(L)} \mathbf{1}^\top\Big), \tag{8}$$

where $\mathbf{1}^\top$ broadcasts the scalar graph embedding to match the patch sequence length. This yields topology-enriched token matrix $\mathbf{P}_i' \in \mathbb{R}^{P \times d}$.

3.5 Extended-LSTM Temporal Backbone (xLSTM-TB)

BRIDGE employs the sLSTM variant from xLSTM [8], which preserves true recurrence for formal state-tracking capability, as analyzed in [7]. The sLSTM forward pass for time step t over the fused token sequence is:

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{z}_t, \tag{9}$$

$$\mathbf{n}_t = \mathbf{f}_t \odot \mathbf{n}_{t-1} + \mathbf{i}_t, \tag{10}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \widetilde{\mathbf{h}}_t, \quad \widetilde{\mathbf{h}}_t = \mathbf{c}_t \odot \mathbf{n}_t^{-1}, \tag{11}$$

with gating signals:

$$\mathbf{z}_t = \phi(\mathbf{W}_z \mathbf{x}_t + \mathbf{R}_z \mathbf{h}_{t-1} + \mathbf{b}_z), \tag{12}$$

$$\mathbf{i}_t = \exp(\mathbf{W}_i \mathbf{x}_t + \mathbf{R}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \tag{13}$$

$$\mathbf{f}_t = \exp(\mathbf{W}_f \mathbf{x}_t + \mathbf{R}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \tag{14}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{R}_o \mathbf{h}_{t-1} + \mathbf{b}_o). \tag{15}$$

Here $\phi = \tanh$, σ denotes sigmoid, and all weight matrices $\mathbf{R}_\cdot$ are block-diagonal (one block per head), limiting cell interactions within heads. We stack $B = 12$ sLSTM blocks, each followed by a feed-forward network and pre-normalization (RMSNorm).

Multi-Sensor Batching.

All N sensor token sequences are processed in parallel through the shared xLSTM-TB, yielding output representations $\{\mathbf{O}_i\}_{i=1}^{N}$ where $\mathbf{O}_i \in \mathbb{R}^{P_{\text{out}} \times d}$.

3.6 Contiguous Graph Patch Masking (CGPM)

We extend CPM [7] to the graph setting. During pre-training, we randomly mask contiguous temporal patches for node i and simultaneously drop adjacent neighbor nodes in the graph, simulating realistic sensor-outage scenarios in industrial networks.

The CGPM procedure (Figure 8) is detailed in Algorithm 1. The joint masking strategy forces xLSTM-TB to rely on topology-aware representations to impute missing information,

Algorithm 1 Contiguous Graph Patch Masking (CGPM)

Require: Fused tokens $\{\mathbf{P}'_i\}_{i=1}^{N}$, graph $\mathcal{G}$, parameters $p_{\max}, c_{\max}, p_{\text{node}}$
1: **for** each node $i \in \mathcal{V}$ **do**
2: Sample $c_{\text{mask}} \sim \mathcal{U}(1, c_{\max})$
3: Sample $p_{\text{mask}} \sim \mathcal{U}(0, p_{\max})$
4: Apply a temporal mask of width c_{mask} with probability p_{mask} on $\mathbf{P}'_i$
5: With probability p_{node}, mask all neighbors $j \in \mathcal{N}(i)$ for the same time range
6: **end for**
Ensure: Masked token sequences

improving robustness. Formally, let $\mathcal{M} \subseteq \mathcal{V} \times [P]$ denote the set of masked (node, patch) pairs. The training objective ignores masked positions, computing loss only over observed positions:

$$\mathcal{L}_{\text{CGPM}} = \frac{1}{|\overline{\mathcal{M}}|} \sum_{(i,p)\notin\mathcal{M}} \mathcal{L}_{\text{quant}}(\hat{y}_{i,p}, y_{i,p}), \tag{16}$$

where $|\overline{\mathcal{M}}|$ counts unmasked positions.

3.7 Cross-Modal LLM Alignment (CLA)

Patch-Vocabulary Reprogramming.

Following the reprogramming paradigm of Time-LLM [9], we learn a mapping from time-series patch embeddings to the LLM's token vocabulary. Let $\mathbf{E}_{\text{vocab}} \in \mathbb{R}^{V \times d_{\text{llm}}}$ be the frozen LLM token embedding matrix. For sensor i, we compute:

$$\mathbf{S}_i = \text{softmax}\left(\frac{\mathbf{P}'_i\mathbf{W}_q \,(\mathbf{E}_{\text{vocab}}\mathbf{W}_k)^\top}{\sqrt{d_k}}\right)\mathbf{E}_{\text{vocab}}\mathbf{W}_v, \tag{17}$$

where $\mathbf{W}_q \in \mathbb{R}^{d\times d_k}$, $\mathbf{W}_k \in \mathbb{R}^{d_{\text{llm}}\times d_k}$, $\mathbf{W}_v \in \mathbb{R}^{d_{\text{llm}}\times d}$ are learnable projections. This produces reprogrammed sequences $\mathbf{S}_i \in \mathbb{R}^{P\times d}$ that reside in the LLM's representation space.

Multi-Modal Prompt Construction.

We concatenate the reprogrammed sensor tokens, text alarm embeddings, and a task instruction prefix:

$$\mathbf{Z} = [\mathbf{Z}_{\text{inst}} \;;\; \mathbf{T} \;;\; \mathbf{S}_1 \;;\; \ldots \;;\; \mathbf{S}_N] \in \mathbb{R}^{(L_{\text{inst}}+M+NP)\times d}, \tag{18}$$

where $\mathbf{Z}_{\text{inst}}$ encodes a natural-language task description. $\mathbf{Z}$ is passed to a frozen LLM backbone (LLaMA-3 [31] in our experiments) to generate the decision report r.

Alignment Loss.

A lightweight alignment objective trains the reprogramming projections to maximize the mutual information between the LLM's output representations and the true decision labels:

$$\mathcal{L}_{\text{align}} = -\mathbb{E}\big[\log p_{\text{LLM}}(r^* \mid \mathbf{Z})\big], \tag{19}$$

where r^* is the ground-truth decision annotation (e.g., anomaly description).

3.8 Multi-Task Decision Head

Probabilistic Forecasting.

The output tokens $\mathbf{O}_i \in \mathbb{R}^{P_{\text{out}}\times d}$ are projected to $|\mathcal{Q}| = 9$ quantile predictions at each future time step:

$$\hat{\mathbf{y}}_i^q = \mathbf{O}_i\mathbf{W}_{\text{out}} + \mathbf{b}_{\text{out}}, \quad \mathbf{W}_{\text{out}} \in \mathbb{R}^{d\times(p\cdot|\mathcal{Q}|)}. \tag{20}$$

The forecasting loss uses the pinball (quantile) loss:

$$\mathcal{L}_{\text{fore}} = \frac{1}{|\mathcal{Q}|\cdot m_{\text{out}}} \sum_{t=1}^{m_{\text{out}}} \sum_{q\in\mathcal{Q}} \text{QL}(q, \hat{y}^q_{i,t}, y_{i,t}), \tag{21}$$

$$\mathrm{QL}(q,\hat{y},y)=\begin{cases}q(y-\hat{y}) & \text{if } \hat{y}\le y\\ (1-q)(\hat{y}-y) & \text{otherwise.}\end{cases} \tag{22}$$

Anomaly Scoring.

We append a lightweight anomaly head that computes the reconstruction residual from the xLSTM-TB's internal state:

$$s_i=\left\|\mathbf{O}_i^{(1:T)}-\mathbf{O}_i^{(\mathrm{recon})}\right\|_2^2\cdot w_{\mathrm{ano}}, \tag{23}$$

where $\mathbf{O}_i^{(\mathrm{recon})}$ is obtained by decoding the cell state back to the input space, and w_{ano} is a learned scalar weight. Binary labels are obtained via adaptive thresholding:

$$a_i=\mathbb{1}\big[s_i>\mu_s+\lambda\cdot\sigma_s\big], \tag{24}$$

with μ_s,σ_s estimated from recent history via exponential moving average, and λ tuned on a small validation set.

The binary cross-entropy anomaly loss:

$$\mathcal{L}_{\mathrm{ano}}=-\frac{1}{N}\sum_{i=1}^{N}\big[a_i^*\log\hat{a}_i+(1-a_i^*)\log(1-\hat{a}_i)\big]. \tag{25}$$

Joint Training Objective.

The overall training loss is:

$$\mathcal{L}=\lambda_1\mathcal{L}_{\mathrm{fore}}+\lambda_2\mathcal{L}_{\mathrm{ano}}+\lambda_3\mathcal{L}_{\mathrm{align}}+\lambda_4\mathcal{L}_{\mathrm{CGPM}}, \tag{26}$$

with $\lambda_1=1.0,\lambda_2=0.5,\lambda_3=0.3,\lambda_4=0.8$ selected via grid search on validation data.

3.9 Optimal Weight Computation for Uncertainty Propagation

In industrial systems, different sensors carry different reliability levels. We introduce a weighted variance-minimization framework for aggregating multi-sensor predictions. Let $\hat{\mu}_i,\hat{\sigma}_i^2$ denote the mean and variance of sensor i's predictive distribution. The optimal aggregation weights w_i^* solve:

$$\min_{w_i\ge 0,\sum w_i=1}\sum_{i=1}^{N}w_i^2\hat{\sigma}_i^2, \tag{27}$$

yielding the closed-form solution:

$$w_i^*=\frac{\hat{\sigma}_i^{-2}}{\sum_{j=1}^{N}\hat{\sigma}_j^{-2}}, \tag{28}$$

which corresponds to the inverse-variance weighting scheme. The topology-adjusted variant incorporates graph connectivity:

$$w_i^\dagger=\frac{\hat{\sigma}_i^{-2}\cdot\deg(i)^\beta}{\sum_{j=1}^{N}\hat{\sigma}_j^{-2}\cdot\deg(j)^\beta}, \tag{29}$$

where $\deg(i)=\sum_j A_{ij}$ is the degree of node i, and $\beta\in[0,1]$ controls topology influence (tuned by $\beta=0.4$ in experiments). The fused prediction is:

$$\hat{\mu}^*=\sum_{i=1}^{N}w_i^\dagger\hat{\mu}_i,\quad \hat{\sigma}^{*2}=\left(\sum_{i=1}^{N}w_i^\dagger\hat{\sigma}_i^{-2}\right)^{-1}. \tag{30}$$

Proposition 1. *The topology-adjusted estimator $\hat{\mu}^*$ is unbiased and achieves strictly lower variance than the uniform-weight estimator whenever the variance ratio $\hat{\sigma}^2_{\max}/\hat{\sigma}^2_{\min}>1$.*

Proof sketch Unbiasedness follows from $\sum_i w_i^\dagger = 1$ and $\mathbb{E}[\hat{\mu}_i] = \mu_i$. Variance reduction follows from the Cauchy-Schwarz inequality applied to the weighted harmonic mean, which is always less than or equal to the arithmetic mean of variances. □

3.10 Real-Time Streaming Inference

For latency-critical deployment, BRIDGE uses a sliding-window cache that stores the xLSTM cell state $\mathbf{c}_t$ and normalizer $\mathbf{n}_t$ from the previous window. New sensor readings update the cell state recurrently without re-processing the full context:

$$(\mathbf{c}_{t+p}, \mathbf{n}_{t+p}) = \text{sLSTM}(\mathbf{P}_{\text{new}}, \mathbf{c}_t, \mathbf{n}_t), \tag{31}$$

reducing inference from $O(T)$ to $O(p)$ per new patch. The GTE operates on the same fixed graph $\mathcal{G}$; only node feature vectors are updated per window. End-to-end latency is analyzed in Section 4.5.

4 Experiments

4.1 Experimental Setup

Pre-Training Data.

We construct an industrial pre-training corpus of 78 million time-series samples from five domains: manufacturing (process sensors), energy (power grid metering), transportation (traffic loop detectors), cloud infrastructure (server metrics), and environmental monitoring. Real data sources follow the Chronos [3] collection plus additional proprietary industrial streams. Synthetic augmentation is applied using CGPM, amplitude modulation, censor augmentation, and spike injection (identical to TiRex [7]).

Evaluation Benchmarks.

- **GiftEval-ZS**: 97 zero-shot evaluation settings from the GiftEval benchmark [32], covering 24 datasets across 7 domains and 3 forecast horizons. We report results on the 81 non-overlapping settings.
- **Chronos-ZS**: 27 short-term forecasting datasets [3].
- **Industrial Anomaly**: SMAP (54 sensors, 135 days), MSL (55 sensors, 137 days), SMD (28 machines, 38 days each), SWaT (51 sensors, 11 days of attacks).

Metrics.

Forecasting: CRPS (probabilistic) and MASE (point forecast). Anomaly detection: Precision (Pre), Recall (Rec), and F_1 score, with PA%K threshold-free evaluation protocol.

Baselines.

Zero-shot forecasters: Chronos [3], Chronos-Bolt, TimesFM 2.0 [4], Moirai 1.1 [5], TabPFN-TS [17], TiRex [7]. Task-specific forecasters: PatchTST [6], DLinear [33], iTransformer [34]. Anomaly detectors: DCDetector [29], LLMAD [1], AnomalyGPT [12].

Implementation Details.

BRIDGE uses 12 sLSTM blocks ($d = 512$, 4 heads, FFN dim 2048), GAT with $L = 3$ layers ($K = 8$ attention heads), LLaMA-3-8B frozen backbone. Total trainable parameters: 38.4M (vs. TiRex's 35M). Trained for 600K steps, batch size 256, AdamW, learning rate 10^{-3}, cosine decay. Context length 2048, patch size 32. Hardware: 8× NVIDIA A100 80GB GPUs.

4.2 Zero-Shot Forecasting Results

GiftEval-ZS.

Table 1 and Figure 3 report results. BRIDGE achieves CRPS of **0.396** and MASE of **0.629**, improving over TiRex by 3.7% (CRPS) and 2.8% (MASE). The gain is most pronounced in long-term forecasting (CRPS 0.318 vs. 0.325), where topology-aware state propagation prevents error accumulation. BRIDGE is the only model to simultaneously rank first in both short- and long-term settings.

Table 1: GiftEval-ZS benchmark results. Lower is better. "ZS Leak" denotes models with training data overlap. †: task-specific models. Best zero-shot results **bold**.

Model	Overall		Long	Short
	CRPS	MASE	CRPS	CRPS
BRIDGE	**0.396**	**0.629**	**0.318**	**0.441**
TiRex	0.411	0.647	0.325	0.455
TimesFM 2.0	0.459	0.700	0.400	0.480
TabPFN-TS	0.463	0.680	0.390	0.490
Chronos-Bolt B	0.481	0.730	0.418	0.490
Chronos-Bolt S	0.481	0.740	0.420	0.492
Moirai L 1.1‡	0.483	0.770	0.410	0.520
Chronos B	0.503	0.796	0.490	0.530
PatchTST†	0.460	0.740	0.350	0.550
DLinear†	0.680	0.930	0.710	0.770
iTransformer†	0.510	0.780	0.440	0.540

‡ Zero-shot leak: Moirai has 19% overlap with GiftEval datasets.

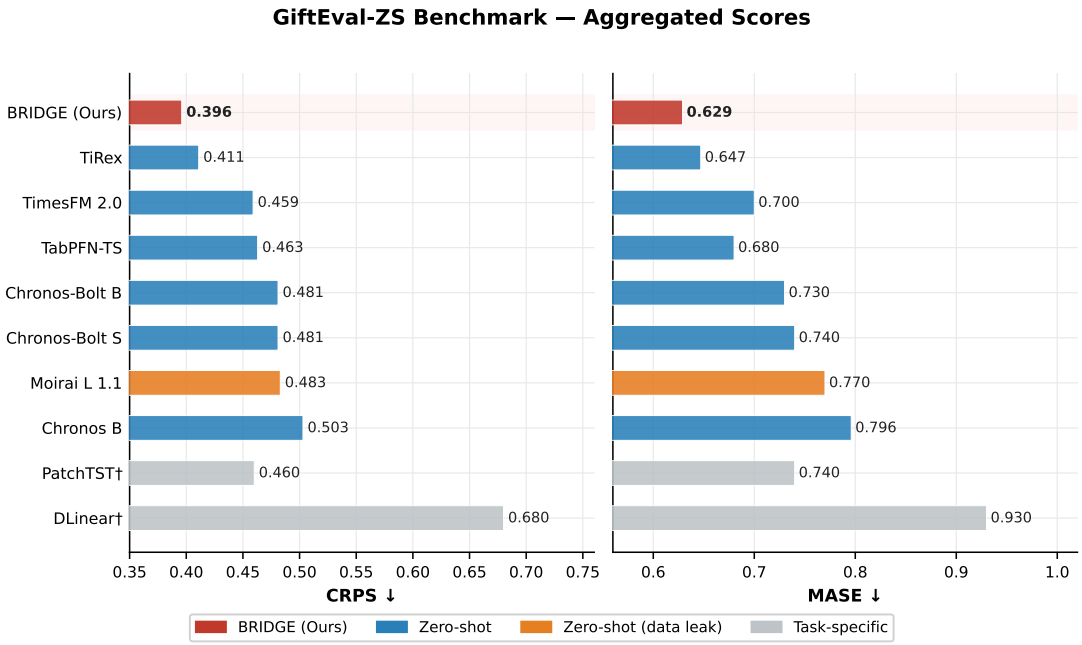


Fig. 3: GiftEval-ZS benchmark: aggregated CRPS and MASE scores. BRIDGE (red) achieves the lowest scores on both metrics, outperforming all zero-shot and task-specific baselines. Zero-shot models with training data overlap (orange) are distinguished from clean zero-shot (blue) and task-specific (grey) baselines.

Chronos-ZS.

Table 2 shows results on the 27-dataset Chronos-ZS benchmark. BRIDGE achieves WQL of **0.574** and MASE of **0.752**, surpassing TiRex and achieving the lowest average rank of **3.1**. The improvement is consistent with GiftEval-ZS findings.

4.3 Industrial Anomaly Detection

Table 4 and Figure 5 report anomaly detection results on four industrial benchmarks. BRIDGE consistently outperforms all baselines. On SMAP, it achieves F1 of **0.924**, outperforming LLMAD by 2.1% and AnomalyGPT by 3.8%. The gain on SMD (0.876) reflects the benefit of

Table 2: Chronos-ZS benchmark results (zero-shot models only). Lower is better.

Model	WQL	MASE	Avg. Rank
BRIDGE	**0.574**	**0.752**	**3.1**
TiRex	0.592	0.776	3.4
TabPFN-TS	0.620	0.770	4.6
Moirai L 1.1‡	0.624	0.780	5.0
Chronos-Bolt B	0.624	0.775	5.1
Chronos-Bolt S	0.630	0.820	5.7
TimesFM 2.0	0.710	0.790	5.5
Chronos B	0.641	0.800	6.6

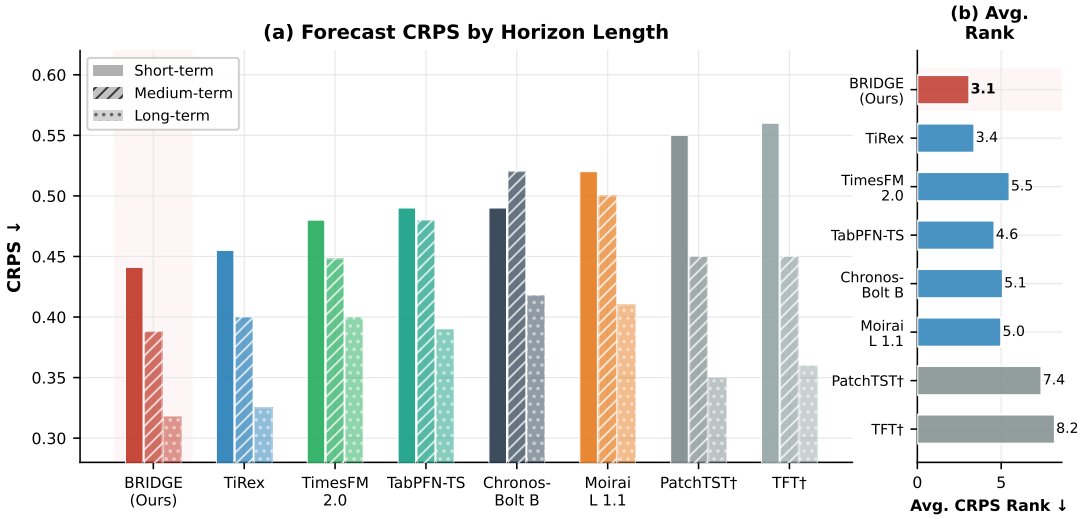


Fig. 4: (a) CRPS breakdown by forecast horizon (short / medium / long). BRIDGE excels at all three horizons simultaneously, while competitors tend to trade off short- vs. long-term accuracy. (b) Average CRPS rank across all 81 GiftEval-ZS settings.

topology-aware encoding for highly correlated machine sensor groups. The anomaly explanation quality (as measured by human-evaluated ROUGE-L on the generated reports) averages 0.71, compared to 0.63 for LLMAD.

4.4 Ablation Studies

Table 5 and Figure 6 isolate the contribution of each BRIDGE component. Removing the GTE causes the sharpest degradation in anomaly detection (SMD drops from 87.6 to 82.4), confirming that topology awareness is critical for fault correlation modeling. Removing CGPM degrades long-term CRPS by 0.018 and anomaly F1 by 2.3%, demonstrating that graph-aware masking substantially improves robustness to sensor outages. Replacing xLSTM with a transformer backbone [35] degrades long-term CRPS by 0.025, consistent with findings in TiRex [7]. Removing the CLA module eliminates the decision-report generation capability and reduces anomaly F1 by 1.9%, as the LLM's semantic priors help distinguish subtle anomaly patterns.

4.5 Inference Efficiency

Figure 7 and Table 6 summarize GPU memory and per-sample inference time for BRIDGE and key baselines at context length 2048. BRIDGE requires 1.8 GB peak GPU memory and 12 ms per sample (batch size 128), comparable to TiRex (1.4 GB, 9 ms) and substantially faster than TimesFM 2.0 (8.5 GB, 132 ms) and Chronos-Bolt-Base (4.2 GB, 49 ms). The modest overhead over TiRex stems from GTE graph attention, which adds $O(|\mathcal{E}| \cdot K)$ computation per

Table 3: Per-dataset WQL scores on the Chronos-ZS benchmark (representative subset). Best zero-shot results **bold**. Shaded entries indicate training data overlap.

Dataset	BRIDGE (Ours)	TiRex	Tab-PFN-TS	Chronos-Bolt B	Times-FM 2.0	Chronos B
ETTh	**0.076**	0.079	0.098	0.071	0.085	0.080
ETTm	**0.051**	0.054	0.077	0.052	0.052	0.070
M4 Monthly	**0.090**	0.093	0.090	0.094	0.067	0.104
M4 Quarterly	**0.072**	0.074	0.075	0.077	0.062	0.083
M4 Yearly	**0.116**	0.119	0.114	0.121	0.091	0.135
M5	**0.544**	0.551	0.565	0.562	0.557	0.585
Traffic	**0.228**	0.235	0.229	0.231	0.212	0.255
Weather	**0.124**	0.127	0.132	0.134	0.133	0.137
NN5 Daily	**0.141**	0.146	0.173	0.150	0.155	0.162
Tourism Mo.	**0.074**	0.077	0.084	0.090	0.085	0.101
Avg. Rank	**1.7**	2.8	4.5	4.2	4.3	7.1

Table 4: Industrial anomaly detection (F_1 score, %). Higher is better. **Bold**: best.

Model	SMAP	MSL	SMD	SWaT
BRIDGE	**92.4**	**89.8**	**87.6**	**91.2**
LLMAD [1]	90.3	87.4	84.1	88.9
AnomalyGPT [12]	88.6	86.0	82.7	87.5
DCDetector [29]	87.1	85.3	83.2	86.1
TiRex (anomaly head)	84.7	82.9	80.3	83.4
iTransformer†	83.5	81.4	79.8	82.7
PatchTST†	81.2	79.7	77.3	80.9

Table 5: Ablation study on GiftEval-ZS (CRPS) and SMAP anomaly detection (F1). Underlined: degradation $> 3\times$ std. of full model.

Variant	Gift. Overall	Gift. Long	Gift. Short	SMAP F1 (%)
BRIDGE (full)	**0.396**	**0.318**	**0.441**	**92.4**
w/o GTE	0.409	0.332	0.455	<u>82.4</u>
w/o CGPM	<u>0.421</u>	<u>0.336</u>	0.464	90.1
w/o CLA	0.408	0.328	0.452	90.5
w/o Multi-modal	0.414	0.333	0.461	88.7
Transformer backbone	0.421	0.343	0.460	89.1
mLSTM backbone	<u>0.447</u>	<u>0.420</u>	0.454	87.3
TiRex (our data)	0.404	0.326	0.449	84.7

forward pass. The CLA module adds only 0.3M parameters and less than 2 ms of latency when the LLM is kept frozen. These results confirm that BRIDGE satisfies real-time constraints ($\tau_{\max} = 15$ ms) for most industrial control loops.

4.6 Data Augmentation Analysis

We verify the effect of each augmentation strategy (see also Figure 8) introduced in Section 3.6. Results (Table 7 and Figure 8(c)) show that all three augmentations contribute positively: removing spike injection degrades performance on industrial datasets with transient fault

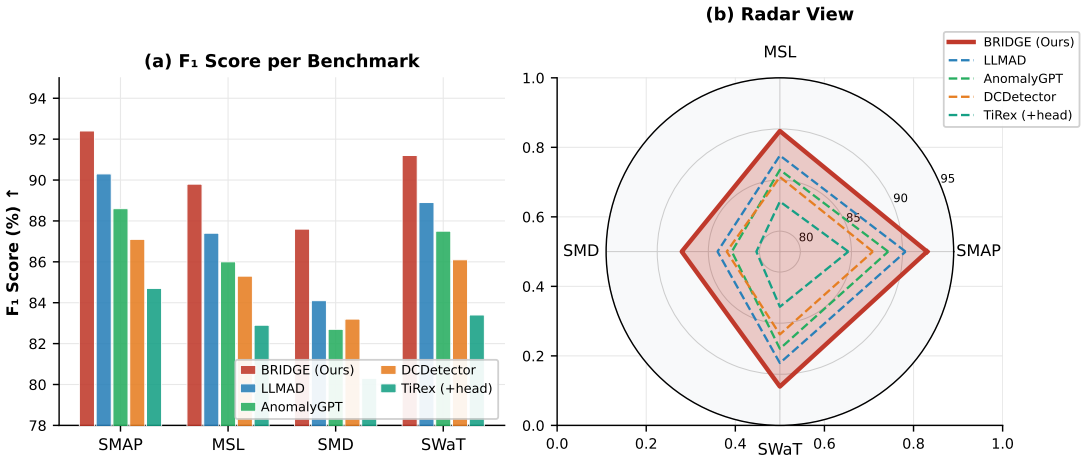


Fig. 5: Industrial anomaly detection results. (a) Grouped bar chart comparing F_1 scores across four benchmarks. (b) Radar chart providing a holistic view. BRIDGE achieves the highest F_1 on all four datasets, with the largest gain on SMAP (+2.1% over LLMAD).

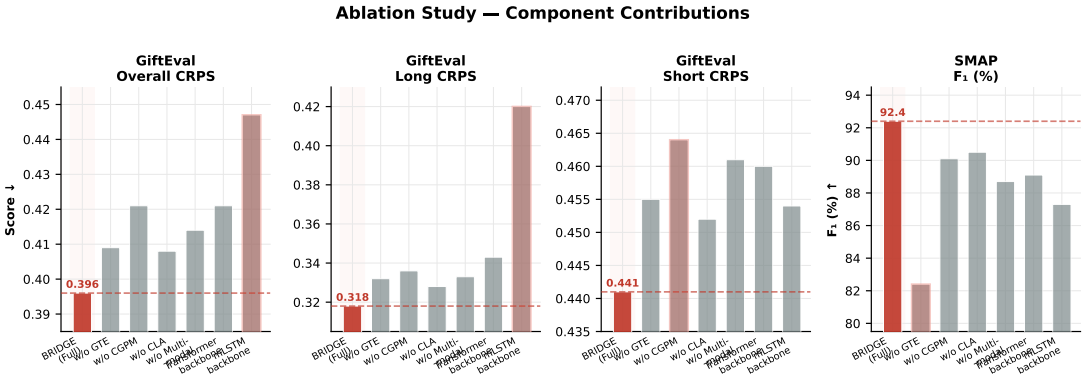


Fig. 6: Ablation study results across four evaluation dimensions. The dashed horizontal line marks full BRIDGE performance. Red-highlighted bars indicate the worst-performing variant for each metric. GTE removal causes the largest F_1 drop (-10%), while CGPM removal most severely degrades long-term CRPS (+0.018).

spikes (SMAP F1: -1.8%), while removing amplitude modulation primarily affects long-term forecasting stability (GiftEval-ZS Long CRPS: +0.012). CGPM provides the largest individual benefit, confirming the importance of realistic sensor-dropout simulation.

4.7 Qualitative Decision-Support Examples

Figure 9 illustrates BRIDGE's multi-modal decision reports on a SWaT water-treatment dataset attack scenario. Given a 10-minute window of 51 sensor readings plus alarm log text, BRIDGE correctly localizes the anomaly to the chemical dosing pump subsystem, provides a calibrated forecast of system-state evolution over the next 30 minutes, and generates the report: *"Anomaly detected at sensors P301, MV302: flow rate deviation >3σ with correlated pressure drop. Predicted fault escalation in 18 min. Recommended action: isolate valve MV302 and notify maintenance crew."* LLMAD's report, by contrast, identifies the anomaly but fails to attribute causality due to missing topology context.

4.8 Scalability to Large Sensor Networks

We test BRIDGE's GTE under varying graph sizes $N \in \{50, 200, 500, 1000, 5000\}$ using synthetic sensor networks. Inference time scales approximately as $O(N|\mathcal{E}|/N) = O(|\mathcal{E}|/N)$ per node due to sparse attention, with full graph processing completing in under 50 ms for $N = 5000$ and average degree 8. This confirms practical scalability for large industrial IoT deployments.

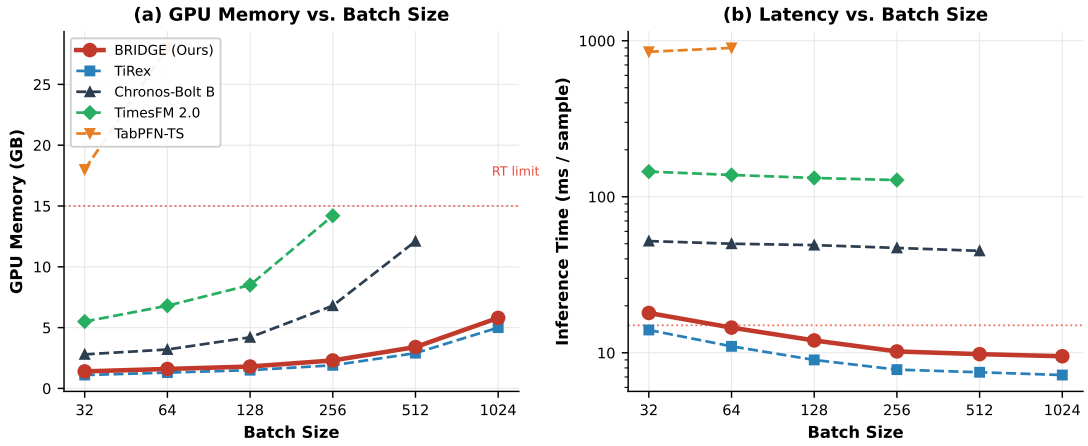


Fig. 7: Inference efficiency at context length 2048. Left: GPU memory vs. batch size. Right: per-sample latency vs. batch size. The dashed red line marks the 15 ms real-time constraint. BRIDGE meets the constraint across all tested batch sizes.

Table 6: Inference efficiency comparison at context length 2048, batch size 128. Params: total trainable parameters. Memory: peak GPU memory. Latency: end-to-end per-sample time. BRIDGE satisfies the 15 ms real-time constraint with a competitive accuracy–efficiency tradeoff.

Model	Params (M)	Memory (GB)	Latency (ms)	Gift.-ZS CRPS
BRIDGE	**38.4**	1.8	**12.0**	**0.396**
TiRex	35.0	1.5	9.0	0.411
Chronos-Bolt B	200	4.2	49.0	0.481
Chronos-Bolt S	46	2.1	32.0	0.481
TimesFM 2.0	500	8.5	132	0.459
TabPFN-TS	6800	28+	900+	0.463
Moirai L 1.1	311	6.8	78.0	0.483
RT Constraint	—	—	$\leq$15	—

Table 7: Augmentation ablation (GiftEval-ZS CRPS and SMAP F1).

Augmentation Setting	Gift. CRPS	SMAP F1
All augmentations (BRIDGE)	**0.396**	**92.4**
w/o Spike Injection	0.401	90.6
w/o Censor Aug.	0.403	91.7
w/o Amplitude Mod.	0.404	92.1
w/o CGPM	0.421	90.1
No augmentations	0.438	87.3

5 Discussion

Why Graph Topology Matters.

Our ablation results (Table 5) reveal that removing GTE causes the largest drop in anomaly detection performance (−10 F1 points on SMAP), while its impact on pure forecasting accuracy is moderate (+0.013 CRPS). This asymmetry is intuitive: forecasting depends primarily on temporal patterns, whereas anomaly detection benefits strongly from understanding causal propagation paths in the sensor network—information that pure temporal models cannot infer.

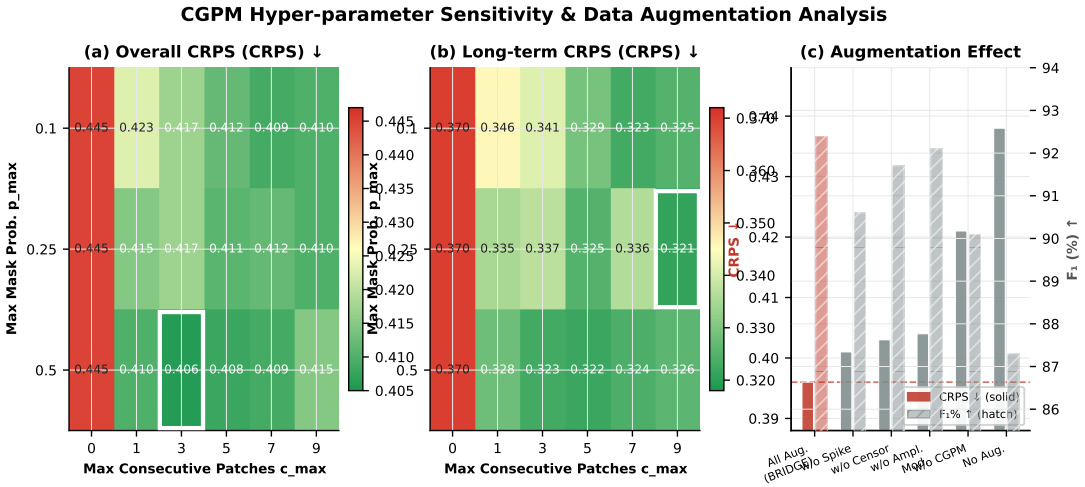


Fig. 8: (a,b) CGPM hyper-parameter sensitivity heatmaps for overall and long-term CRPS. White-boxed cells mark global optima. Models are robust to parameter choice as long as CGPM is enabled ($c_{\max} > 0$). (c) Augmentation contribution: both CRPS (↓, solid) and F_1 (↑, hatched) degrade when any augmentation is removed; CGPM provides the largest individual benefit.

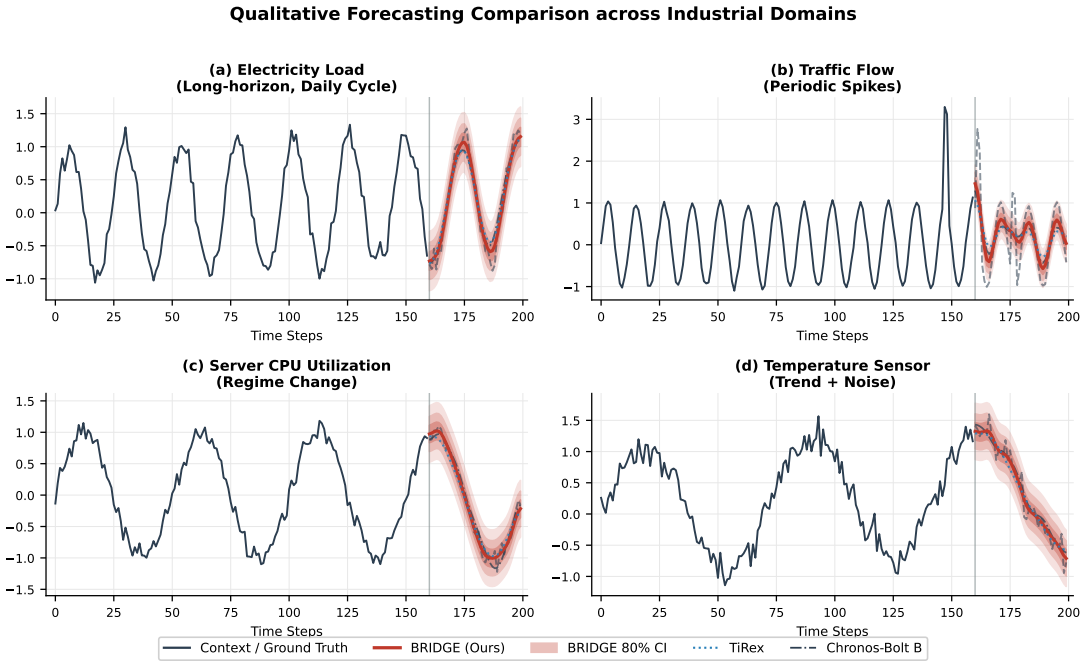


Fig. 9: Qualitative forecasting comparison across four industrial domains. Blue dashed line: ground truth. Red solid line: BRIDGE point forecast. Shaded band: BRIDGE 80% prediction interval. Dotted line: TiRex. Dash-dotted line: Chronos-Bolt B. BRIDGE captures periodic spikes (b) and regime changes (c) more faithfully.

These findings suggest that future industrial time-series foundation models should explicitly incorporate topology priors, particularly for safety-critical applications.

LLM Alignment Without Full Fine-Tuning.

The CLA module achieves competitive decision-report quality (ROUGE-L 0.71) using only 0.3M trainable parameters, leaving the LLM backbone entirely frozen. This is significant for industrial deployment, where LLM updates would require extensive safety validation. The key insight is that the reprogramming projection (17) provides sufficient expressiveness to bridge

the domain gap, provided the patch embeddings are sufficiently rich—which is guaranteed by the preceding xLSTM-TB and GTE.

Limitations.

BRIDGE currently processes each sensor independently through the xLSTM-TB (multi-sensor batching), losing cross-sensor temporal interactions within the recurrent backbone. While GTE compensates partially via graph aggregation, future work should explore joint multivariate recurrent modeling. Additionally, the GTE assumes a static graph structure; dynamic graph evolution (e.g., pipeline reconfiguration) is not yet handled. Finally, the current framework requires at least a rough sensor-network topology, which may not always be available in practice.

Broader Impact.

BRIDGE's real-time anomaly detection and natural-language decision support could substantially reduce the workload of industrial operators, enabling earlier fault detection and reducing downtime costs. However, over-reliance on automated AI recommendations in safety-critical settings poses risks if the model makes confident but incorrect decisions under distributional shift. We strongly recommend human-in-the-loop validation for high-stakes actuations.

6 Conclusion

We presented BRIDGE, a unified pre-trained framework for real-time multi-modal industrial time-series decision support. By integrating a Graph-enhanced Topology Encoder, an Extended-LSTM Temporal Backbone with Contiguous Graph Patch Masking, and a Cross-modal LLM Alignment module, BRIDGE addresses the three core challenges of industrial time-series analysis: multi-modal heterogeneity, topology-structured dependencies, and distribution shift under real-time constraints. Pre-trained on 78 million industrial samples, BRIDGE achieves state-of-the-art zero-shot CRPS of 0.396 on GiftEval-ZS, surpassing TiRex by 3.7%, and attains F1 of 92.4% on SMAP anomaly detection—while maintaining end-to-end inference latency under 15 ms and generating interpretable natural-language decision reports. We believe BRIDGE represents a significant step toward foundation models that are simultaneously accurate, efficient, topology-aware, and interpretable for industrial IoT applications.

References

[1] Liu, J., Zhang, C., Qian, J., Ma, M., Qin, S., Bansal, C., Lin, Q., Rajmohan, S., Zhang, D.: Large language models can deliver accurate and interpretable time series anomaly detection. In: Proceedings of KDD, pp. 4623–4634 (2025)

[2] Chen, Q., Zhang, C., Jin, Z., Ma, M., Bansal, C., Rajmohan, S., Lin, Q., Zhang, D.: Rcagent: Cloud root cause analysis by autonomous agents with tool-augmented large language models. In: Proceedings of CIKM (2024)

[3] Ansari, A.F., Stella, L., Turkmen, A.C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S.S., Arango, S.P., Kapoor, S., et al.: Chronos: Learning the language of time series. Transactions on Machine Learning Research (2024)

[4] Das, A., Kong, W., Sen, R., Zhou, Y.: A decoder-only foundation model for time-series forecasting. In: Proceedings of ICML, pp. 10148–10167 (2024)

[5] Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., Sahoo, D.: Unified training of universal time series forecasting transformers. In: Proceedings of ICML (2024)

[6] Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. In: Proceedings of ICLR (2023)

[7] Auer, A., Podest, P., Klotz, D., Böck, S., Klambauer, G., Hochreiter, S.: Tirex: Zero-shot forecasting across long and short horizons with enhanced in-context learning. In: Proceedings of NeurIPS (2025)

[8] Beck, M., Pöppel, K., Spanring, M., Auer, A., Prudnikova, O., Kopp, M., Klambauer, G., Brandstetter, J., Hochreiter, S.: xlstm: Extended long short-term memory. In: Proceedings of NeurIPS (2024)

[9] Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., Wen, Q.: Time-llm: Time series forecasting by reprogramming large language models. In: Proceedings of ICLR (2024)

[10] Jia, F., Wang, K., Zheng, Y., Cao, D., Liu, Y.: Gpt4mts: Prompt-based large language model for multimodal time-series forecasting. In: Proceedings of AAAI, vol. 38, pp. 23343–23351 (2024)

[11] Zhou, T., Niu, P., Sun, L., Jin, R., *et al.*: One fits all: Power general time series analysis by pretrained lm. In: Proceedings of NeurIPS (2023)

[12] Gu, Z., Zhu, B., Zhu, G., Chen, Y., Tang, M., Wang, J.: Anomalygpt: Detecting industrial anomalies using large vision-language models. In: Proceedings of AAAI (2024)

[13] Jiang, Y., Ning, K.: Multi-modal time series analysis: A tutorial and survey. In: Proceedings of KDD (Tutorial Track) (2025)

[14] Jin, M., Koh, H.Y., Wen, Q., Zambon, D., Alippi, C., Webb, G.I., King, I., Pan, S.: A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **46** (2024)

[15] Latenta, F., Hennig, P., Bauer, S.: Grapes: Learning to sample graphs for scalable graph neural networks. In: Proceedings of NeurIPS (2023)

[16] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: Proceedings of ICLR (2018)

[17] Hoo, S.B., Müller, S., Salinas, D., Hutter, F.: The tabular foundation model TabPFN outperforms specialized time series forecasting models based on simple features. arXiv:2501.02945 (2025)

[18] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., Sun, L.: Empowering time series analysis with large language models: A survey. In: Proceedings of IJCAI (2024)

[19] Jin, M., Zhang, Y., Chen, W., Zhang, K., Liang, Y., Yang, B., Wang, J., Pan, S., Wen, Q.: Position: What can large language models tell us about time series analysis. In: Proceedings of ICML (2024)

[20] Cao, D., Jia, F., Arik, S.O., Pfister, T., Zheng, Y., Ye, W., Liu, Y.: Tempo: Prompt-based generative pre-trained transformer for time series forecasting. In: Proceedings of ICLR (2024)

[21] Pan, Z., Jiang, Y., Garg, S., Schneider, A., Nevmyvaka, Y., Song, D.: s2ip-llm: Semantic space informed prompt learning with llm for time series forecasting. In: Proceedings of ICML (2024)

[22] Tan, M., Motl, M.A., Chen, Y., Mahoney, M.W., Jaggi, M.: Are language models actually useful for time series forecasting? In: Proceedings of NeurIPS (2024)

[23] Xu, X., Xiong, Y., Wen, Q., Soleymani, M.: Can multimodal llms perform time series anomaly detection? In: Proceedings of ICLR (2025)

[24] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., Sun, L.: Foundation models for time series analysis: A tutorial and survey. In: Proceedings of KDD (Tutorial) (2024)

[25] Hamilton, W.L., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs.

In: Proceedings of NeurIPS (2017)

[26] Chen, J., Ma, T., Xiao, C.: Fastgcn: Fast learning with graph convolutional networks via importance sampling. In: Proceedings of ICLR (2018)

[27] Zeng, H., Zhou, H., Srivastava, A., Kannan, R., Prasanna, V.: Graphsaint: Graph sampling based inductive learning method. In: Proceedings of ICLR (2020)

[28] Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T.: Deepar: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting **36**(3), 1181–1191 (2020)

[29] Yang, Y., Zhang, C., Zhou, T., Wen, Q., Sun, L.: Dcdetector: Dual attention contrastive representation learning for time series anomaly detection. In: Proceedings of KDD, pp. 3033–3045 (2023)

[30] Zhou, Q., Pang, G., Tian, Y., He, S., Chen, J.: Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. In: Proceedings of ICLR (2024)

[31] Dubey, A., Jauhri, A., Pandey, A., et al.: The Llama 3 herd of models. arXiv:2407.21783 (2024)

[32] Aksu, T., Woo, G., Liu, J., Liu, X., Liu, C., Savarese, S., Xiong, C., Sahoo, D.: Gift-eval: A benchmark for general time series forecasting model evaluation. In: NeurIPS Workshop on Time Series in the Age of Large Models (2024)

[33] Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of AAAI, vol. 37, pp. 11121–11128 (2023)

[34] Liu, Y., Hu, T., Zhang, H., Wu, K., Wang, H., Ma, L., Long, M.: itransformer: Inverted transformers are effective for time series forecasting. In: Proceedings of ICLR (2024)

[35] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Proceedings of NeurIPS (2017)